allref

October 12, 2009

References

[1] M. A-Razzak and R. C. Glen. Applications of rule-induction in the derivation of quantitative structure-activity relationships. *J Comput Aided Mol Des*, 6(4):349–383, Aug 1992.

[2] J. Abernethy, F. Bach, T. Evgeniou, and J.-P. Vert. Low-rank matrix factorization with attributes. Technical Report cs/0611124, arXiv, 2006.

[3] J. Abernethy, F. Bach, T. Evgeniou, and J.-P. Vert. A new approach to collaborative filtering: operator estimation with spectral regularization. *J. Mach. Learn. Res.*, 2008. In press.

[4] J. Abernethy, F. Bach, T. Evgeniou, and J.-P. Vert. A new approach to collaborative filtering: Operator estimation with spectral regularization. Technical Report 00250231, HAL, 2008.

[5] J. Abernethy, T. Evgeniou, O. Toubia, and J.-P. Vert. Eliciting consumer preferences using robust adaptive choice questionnaires. *IEEE Trans. Knowl. Data Eng.*, 20(2):145–155, 2008.

[6] J. Abernethy, T. Evgeniou, and J.-P. Vert. An optimization framework for adaptive conjoint questionnaire design. Technical report, INSEAD, 2004.

[7] E. Abrahamian, P. C. Fox, L. Naerum, I. T. Christensen, H. Thøgersen, and R. D. Clark. Efficient generation, storage, and manipulation of fully flexible pharmacophore multiplets and their use in 3-D similarity searching. *J. Chem. Inf. Comput. Sci.*, 43(2):458–468, 2003.

[8] F. Abramovich, Y. Benjamini, D. L. Donoho, and I. M. Johnstone. Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Stat.*, 34(2):584–653, 2006.

[9] E. A. Adie, R. R. Adams, K. L. Evans, D. J. Porteous, and B. S. Pickard. Speeding disease gene discovery by sequence based candidate prioritization. *BMC Bioinformatics*, 6:55, 2005.

[10] R. Aebersold and M. Mann. Mass spectrometry-based proteomics. *Nature*, 422(6928):198–207, Mar 2003.

[11] S. Aerts, D. Lambrechts, S. Maity, P. Van Loo, B. Coessens, F. De Smet, L.-C. Tranchevent, B. De Moor, P. Marynen, B. Hassan, P. Carmeliet, and Y. Moreau. Gene prioritization through genomic data fusion. *Nat. Biotechnol.*, 24(5):537–544, May 2006.

[12] A. Aguilera and B. Gómez-González. Genome instability: a mechanistic view of its causes and consequences. *Nat. Rev. Genet.*, 9(3):204–217, Mar 2008.

[13] J. Aires-de Sousa and J. Gasteiger. Prediction of enantiomeric excess in a combinatorial library of catalytic enantioselective reactions. *J Comb Chem*, 7(2):298–301, 2005.

[14] M. A. Aizerman, E. M. Braverman, and L. I. Rozonoér. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837, 1964.

[15] Hirotogu Akaike. Information theory and an extension of the maximum likelihood principle. In Petrov B. N. and Csaki F., editors, *Proc. of the 2nd Int. Symp. on Information Theory*, pages 267–281, 1973.

[16] T. Akutsu, S. Miyano, and S. Kuhara. Algorithms for identifying Boolean networks and related biological networks based on matrix multiplication and fingerprint function. *J. Comput. Biol.*, 7(3-4):331–343, 2000.

[17] T. Akutsu, S. Miyano, and S. Kuhara. Inferring qualitative relations in genetic networks and metabolic pathways. *Bioinformatics*, 16(8):727–734, 2000.

[18] R. Albert and A.L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, 2002.

[19] R. Albert, H. Jeong, and A.-L. Barabási. Diameter of the World-Wide Web. *Nature*, 401:130–131, 1999.

[20] R. Albert, H. Jeong, and A.-L. Barabási. Attack and error tolerance in complex networks. *Nature*, 406:378–381, 2000.

[21] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell.* Garland Science, Taylor & Francis Group, LLC, 2002. Fourth Edition.

[22] D. G. Albertson, C. Collins, F. McCormick, and J. W. Gray. Chromosome aberrations in solid tumors. *Nat. Genet.*, 34(4):369–376, Aug 2003.

[23] D. G. Albertson and D. Pinkel. Genomic microarrays in human genetic disease and cancer. *Hum. Mol. Genet.*, 12 Spec No 2:R145–R152, Oct 2003.

[24] E. Aldea, J. Atif, and I. Bloch. Image classification using marginalized kernels for graphs. In *Graph-Based Representations in Pattern Recognition*, volume 4538/2007 of *Lecture Notes in Computer Science*, pages 103–113. Springer Berlin / Heidelberg, 2007.

[25] M. Alexandersson, S. Cawley, and L. Pachter. SLAM: cross-species gene finding and alignment with a generalized pair hidden Markov model. *Genome Res.*, 13(3):496–502, Mar 2003.

[26] P.H. Algoet. The strong law of large numbers for sequential decisions under uncertainty. *IEEE Trans. Inform. Theory*, 40(3):609–633, May 1994.

[27] C.F. Aliferis, D.P. Hardin, and P. Massion. Machine Learning Models For Lung Cancer Classification Using Array Comparative Genomic Hybridization. In *Proceedings of the 2002 American Medical Informatics Association (AMIA) Annual Symposium*, pages 7–11, 2002.

[28] A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson, L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, and L. M. Staudt. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, 403(6769):503–511, Feb 2000.

[29] Can Alkan, Jeffrey M Kidd, Tomas Marques-Bonet, Gozde Aksay, Francesca Antonacci, Fereydoun Hormozdiari, Jacob O Kitzman, Carl Baker, Maika Malig, Onur Mutlu, S. Cenk Sahinalp, Richard A Gibbs, and Evan E Eichler. Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet*, 41(10):1061–1067, Oct 2009.

[30] E.L. Allgower and K.Georg. *Numerical continuation methods.* Springer, 1990.

[31] H.A. Almohamad and S.O. Duffuaa. A linear programming approach for the weighted graph matching problem. *IEEE Trans. Inform. Theory*, 15(5):522–525, May 1993.

[32] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. A basic local alignment search tool. *J. Mol. Biol.*, 215:403–410, 1990.

[33] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, 25:3389–3402, 1997.

[34] Y. Altun and T. Hofmann. Large Margin Methods for Label Sequence Learning. In *8th European Conference on Speech Communication and Technology (EuroSpeech)*, 2003.

[35] Y. Altun, T. Hofmann, and A.J. Smola. Gaussian process classification for segmenting and annotating sequences. In *Twenty-first international conference on Machine learning*. ACM Press, 2004.

[36] Y. Altun, A. Smola, and T. Hofmann. Exponential Families for Conditional Random Fields. In *20th Conference on Uncertainty in Artificial Intelligennce*, 2004.

[37] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley. Classes of small-world networks. *Proc. Natl. Acad. Sci. USA*, 97(21):11149–11152, 2000.

[38] S.-I. Amari. Natural Gradient Works Efficiently in Learning. *Neural Computation*, 10(2):251–276, 1998.

[39] S.-I. Amari. Information geometry on hierarchy of probability distributions. *IEEE Trans. Inform. Theory*, 47(5):1701–1711, July 2001.

[40] S.-I. Amari and H. Nagaoka. *Methods of information geometry.* AMS vol. 191, 2001.

[41] S.-I. Amari and S. Wu. Improving support vector machine classifiers by modifying kernel functions. *Neural Networks*, 12(6):783–789, Jul 1999.

[42] M. Amarzguioui and H. Prydz. An algorithm for selection of functional siRNA sequences. *Biochem. Biophys. Res. Commun.*, 316(4):1050–8, Apr 2004.

[43] R. Ambauen, Stefan Fischer 0002, and Horst Bunke. Graph edit distance with node splitting and merging, and its application to diatom idenfication. In *GbRPR*, pages 95–106, 2003.

[44] R. Ambauen, S. Fischer, and H. Bunke. Graph edit distance with node splitting and merging and its application to diatom identification. In *Proceedings of the 4th International Workshop on Graph Based Representations in Pattern Recognition*, pages 95–106, 2003.

[45] C. Ambroise and G.J. McLachlan. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc. Natl. Acad. Sci. USA*, 99(10):6562–6566, 2002.

[46] Yonatan Amit, Michael Fink, Nathan Srebro, and Shimon Ullman. Uncovering shared structures in multiclass classification. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pages 17–24, New York, NY, USA, 2007. ACM.

[47] D.C. Anderson, W. Li, D.G. Payan, and W.S. Noble. A new algorithm for the evaluation of shotgun peptide sequencing in proteomics: support vector machine classification of peptide MS/MS spectra and SEQUEST scores. *J Proteome Res*, 2(2):137–146, 2003.

[48] C. David Andersson, Elin Thysell, Anton Lindström, Max Bylesjö, Florian Raubacher, and Anna Linusson. A multivariate approach to investigate docking parameters' effects on docking performance. *J Chem Inf Model*, 47(4):1673–1687, 2007.

[49] Rie Kubota Ando, Tong Zhang, and Peter Bartlett. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6:1817–1853, 2005.

[50] T. A. Andrea and H. Kalayeh. Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. *J Med Chem*, 34(9):2824–2836, Sep 1991.

[51] S. Andrews, T. Hofmann, and I. Tsochantaridis. Multiple Instance Learning with Generalized Support Vector Machines. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pages 943–944. American Association for Artificial Intelligence, 2002.

[52] Davide Anguita, Sandro Ridella, Fabio Rivieccio, and Rodolfo Zunino. Quantum optimization for training support vector machines. *Neural Netw.*, 16(5-6):763–70, 2003.

[53] K. M. Anstreicher and N. W. Brixius. A new bound for the quadratic assignment problem based on convex quadratic programming. *Mathematical Programming*, 89(3):341–357, 2001.

[54] Iris Antes, Shirley W I Siu, and Thomas Lengauer. DynaPred: a structure and sequence based method for the prediction of MHC class I binding peptide sequences and conformations. *Bioinformatics*, 22(14):e16–e24, Jul 2006.

[55] K. F. Aoki, H. Mamitsuka, T. Akutsu, and M. Kanehisa. A score matrix to reveal the hidden links in glycans. *Bioinformatics*, 21(8):1457–63, Apr 2005.

[56] K. F. Aoki, A. Yamaguchi, N. Ueda, T. Akutsu, H. Mamitsuka, S. Goto, and M. Kanehisa. KCaM (KEGG Carbohydrate Matcher): a software tool for analyzing the structures of carbohydrate sugar chains. *Nucleic Acids Res.*, 32(Web Server issue):W267–72, Jul 2004.

[57] A. Aouba, F. Péquignot, A. Le Toullec, and E. Jougla. Les causes mdicales de décès en France en 2004 et leur volution 1980-2004. *Bulletin épidémiologique hebdomadaire*, 35-36:308–314, 2007.

[58] T. Aoyama, Y. Suzuki, and H. Ichikawa. Neural networks applied to quantitative structure-activity relationship analysis. *J Med Chem*, 33(9):2583–2590, Sep 1990.

[59] Yindalon Aphinyanaphongs, Ioannis Tsamardinos, Alexander Statnikov, Douglas Hardin, and Constantin F Aliferis. Text categorization models for high-quality article retrieval in internal medicine. *J. Am. Med. Inform. Assoc.*, 12(2):207–16, 2005.

[60] M. Arakawa, K. Hasegawa, and K. Funatsu. Application of the novel molecular alignment method using the Hopfield Neural Network to 3D-QSAR. *J Chem Inf Comput Sci*, 43(5):1396–1402, 2003.

[61] A. Argyriou, T. Evgeniou, and M. Pontil. Multi-task feature learning. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Adv. Neural. Inform. Process Syst. 19*, pages 41–48, Cambridge, MA, 2007. MIT Press.

[62] A. Argyriou, T. Evgeniou, and M. Pontil. Convex multi-task feature learning. *Mach. Learn.*, 73(3):243–272, 2008. To appear.

[63] Andreas Argyriou, Charles A. Micchelli, and Massimiliano Pontil. When is there a representer theorem? vector versus matrix regularizers. *CoRR*, abs/0809.1590, 2008.

[64] Andreas Argyriou, Charles A. Micchelli, Massimiliano Pontil, and Yiming Ying. A spectral regularization framework for multi-task structure learning. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 25–32. MIT Press, Cambridge, MA, 2008.

[65] Rieko Arimoto, Madhu-Ashni Prasad, and Eric M Gifford. Development of CYP3A4 inhibition models: comparisons of machine-learning techniques and molecular descriptors. *J Biomol Screen*, 10(3):197–205, Apr 2005.

[66] J. W. Armstrong. A review of high-throughput screening approaches for drug discovery. Application note, 1999.

[67] Tomasz Arodź, Marcin Kurdziel, Erik O D Sevre, and David A Yuen. Pattern recognition techniques for automatic detection of suspicious-looking anomalies in mammograms. *Comput. Methods Programs Biomed.*, 79(2):135–49, Aug 2005.

[68] A. M. Aronov. Predictive in silico modeling for hERG channel blockers. *Drug Discov. Today*, 10(2):149–155, Jan 2005.

[69] N. Aronszajn. Theory of reproducing kernels. *Trans. Am. Math. Soc.*, 68:337 – 404, 1950.

[70] Tirusew Asefa, Mariush Kemblowski, Gilberto Urroz, and Mac McKee. Support vector machines (SVMs) for monitoring network design. *Ground Water*, 43(3):413–22, 2005.

[71] V. Atalay and R. Cetin-Atalay. Implicit motif distribution based hybrid computational kernel for sequence classification. *Bioinformatics*, 21(8):1429–1436, Apr 2005.

[72] T. K. Attwood, P. Bradley, D. R. Flower, A. Gaulton, N. Maudling, A. L. Mitchell, G. Moulton, A. Nordle, K. Paine, P. Taylor, A. Uddin, and C. Zygouri. Prints and its automatic supplement, preprints. *Nucleic Acids Res*, 31(1):400–402, Jan 2003.

[73] Cédric Auliac, Vincent Frouin, Xavier Gidrol, and Florence d'Alché Buc. Evolutionary approaches for the reverse-engineering of gene regulatory networks: A study on a biologically realistic dataset. *BMC Bioinformatics*, 9, 2008.

[74] Shai Avidan. Support vector tracking. *IEEE Trans Pattern Anal Mach Intell*, 26(8):1064–72, Aug 2004.

[75] V. A. Avlani, K. J. Gregory, C. J. Morton, M. W. Parker, P. M. Sexton, and A. Christopoulos. Critical role for the second extracellular loop in the binding of both orthosteric and allosteric g protein-coupled receptor ligands. *J. Biol. Chem.*, 282(35):25677–25686, Aug 2007.

[76] C.-A. Azencott, A. Ksikes, S. J. Swamidass, J. H. Chen, L. Ralaivola, and P. Baldi. One- to four-dimensional kernels for virtual screening and the prediction of physical, chemical, and biological properties. *J. Chem. Inf. Model.*, 47(3):965–974, 2007.

[77] O Babur, E Demir, A Ayaz, U Dogrusoz, and O Sakarya. Pathway activity inference using microarray data. *Technical report, Bilkent Center for Bioinformatics (BCBI)*, 2004.

[78] F. Bach. Consistency of the group lasso and multiple kernel learning. *J. Mach. Learn. Res.*, 9:1179–1225, 2008.

[79] F. Bach. Exploring large feature spaces with hierarchical multiple kernel learning. In *Adv. Neural. Inform. Process Syst.*, volume 21, 2009.

[80] F. R. Bach. Consistency of trace norm minimization. *J. Mach. Learn. Res.*, 9:1019–1048, 2008.

[81] F. R. Bach and M. I. Jordan. Predictive low-rank decomposition for kernel methods. In *ICML '05: Proceedings of the 22nd international conference on Machine learning*, pages 33–40, New York, NY, USA, 2005. ACM.

[82] F. R. Bach, G. Lanckriet, and M. I. Jordan. Fast kernel learning using sequential minimal optimization. Technical Report UCB/CSD-04-1307, Computer Science Division, UC Berkeley, February 2004.

[83] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan. Multiple kernel learning, conic duality, and the SMO algorithm. In *ICML '04: Proceedings of the twenty-first international conference on Machine learning*, page 6, New York, NY, USA, 2004. ACM.

[84] F. R. Bach, R. Thibaux, and M. I. Jordan. Computing regularization paths for learning multiple kernels. In L. K. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems 17*, pages 73–80, Cambridge, MA, 2005. MIT Press.

[85] F.R. Bach and M.I. Jordan. Kernel independent component analysis. *J. Mach. Learn. Res.*, 3:1–48, 2002.

[86] Francis R. Bach. Bolasso: model consistent lasso estimation through the bootstrap. In *ICML '08: Proceedings of the 25th international conference on Machine learning*, pages 33–40, New York, NY, USA, 2008. ACM.

[87] Francis R. Bach and Michael I. Jordan. Learning spectral clustering. In *Advances in Neural Information Processing Systems 16*. MIT Press, 2003.

[88] Magdalena Bacilieri and Stefano Moro. Ligand-based drug design methodologies in drug discovery process: an overview. *Curr Drug Discov Technol*, 3(3):155–165, Sep 2006.

[89] Harmohina Bagga, David S Greenfield, and William J Feuer. Quantitative assessment of atypical birefringence images using scanning laser polarimetry with variable corneal compensation. *Am J Ophthalmol*, 139(3):437–46, Mar 2005.

[90] A. M. Bagirov, B. Ferguson, S. Ivkovic, G. Saunders, and J. Yearwood. New algorithms for multi-class cancer diagnosis using tumor gene expression signatures. *Bioinformatics*, 19(14):1800–7, Sep 2003.

[91] Pantelis G Bagos, Theodore D Liakopoulos, and Stavros J Hamodrakas. Evaluation of methods for predicting the topology of beta-barrel outer membrane proteins and a consensus prediction method. *BMC Bioinformatics*, 6(1):7, Jan 2005.

[92] T. L. Bailey and C. Elkan. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology*, 2:28–36, 1994.

[93] Timothy L. Bailey, Nadya Williams, Chris Misleh, and Wilfred W. Li. Meme: discovering and analyzing dna and protein sequence motifs. *Nucl. Acids Res.*, 34(suppl_2):W369–373, July 2006.

[94] H. Baird. Document image defect models and their uses. In *Proceedings of the Second International Conference on Document Analysis and Recognition ICDAR-93*, 1993.

[95] J. Bajorath. Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *J Chem Inf Comput Sci*, 41(2):233–245, 2001.

[96] J. Bajorath. Integration of virtual and high-throughput screening. *Nat Rev Drug Discov*, 1(11):882–894, Nov 2002.

[97] B. Bakker and T. Heskes. Task clustering and gating for bayesian multitask learning. *J. Mach. Learn. Res.*, 4:83–99, 2003.

[98] K. V. Balakin, S. E. Tkachenko, S. A. Lang, I. Okun, A. A. Ivashchenko, and N. P. Savchuk. Property-based design of GPCR-targeted library. *J. Chem. Inf. Comput. Sci.*, 42(6):1332–1342, 2002.

[99] M. Balasubramanian and E. L. Schwartz. The isomap algorithm and topological stability. *Science*, 295(5552):7, Jan 2002.

[100] P. Baldi, S. Brunak, P. Frasconi, G. Soda, and G. Pollastri. Exploiting the past and the future in protein secondary structure prediction. *Bioinformatics*, 15:937–946, 1999.

[101] P. Baldi, Y. Chauvin, T. Hunkapiller, and M.A. McClure. Hidden Markov models of biological primary sequence information. *Proc. Natl. Acad. Sci. USA*, 91(3):1053–1063, 1994.

[102] J. Ballesteros and K. Palczewski. G protein-coupled receptor drug discovery: implications from the crystal structure of rhodopsin. *Curr. Opin. Drug Discov. Devel.*, 4(5):561–574, Sep 2001.

[103] A. Balmain, J. Gray, and B. Ponder. The genetics and genomics of cancer. *Nat. Genet.*, 33:238–244, 2003.

[104] S. Bandyopadhyay, R. Sharan, and T. Ideker. Systematic identification of functional orthologs based on protein network comparison. *Genome Res.*, 16(3):428–435, Mar 2006.

[105] M. Bansal, G. Della Gatta, and D. Bernardo. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics*, 22(7):815–822, Apr 2006.

[106] Mukesh Bansal, Vincenzo Belcastro, Alberto Ambesi-Impiombato, and Diego di Bernardo. How to infer gene networks from expression profiles. *Mol Syst Biol*, 3:78, 2007.

[107] L. Bao and Z. Sun. Identifying genes related to drug anticancer mechanisms using support vector machine. *FEBS Lett.*, 521:109–114, 2002.

[108] Lei Bao. Identifying genes related to chemosensitivity using support vector machine. *Methods Mol Med*, 111:233–40, 2005.

[109] Lei Bao and Yan Cui. Prediction of the phenotypic effects of non-synonymous single nucleotide polymorphisms using structural and evolutionary information. *Bioinformatics*, 21(10):2185–90, May 2005.

[110] A.-L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286:509–512, 1999.

[111] A.-L. Barabási and E. Ravasz. Deterministic scale-free networks. E-print cond-mat/0107419, 2001.

[112] Yannick Baraud, Christophe Giraud, and Sylvie Huet. Gaussian model selection with an unknown variance. *Annals Of Statistics, to appear*, 37:630, 2009.

[113] A. Barron, J. Rissanen, and Bin Yu. The minimum description length principle in coding and modeling. *IEEE Trans. Inform. Theory*, 44(6):2743–2760, Oct 1998.

[114] A.R. Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inform. Theory*, 39(3):930–945, May 1993.

[115] A.R. Barron and T.M. Cover. A bound on the financial value of information. *IEEE Trans. Inform. Theory*, 34(5):1097–1100, Sep 1988.

[116] A.R. Barron and T.M. Cover. Minimum complexity density estimation. *IEEE Trans. Inform. Theory*, 37(4):1034–1054, Jul 1991.

[117] A.R. Barron, L. Györfi, and E.C. van der Meulen. Distribution estimation consistent in total variation and in two types of information divergence. *IEEE Trans. Inform. Theory*, 38(5):1437–1454, Sep 1992.

[118] W. T. Barry, A. B. Nobel, and F. A. Wright. Significance analysis of functional categories in gene expression studies: a structured permutation approach. *Bioinformatics*, 21(9):1943–1949, May 2005.

[119] P. L. Bartlett, O. Bousquet, and S. Mendelson. Local Rademacher Complexities. *Ann. Stat.*, 33(4):1497–1537, 2005.

[120] P. L. Bartlett and A. Tewari. Sparseness vs estimating conditional probabilities: Some asymptotic results. In *Lecture Notes in Computer Science*, volume 3120, pages 564–578. Springer, 2004.

[121] P.I. Bartlett, M.I. Jordan, and J.D. McAuliffe. Convexity, classification and risk bounds. Technical Report 638, UC Berkeley Statistics, 2003.

[122] S.C. Basak, V.R. Magnuson, G.J. Niemi, and R.R. Regal. Determining Structural Similarity of Chemicals Using Graph Theoretic Indices. *Discrete Appl. Math.*, 19:17–44, 1988.

[123] Ali Bashir, Stanislav Volik, Colin Collins, Vineet Bafna, and Benjamin J Raphael. Evaluation of paired-end sequencing strategies for detection of genome rearrangements in cancer. *PLoS Comput Biol*, 4(4):e1000051, Apr 2008.

[124] G. Baudat and F. Anouar. Generalized discriminant analysis using a kernel approach. *Neural Comput*, 12(10):2385–404, Oct 2000.

[125] H. Bauknecht, A. Zell, H. Bayer, P. Levi, M. Wagener, J. Sadowski, and J. Gasteiger. Locating biologically active compounds in medium-sized heterogeneous datasets by topological autocorrelation vectors: dopamine and benzodiazepine agonists. *J Chem Inf Comput Sci*, 36(6):1205–1213, 1996.

[126] D. C. Baulcombe. RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. *Plant Mol. Biol.*, 32(1-2):79–88, Oct 1996.

[127] C. Baumgartner, C. Bohm, D. Baumgartner, G. Marini, K. Weinberger, B. Olgemoller, B. Liebl, and A. A. Roscher. Supervised machine learning techniques for the classification of metabolic disorders in newborns. *Bioinformatics*, 20(17):2985–2996, 2004.

[128] R. Baumgartner, R. Somorjai, C. Bowman, T. C. Sorrell, C. E. Mountford, and U. Himmelreich. Unsupervised feature dimension reduction for classification of MR spectra. *Magn Reson Imaging*, 22(2):251–6, Feb 2004.

[129] Jonathan Baxter. A bayesian/information theoretic model of bias learning. In *COLT '96: Proceedings of the ninth annual conference on Computational learning theory*, pages 77–88, New York, NY, USA, 1996. ACM Press.

[130] Jonathan Baxter. Learning model bias. In *Advances in Neural Information Processing Systems*, pages 169–175. MIT Press, 1996.

[131] Jonathan Baxter. A bayesian/information theoretic model of learning to learn via multiple task sampling. In *Machine Learning*, pages 7–39, 1997.

[132] Jonathan Baxter. A model of inductive bias learning. *Journal of Artificial Intelligence Research*, 12:149–198, 2000.

[133] A. Bazzani, A. Bevilacqua, D. Bollini, R. Brancaccio, R. Campanini, N. Lanconelli, A. Riccardi, and D. Romani. An SVM classifier to separate false signals from microcalcifications in digital mammograms. *Phys Med Biol*, 46(6):1651–63, Jun 2001.

[134] M. J. Beal, F. Falciani, Z. Ghahramani, C. Rangel, and D. L. Wild. A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. *Bioinformatics*, 21(3):349–356, Feb 2005.

[135] O. M. Becker, Y. Marantz, S. Shacham, B. Inbal, A. Heifetz, O. Kalid, S. Bar-Haim, D. Warshaviak, M. Fichman, and S. Noiman. G protein-coupled receptors: in silico drug discovery in 3D. *Proc. Natl. Acad. Sci. USA*, 101(31):11304–11309, Aug 2004.

[136] M. A. Beer and S. Tavazoie. Predicting gene expression from sequence. *Cell*, 117:185–198, 2004.

[137] N. Beerenwinkel, T. Lengauer, M. Daumer, R. Kaiser, H. Walter, K. Korn, D. Hoffmann, and J. Selbig. Methods for optimizing antiviral combination therapies. *Bioinformatics*, 19(Suppl. 1):i16–i25, 2003.

[138] N. Beerenwinkel, B. Schmidt, H. Walter, R. Kaiser, T. Lengauer, D. Hoffman, K. Korn, and J. Selbig. Geno2pheno: Interpreting Genotypic HIV Drug Resistance Tests. *IEEE Intelligent Systems*, 6(6):35–41, 2001.

[139] R. Begg and J. Kamruzzaman. A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. *J Biomech*, 38(3):401–8, Mar 2005.

[140] Rezaul K Begg, Marimuthu Palaniswami, and Brendan Owen. Support vector machines for automated gait classification. *IEEE Trans Biomed Eng*, 52(5):828–38, May 2005.

[141] G. Bejerano and G. Yona. Modeling protein families using probabilistic suffix trees. In *Proceedings of RECOMB 1999*, pages 15–24. ACM Press, 1999.

[142] G. Bejerano and G. Yona. Variations on probabilistic suffix trees: statistical modeling and prediction of protein families. *Bioinformatics*, 17:23–43, 2001.

[143] M. Belkin and P. Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Comput.*, 15(6):1373–1396, 2003.

[144] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *Transaction on pattern analysis and machine intelligence*, 24, 2002.

[145] S. Ben-David and R. Schuller. Exploiting task relatedness for multiple task learning, 2003.

[146] A. Ben-Dor, L. Bruhn, N. Friedman, I. Nachman, M. Schummer, and Z. Yakhini. Tissue classification with gene expression profiles. *J. Comput. Biol.*, 7(3-4):559–583, 2000.

[147] A. Ben-Hur and D. Brutlag. Remote homology detection: a motif based approach. *Bioinformatics*, 19(Suppl. 1):i26–i33, 2003.

[148] A. Ben-Hur, D. Horn, H.T. Siegelmann, and V. Vapnik. Support Vector Clustering. *J. Mach. Learn. Res.*, 2:125–137, 2001.

[149] A. Ben-Hur and W. S. Noble. Kernel methods for predicting protein-protein interactions. *Bioinformatics*, 21(Suppl. 1):i38–i46, Jun 2005.

[150] A. Ben-Hur and W. S. Noble. Choosing negative examples for the prediction of protein-protein interactions. *BMC Bioinformatics*, 7 Suppl 1:S2, 2006.

[151] Y. Bengio, O. Delalleau, N. Le Roux, J.-F. Paiement, P. Vincent, and M. Ouimet. Learning eigenfunctions links spectral embedding and kernel PCA. *Neural Comput.*, 16(10):2197–219, Oct 2004.

[152] Monica Benito, Joel Parker, Quan Du, Junyuan Wu, Dong Xiang, Charles M Perou, and J. S. Marron. Adjustment of systematic microarray data biases. *Bioinformatics*, 20(1):105–14, Jan 2004.

[153] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B*, 57:289–300, 1995.

[154] C. Berg, J. P. R. Christensen, and P. Ressel. *Harmonic analysis on semigroups*. Springer-Verlag, New-York, 1984.

[155] J. Berg and M. Lässig. Cross-species analysis of biological networks by bayesian alignment. *Proc. Natl. Acad. Sci. USA*, 103(29):10967–10972, Jul 2006.

[156] J.O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, 1985.

[157] A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer, 2003.

[158] M. Bern, D. Goldberg, W. H. McDonald, and III Yates, J. R. Automatic Quality Assessment of Peptide Tandem Mass Spectra. *Bioinformatics*, 20(Suppl. 1):i49–i54, 2004.

[159] D. Bernardo, M. J. Thompson, T. S. Gardner, S. E. Chobot, E. L. Eastwood, A. P. Wojtovich, S. J. Elliott, S. E. Schaus, and J. J. Collins. Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nat. Biotechnol.*, 23(3):377–383, Mar 2005.

[160] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. The protein data bank: a computer-based archival file for macromolecular structures. *J Mol Biol*, 112(3):535–542, May 1977.

[161] Stefano Berretti, Alberto Del Bimbo, and Pietro Pala. A graph edit distance based on node merging. In *Proc. of ACM International Conference on Image and Video Retrieval (CIVR)*, pages 464–472, Dublin, Ireland, July 2004.

[162] D. Bertsekas. *Nonlinear programming.* Athena Scientific, 1999.

[163] M. Bhasin and G. P. S. Raghava. Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci.*, 13(3):596–607, Mar 2004.

[164] M. Bhasin and G. P. S. Raghava. Classification of Nuclear Receptors Based on Amino Acid Composition and Dipeptide Composition. *J. Biol. Chem.*, 279(22):23262–23266, 2004.

[165] M. Bhasin and G. P. S. Raghava. ESLpred: SVM-based method for subcellular localization of eukaryotic proteins using dipeptide composition and PSI-BLAST. *Nucl. Acids Res.*, 32(Suppl. 2):W414–419, 2004.

[166] M. Bhasin and G. P. S. Raghava. GPCRpred: an SVM-based method for prediction of families and subfamilies of G-protein coupled receptors. *Nucl. Acids Res.*, 32(Supp.2):W383–389, 2004.

[167] M. Bhasin and G. P. S. Raghava. Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine*, 22(23-24):3195–3204, 2004.

[168] M. Bhasin and G. P. S. Raghava. SVM based method for predicting HLA-DRB1*0401 binding peptides in an antigen sequence. *Bioinformatics*, 20(3):421–423, 2004.

[169] M. Bhasin and G. P. S. Raghava. GPCRsclass: a web tool for the classification of amine type of G-protein-coupled receptors. *Nucleic Acids Res.*, 33(Web Server issue):W143–7, Jul 2005.

[170] M. Bhasin and G. P. S. Raghava. Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. *Nucleic Acids Res*, 33(Web Server issue):W202–7, Jul 2005.

[171] Manoj Bhasin, Harpreet Singh, and G. P S Raghava. MHCBN: a comprehensive database of MHC binding and non-binding peptides. *Bioinformatics*, 19(5):665–666, Mar 2003.

[172] A. Bhattacharjee, W. G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, M. Loda, G. Weber, E. J. Mark, E. S. Lander, W. Wong, B. E. Johnson, T. R. Golub, D. A. Sugarbaker, and M. Meyerson. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA*, 98(24):13790–13795, Nov 2001.

[173] S. Bhavani, A. Nagargadde, A. Thawani, V. Sridhar, and N. Chandra. Substructure-based support vector machine classifiers for prediction of adverse effects in diverse classes of drugs. *J Chem Inf Model*, 46(6):2478–2486, 2006.

[174] Silvia Biasotti, Simone Marini, Michela Mortara, Giuseppe Patane, Michela Spagnuolo, and Bianca Falcidieno. 3d shape matching through topological structures. In *Discrete Geometry for Computer Imagery*, pages 194–203. Springer Berlin / Heidelberg, 2004.

[175] G. Biau and K. Bleakley. Statistical inference on graphs. *Statistics and Decisions*, 24(2):209–232, 2006.

[176] P. J. Bickel, Y. Ritov, and A. Tsybakov. Simultaneous analysis of lasso and dantzig selector. *Ann. Stat.* To appear.

[177] Steffen Bickel, Jasmina Bogojeska, Thomas Lengauer, and Tobias Scheffer. Multi-task learning for hiv therapy screening. In *ICML*, pages 56–63, 2008.

[178] Steffen Bickel, Michael Brückner, and Tobias Scheffer. Discriminative learning for differing training and test distributions. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pages 81–88. ACM Press, 2007.

[179] A. H. Bild, G. Yao, J. T. Chang, Q. Wang, A. Potti, D. Chasse, M. B. Joshi, D. Harpole, J. M. Lancaster, A. Berchuck, Jr. Olson, J. A., J. R. Marks, H. K. Dressman, M. West, and J. R. Nevins. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature*, 439(7074):353–7, 2006.

[180] Michel Bilello, Salih Burak Gokturk, Terry Desser, Sandy Napel, R. Brooke Jeffrey, and Christopher F Beaulieu. Automatic detection and classification of hypodense hepatic lesions on contrast-enhanced venous-phase CT. *Med Phys*, 31(9):2584–93, Sep 2004.

[181] C. Billerey and L. Boccon-Gibod. Etude des variations inter-pathologistes dans l'évaluation du grade et du stade des tumeurs vésicales. *Progrès en Urologie*, 6:49–57, 1996.

[182] L. Birgé and P. Massart. Minimal penalties for gaussian model selection. *Probab. Theory Relat. Fields*, 138:33–73, 2006.

[183] C.M. Bishop. *Pattern recognition and machine learning.* Springer, 2006.

[184] C. Bissantz, P. Bernard, M. Hibert, and D. Rognan. Protein-based virtual screening of chemical databases. II. are homology models of G-protein coupled receptors suitable targets? *Proteins*, 50(1):5–25, Jan 2003.

[185] Andrew Blake and Andrew Zisserman. *Visual Reconstruction.* MIT Press, 1987.

[186] J. F. Blake. Chemoinformatics - predicting the physicochemical properties of 'drug-like' molecules. *Curr Opin Biotechnol*, 11(1):104–107, Feb 2000.

[187] G. Blanchard. *Méthodes de mélange et d'aggregation d'estimateurs en reconnaissance de formes. Applications aux arbres de decision.* PhD thesis, University Paris 13, January 2001.

[188] G. Blanchard, O. Bousquet, and P. Masssart. Statistical Performance of Support Vector Machine. Submitted Ann.Stat., 2004.

[189] M.B. Blaschko and T. Hofmann. Conformal Multi-Instance Kernels. In NIPS 2006 Workshop on Learning to Compare Examples, 2006.

[190] K. Bleakley, G. Biau, and J.-P. Vert. Supervised reconstruction of biological networks with local models. *Bioinformatics*, 23(13):i57–i65, Jul 2007.

[191] K. H. Bleicher, H.-J. Böhm, K. Müller, and A. I. Alanine. Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov*, 2(5):369–378, May 2003.

[192] Nathan Blow. DNA sequencing: generation next-next. *Nature Methods*, 5:267–274, 2008.

[193] J. R. Bock and D. A. Gough. Predicting protein-protein interactions from primary structure. *Bioinformatics*, 17(5):455–460, 2001.

[194] J. R. Bock and D. A. Gough. A New Method to Estimate Ligand-Receptor Energetics. *Mol Cell Proteomics*, 1(11):904–910, 2002.

[195] J. R. Bock and D. A. Gough. Whole-proteome interaction mining. *Bioinformatics*, 19(1):125–134, 2003.

[196] J. R. Bock and D. A. Gough. Virtual screen for ligands of orphan G protein-coupled receptors. *J. Chem. Inf. Model.*, 45(5):1402–1414, 2005.

[197] Mary Ellen Bock, Claudio Garutti, and Conettina Guerra. Effective labeling of molecular surface points for cavity detection and location of putative binding sites. *Comput Syst Bioinformatics Conf*, 6:263–274, 2007.

[198] J. Bockaert and J. P. Pin. Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J.*, 18(7):1723–1729, Apr 1999.

[199] Q. Boese, D. Leake, A. Reynolds, S. Read, S. A. Scaringe, W. S. Marshall, and A. Khvorova. Mechanistic insights aid computational short interfering RNA design. *Methods Enzymol.*, 392:73–96, 2005.

[200] H.-J. Böhm, G. Schneider, R. Mannhold, H. Kubinyi, and G. Folkers. *Protein-ligand interactions.* Wiley, 2003.

[201] F. Bonachéra and D. Horvath. Fuzzy tricentric pharmacophore fingerprints. 2. application of topological fuzzy pharmacophore triplets in quantitative structure-activity relationships. *J. Chem. Inf. Model.*, 48(2):409–425, Feb 2008.

[202] F. Bonachéra, B. Parent, F. Barbosa, N. Froloff, and D. Horvath. Fuzzy tricentric pharmacophore fingerprints. 1. topological fuzzy pharmacophore triplets and adapted molecular similarity scoring schemes. *J. Chem. Inf. Model.*, 46(6):2457–2477, 2006.

[203] Edwin Bonilla, Kian Ming Chai, and Chris Williams. Multi-task gaussian process prediction. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20.* MIT Press, Cambridge, MA, 2008.

[204] Edwin V. Bonilla, Felix V. Agakov, and Christopher K. I. Williams. Kernel multi-task learning using task-specific features. In *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics.* Omnipress, March 2007.

[205] A. Boobis, U. Gundert-Remy, P. Kremers, P. Macheras, and O. Pelkonen. In silico prediction of ADME and pharmacokinetics. Report of an expert meeting organised by COST B15. *Eur J Pharm Sci*, 17(4-5):183–193, Dec 2002.

[206] F.L. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *Transaction on pattern analysis and machine intelligence*, 11:567–585, 1989.

[207] Andrew J Bordner and Ruben Abagyan. Statistical analysis and prediction of protein-protein interfaces. *Proteins*, 60(3):353–66, Aug 2005.

[208] Karsten M. Borgwardt and Hans-Peter Kriegel. Shortest-path kernels on graphs. In *ICDM '05: Proceedings of the Fifth IEEE International Conference on Data Mining*, pages 74–81, Washington, DC, USA, 2005. IEEE Computer Society.

[209] K.M. Borgwardt, C.S. Ong, S. Schönauer, S.V.N. Vishwanathan, A.J. Smola, and H.-P. Kriegel. Protein function prediction via graph kernels. *Bioinformatics*, 21(Suppl. 1):i47–i56, Jun 2005.

[210] J. M. Borwein and A. S. Lewis. *Convex Analysis and Nonlinear Optimization.* Springer-Verlag, New York, 2000.

[211] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the 5th annual ACM workshop on Computational Learning Theory*, pages 144–152, New York, NY, USA, 1992. ACM Press.

[212] H. R. Bosshard. Molecular recognition by induced fit: how fit is the concept? *News Physiol Sci*, 16:171–173, Aug 2001.

[213] J. Boström. Reproducing the conformations of protein-bound ligands: a critical evaluation of several popular conformational searching tools. *J Comput Aided Mol Des*, 15(12):1137–1152, Dec 2001.

[214] J. Boström, J. R. Greenwood, and J. Gottfries. Assessing the performance of OMEGA with respect to retrieving bioactive conformations. *J. Mol. Graph. Model.*, 21(5):449–462, Mar 2003.

[215] L. Bottou, O. Chapelle, D. DeCoste, and J. Weston, editors. *Large-scale kernel machines*. MIT Press, 2007.

[216] Alexandre Bouchard-Côté, Michael I. Jordan, and Dan Klein. Efficient inference in phylogenetic indel trees. In Daphne Koller, Dale Schuurmans, Yoshua Bengio, and Léon Bottou, editors, *NIPS*, pages 177–184. MIT Press, 2008.

[217] S. Boucheron, G. Lugosi, and P. Massart. A sharp concentration inequality with applications. *Random Structures and Algorithms*, 16:277–292, 2000.

[218] Christopher Bowd, Kwokleung Chan, Linda M Zangwill, Michael H Goldbaum, Te-Won Lee, Terrence J Sejnowski, and Robert N Weinreb. Comparing neural networks and linear discriminant functions for glaucoma detection using confocal scanning laser ophthalmoscopy of the optic disc. *Invest Ophthalmol Vis Sci*, 43(11):3444–54, Nov 2002.

[219] Christopher Bowd, Felipe A Medeiros, Zuohua Zhang, Linda M Zangwill, Jiucang Hao, Te-Won Lee, Terrence J Sejnowski, Robert N Weinreb, and Michael H Goldbaum. Relevance vector machine and support vector machine classifier analysis of scanning laser polarimetry retinal nerve fiber layer measurements. *Invest Ophthalmol Vis Sci*, 46(4):1322–9, Apr 2005.

[220] Christopher Bowd, Linda M Zangwill, Felipe A Medeiros, Jiucang Hao, Kwokleung Chan, Te-Won Lee, Terrence J Sejnowski, Michael H Goldbaum, Pamela A Sample, Jonathan G Crowston, and Robert N Weinreb. Confocal scanning laser ophthalmoscopy classifiers and stereophotograph evaluation for prediction of visual field abnormalities in glaucoma-suspect eyes. *Invest Ophthalmol Vis Sci*, 45(7):2255–62, Jul 2004.

[221] J. M. Bower and H. Bolouri. *Computational modeling of genetic and biochemical networks*. MIT Press, Cambridge, MA, 2001.

[222] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, New York, NY, USA, 2004.

[223] P. Boyle and J. Ferlay. Cancer incidence and mortality in europe, 2004. *Ann. Oncol.*, 16(3):481–488, Mar 2005.

[224] L. Boysen, A. Kempe, V. Liebscher, A. Munk, and O. Wittich. Consistencies and rates of convergence of jump-penalized least squares estimators. *Ann. Stat.*, 37(1):157–183, 2009.

[225] Z. Bozdech and H. Ginsburg. Antioxidant defense in Plasmodium falciparum - data mining of the transcriptome. *Malaria Journal*, 3(1):23, 2004.

[226] Z. Bozdech, M. Llinas, B. L. Pulliam, E. D. Wong, J. Zhu, and J. L. DeRisi. The Transcriptome of the Intraerythrocytic Developmental Cycle of Plasmodium falciparum. *PLoS Biology*, 1(1):e5, 2003.

[227] Z. Bozdech, J. Zhu, M. Joachimiak, F. Cohen, B. Pulliam, and J. DeRisi. Expression profiling of the schizont and trophozoite stages of Plasmodium falciparum with a long-oligonucleotide microarray. *Genome Biology*, 4(2):R9, 2003.

[228] James R Bradford and David R Westhead. Improved prediction of protein-protein binding sites using a support vector machines approach. *Bioinformatics*, 21(8):1487–94, Apr 2005.

[229] Sheryl Brahnam, Chao-Fa Chuang, Frank Y Shih, and Melinda R Slack. Machine recognition and representation of neonatal facial displays of acute pain. *Artif. Intell. Med.*, Jun 2005.

[230] M. Bredel and E. Jacoby. Chemogenomics: an emerging strategy for rapid target and drug discovery. *Nat. Rev. Genet.*, 5(4):262–275, Apr 2004.

[231] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *14th Conference on Uncertainty in Artificial Intelligence*, pages 43–52, Madison, W.I., 1998. Morgan Kaufman.

[232] K. Brein, M. Remm, and E. Sonnhammer. Inparanoid: a comprehensive database of eukaryothic orthologs. *Nucleic acids research*, 33, 2005.

[233] Tara M Breslin, Fushen Xu, Gregory M Palmer, Changfang Zhu, Kennedy W Gilchrist, and Nirmala Ramanujam. Autofluorescence and diffuse reflectance properties of malignant and benign breast tissues. *Ann Surg Oncol*, 11(1):65–70, Jan 2004.

[234] Hans Briem and Judith Günther. Classifying "kinase inhibitor-likeness" by using machine-learning methods. *Chembiochem*, 6(3):558–66, Mar 2005.

[235] F.K. Brown. Chemoinformatics : What is it and How does it Impact Drug Discovery. *Annual Reports in Med. Chem.*, 33:375–384, 1998.

[236] M. P. Brown, W. N. Grundy, D. Lin, N. Cristianini, C. W. Sugnet, T. S. Furey, M. Ares, and D. Haussler. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. USA*, 97(1):262–7, Jan 2000.

[237] M.P. Brown, R. Hughey, A. Krogh, I.S. Mian, K. Sjolander, and D. Haussler. Using Dirichlet mixture priors to derive hidden Markov models for protein families. In *Proc. First International Conference on Intelligent Systems for Molecular Biology (ISMB 1993)*, 1993.

[238] P. J. Brown and J. V. Zidek. Adaptive multivariate ridge regression. *Ann. Statist.*, 8(1):64–74, 1980.

[239] P.O. Brown and D. Botstein. Exploring the new world of the genome with DNA microarrays. *Nat. Genet.*, 21:33–37, 2000.

[240] R. D. Brown and Y. C. Martin. The information content of 2D and 3D structural descriptors relevant to ligand-receptor binding. *J Chem Inf Comput Sci*, 37:1–9, 1997.

[241] Robert D. Brown and Yvonne C. Martin. Use of Structure-Activity Data To Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J Chem Inf Comput Sci*, 36:572–584, 1996.

[242] J. P. Brunet, P. Tamayo, T. R. Golub, and J. P. Mesirov. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A*, 101(12):4164–9, 2004.

[243] V. Brusic, N. Petrovsky, G. Zhang, and V. B. Bajic. Prediction of promiscuous peptides that bind HLA class I molecules. *Immunol. Cell Biol.*, 80(3):280–285, Jun 2002.

[244] Samarasena Buchala, Neil Davey, Ray J Frank, Martin Loomes, and Tim M Gale. The role of global and feature based information in gender classification of faces: a comparison of human performance and computational models. *Int J Neural Syst*, 15(1-2):121–8, 2005.

[245] H.-H. Bui, A. J. Schiewe, H. von Grafenstein, and I. S. Haworth. Structural prediction of peptides binding to MHC class I molecules. *Proteins*, 63(1):43–52, Apr 2006.

[246] Huynh-Hoa Bui, John Sidney, Bjoern Peters, Muthuraman Sathiamurthy, Asabe Sinichi, Kelly-Anne Purton, Bianca R Mothé, Francis V Chisari, David I Watkins, and Alessandro Sette. Automated generation and evaluation of specific mhc binding predictive tools: Arb matrix applications. *Immunogenetics*, 57(5):304–314, Jun 2005.

[247] Rogier Buijsman. Structural aspects of kinases and their inhibitors. In *Chemogenomics in Drug Discovery*, chapter 7, pages 191–219. Wiley-VCH, 2005.

[248] P. Bultinck, T. Kuppens, X. Gironès, and R. Carbó-Dorca. Quantum similarity superposition algorithm (QSSA): a consistent scheme for molecular alignment and molecular similarity based on quantum chemistry. *J Chem Inf Comput Sci*, 43(4):1143–1150, 2003.

[249] F. Bunea, A. Tsybakov, and M. Wegkamp. Aggregation and sparsity via l_1 penalized least squares. In G. Lugosi and H. U. Simon, editors, *Proceedings of the 19th Annual Conference on Learning Theory, COLT 2006.*, number 4005 in LNAI, pages 379–391, Berlin Heidelberg, 2006. Springer-Verlag.

[250] F. Bunea, A. Tsybakov, and M. Wegkamp. Sparsity oracle inequalities for the lasso. *Electron. J. Statist.*, 1:169–194, 2007.

[251] R. Bunescu, R. Ge, R. J. Kate, E. M. Marcotte, R. J. Mooney, A. K. Ramani, and Y. W. Wong. Comparative experiments on learning information extractors for proteins and their interactions. *Artif. Intell. Med.*, 33(2):139–55, Feb 2005.

[252] H. Bunke. Inexact graph matching for structural pattern recognition. *Pattern Recognition Letters*, 1(4):245–253, May 1983.

[253] H. Bunke and K. Shearer. A Graph Distance Metric based on the Maximal Common Subgraph. *Pattern Recognition Letters*, 19:255–259, 1998.

[254] R. Burbidge, M. Trotter, B. Buxton, and S. Holden. Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Comput. Chem.*, 26(1):4–15, December 2001.

[255] T. Burckin, R. Nagel, Y. Mandel-Gutfreund, L. Shiue, T. A. Clark, J.-L. Chong, T.-H. Chang, S. Squazzo, G. Hartzog, and M. Ares. Exploring functional relationships between components of the gene expression machinery. *Nat. Struct. Mol. Biol.*, 12(2):175–82, Feb 2005.

[256] C. J. C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Discov.*, 2(2):121–167, 1998.

[257] H. J. Bussemaker, H. Li, and E. D. Siggia. Regulatory element detection using correlation with expression. *Nat. Genet.*, 27:167–174, 2001.

[258] S. Busuttil, J. Abela, and G. J. Pace. Support vector machines with profile-based kernels for remote protein homology detection. *Genome Inform Ser Workshop Genome Inform*, 15(2):191–200, 2004.

[259] D. Butina, M. D. Segall, and K. Frankcombe. Predicting ADME properties in silico: methods and models. *Drug Discov Today*, 7(11 Suppl):S83–S88, Jun 2002.

[260] A. J. Butte, P. Tamayo, D. Slonim, T. R. Golub, and I. S. Kohane. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc. Natl. Acad. Sci. USA*, 97(22):12182–12186, Oct 2000.

[261] S. Buus, S. L. Lauemøller, P. Worning, C. Kesmir, T. Frimurer, S. Corbet, A. Fomsgaard, J. Hilden, A. Holm, and S. Brunak. Sensitive quantitative predictions of peptide-MHC binding by a 'query by committee' artificial neural network approach. *Tissue Antigens*, 62(5):378–384, Nov 2003.

[262] E. Byvatov, U. Fechner, J. Sadowski, and G. Schneider. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J Chem Inf Comput Sci*, 43(6):1882–9, 2003.

[263] E. Byvatov and G. Schneider. Support vector machine applications in bioinformatics. *Appl Bioinformatics*, 2(2):67–77, 2003.

[264] Evgeny Byvatov and Gisbert Schneider. SVM-based feature selection for characterization of focused compound collections. *J Chem Inf Comput Sci*, 44(3):993–9, 2004.

[265] T. Caelli and S. Kosinov. An eigenspace projection clustering method for inexact graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(4):515–519, April 2004.

[266] C. Z. Cai, L. Y. Han, Z. L. Ji, X. Chen, and Y. Z. Chen. SVM-Prot: Web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucleic Acids Res*, 31(13):3692–7, Jul 2003.

[267] C.Z. Cai, L.Y. Han, Z.L. Ji, and Y.Z. Chen. Enzyme family classification by support vector machines. *Proteins*, 55(1):66–76, 2004.

[268] C.Z. Cai, W.L. Wang, L.Z. Sun, and Y.Z. Chen. Protein function classification via support vector machine approach. *Math. Biosci.*, 185(2):111–122, 2003.

[269] Y.-D. Cai, X.-J. Liu, X.-B. Xu, and K.-C. Chou. Support vector machines for prediction of protein subcellular location by incorporating quasi-sequence-order effect. *J. Cell. Biochem.*, 84(2):343–348, 2002.

[270] Y.-D. Cai, X.-J. Liu, X.-B. Xu, and G.-P. Zhou. Support Vector Machines for predicting protein structural class. *BMC Bioinformatics*, 2(3):3, 2001.

[271] Y.-D. Cai, X.-J. Liu, X.-B. Xu, and G.-P. Zhou. Prediction of protein structural classes by support vector machines. *Comput. Chem.*, 26(3):293–296, 2002.

[272] Y.-D. Cai, G.-P. Zhou, and K.-C. Chou. Support Vector Machines for Predicting Membrane Protein Types by Using Functional Domain Composition. *Biophys. J.*, 84(5):3257–3263, 2003.

[273] Y.D. Cai, K.Y. Feng, Y.X. Li, and K.C. Chou. Support vector machine for predicting alpha-turn types. *Peptides*, 24(4):629–630, 2003.

[274] Y.D. Cai and S.L. Lin. Support vector machines for predicting rRNA-, RNA-, and DNA-binding proteins from amino acid sequence. *Biochim. Biophys. Acta*, 1648(1-2):127–133, 2003.

[275] Y.D. Cai, S.L. Lin, and K.C. Chou. Support vector machines for prediction of protein signal sequences and their cleavage sites. *Peptides*, 24(1):159–161, 2003.

[276] Y.D. Cai, X.J. Liu, Y.X. Li, X.B. Xu, and K.C. Chou. Prediction of beta-turns with learning machines. *Peptides*, 24(5):665–669, 2003.

[277] Y.D. Cai, X.J. Liu, X.B. Xu, and K.C. Chou. Support vector machines for prediction of protein subcellular location. *Mol. Cell Biol. Res. Commun.*, 4(4):230–234, 2000.

[278] Y.D. Cai, X.J. Liu, X.B. Xu, and K.C. Chou. Support Vector Machines for predicting HIV protease cleavage sites in protein. *J. Comput. Chem.*, 23(2):267–274, 2002.

[279] Y.D. Cai, X.J. Liu, X.B. Xu, and K.C. Chou. Support vector machines for predicting the specificity of GalNAc-transferase. *Peptides*, 23:205–208, 2002.

[280] Y.D. Cai, X.J. Liu, X.B. Xu, and K.C. Chou. Support vector machines for the classification and prediction of beta-turn types. *J. Pept. Sci.*, 8(7):297–301, 2002.

[281] Y.D. Cai, X.J. Liu, X.B. Xu, and K.C. Chou. Support vector machines for prediction of protein domain structural class. *J. Theor. Biol.*, 221(1):115–120, 2003.

[282] Y.D. Cai, P.W. Ricardo, C.H. Jen, and K.C. Chou. Application of SVM to predict membrane protein types. *J. Theor. Biol.*, 226(4):373–376, 2004.

[283] Y.D. Cai, G.P. Zhou, C.H. Jen, S.L. Lin, and K.C. Chou. Identify catalytic triads of serine hydrolases by support vector machines. *J. Theor. Biol.*, 228(4):551–557, 2004.

[284] Yu-Dong Cai and Andrew J Doig. Prediction of Saccharomyces cerevisiae protein functional class from functional domain composition. *Bioinformatics*, 20(8):1292–300, May 2004.

[285] Yu-Dong Cai, Kai-Yan Feng, Wen-Cong Lu, and Kuo-Chen Chou. Using LogitBoost classifier to predict protein structural classes. *J Theor Biol*, Jul 2005.

[286] J. Caldwell, I. Gardner, and N. Swales. An introduction to drug disposition: the basic principles of absorption, distribution, metabolism, and excretion. *Toxicol. Pathol.*, 23(2):102–114, 1995.

[287] G.A. Calin and C. M. Croce. MicroRNA-cancer connection: the beginning of a new tale. *Cancer Res.*, 66(15):7390–7394, Aug 2006.

[288] L. Calzone, A. Gelay, A. Zinovyev, F. Radvanyi, and E. Barillot. A comprehensive modular map of molecular interactions in RB/E2F pathway. *Mol. Syst. Biol.*, 4:173, 2008.

[289] Francesco Camastra and Alessandro Verri. A novel kernel method for clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(5):801–5, May 2005.

[290] Renato Campanini, Danilo Dongiovanni, Emiro Iampieri, Nico Lanconelli, Matteo Masotti, Giuseppe Palermo, Alessandro Riccardi, and Matteo Roffilli. A novel featureless approach to mass detection in digital mammograms based on support vector machines. *Phys Med Biol*, 49(6):961–75, Mar 2004.

[291] Peter J Campbell, Philip J Stephens, Erin D Pleasance, Sarah O'Meara, Heng Li, Thomas Santarius, Lucy A Stebbings, Catherine Leroy, Sarah Edkins, Claire Hardy, Jon W Teague, Andrew Menzies, Ian Goodhead, Daniel J Turner, Christopher M Clee, Michael A Quail, Antony Cox, Clive Brown, Richard Durbin, Matthew E Hurles, Paul A W Edwards, Graham R Bignell, Michael R Stratton, and P. Andrew Futreal. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet*, 40(6):722–729, Jun 2008.

[292] G. Camps-Valls, A.M. Chalk, A.J. Serrano-Lopez, J.D. Martin-Guerrero, and E.L. Sonnhammer. Profiled support vector machines for antisense oligonucleotide efficacy prediction. *BMC Bioinformatics*, 5(135):135, 2004.

[293] E. Candes. The restricted isometry property. *Compte Rendus de l'Académie des Sciences, Paris*, 1(346):589–592, 2008.

[294] E. Candes, J. K. Romberk, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[295] E. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

[296] E. Candes and T. Tao. The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Stat.*, 35(6):2313–2351, 2007.

[297] N. J. Caplen, S. Parrish, F. Imani, A. Fire, and R. A. Morgan. Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. *Proc. Natl. Acad. Sci. USA*, 98(17):9742–9747, Aug 2001.

[298] E. Capriotti, P. Fariselli, and R. Casadio. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.*, 33(Web Server issue):W306–10, Jul 2005.

[299] R. Carbó, L. Leyda, and M. Arnau. How similar is a molecule to another - an electron-density measure of similarity between 2 molecular structures. *Int. J. Qantum Chem.*, 17:1185–1189, 1980.

[300] M. Carcassoni and E. Hancock. Spectral correspondence for point pattern matching. *Pattern Recognition*, 36(1):193–204, January 2003.

[301] R.E. Carhart, D.H. Smith, and R. Venkataraghavan. Atom Pairs as Molecular Features in Structure-Activity Studies: Definitions and Applications. *J Chem Inf Comput Sci*, 25:64–73, 1985.

[302] Monte Carlo, Shuying Li, Dennis K. Pearl, and Hani Doss. Phylogenetic tree construction using markov chain monte carlo. *Journal of the American Statistical Association*, 95:493–508, 1999.

[303] R. J. Carter, I. Dubchak, and S. R. Holbrook. A computational approach to identify genes for functional RNAs in genomic sequences. *Nucl. Acids Res.*, 29(19):3928–3938, 2001.

[304] Rich Caruana. Multitask learning. *Machine Learning*, 28(1):41–75, 1997.

[305] Richard Caruana. Multitask learning: A knowledge-based source of inductive bias. In *Proceedings of the Tenth International Conference on Machine Learning*, pages 41–48. Morgan Kaufmann, 1993.

[306] L. A. Catapano and H. K. Manji. G protein-coupled receptors in major psychiatric disorders. *Biochim. Biophys. Acta*, 1768(4):976–993, Apr 2007.

[307] O. Catoni. Gibbs estimators. Revised version.

[308] O. Catoni. Data Compression and Adaptive Histograms. In Felipe Cucker and J. Maurice Rojas, editors, *Foundations of Computational Mathematics, Proceedings of Smalefest 2000*. World Scientific, 2002.

[309] D. Cavalieri and C. De Filippo. Bioinformatic methods for integrating whole-genome expression results into cellular networks. *Drug Discov Today*, 10(10):727–34, 2005.

[310] A. Cavalli, E. Poluzzi, F. De Ponti, and M. Recanatini. Toward a pharmacophore for drugs inducing the long QT syndrome: insights from a CoMFA study of HERG K(+) channel blockers. *J. Med. Chem.*, 45(18):3844–3853, Aug 2002.

[311] C. N. Cavasotto, A. J. W. Orry, and R. A. Abagyan. Structure-based identification of binding sites, native ligands and potential inhibitors for G-protein coupled receptors. *Proteins*, 51(3):423–433, May 2003.

[312] C. N. Cavasotto, A. J. W. Orry, N. J. Murgolo, M. F. Czarniecki, S. A. Kocsi, B. E. Hawes, K. A. O'Neill, H. Hine, M. S. Burton, J. H. Voigt, R. A. Abagyan, M. L. Bayne, and F. J. Monsma. Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. *J. Med. Chem.*, 51(3):581–588, Feb 2008.

[313] Gavin C Cawley and Nicola L C Talbot. Fast exact leave-one-out cross-validation of sparse least-squares support vector machines. *Neural Netw*, 17(10):1467–75, Dec 2004.

[314] A. Cayley. On the theory of the analytical forms called threes. *Philosophical magazine*, 37(18):374–378, 1859.

[315] A. Cayley. On the mathematical theory of isomers. *Philosophical magazine*, 10(47):444–446, 1874.

[316] A. Cayley. On the theory of the analytical forms called threes, with application to the theory of chemical combinations. *Rep. Brit. Assoc. Sci.*, 4(45):257–305, 1875.

[317] A. Cayley. On the number of univalent radicals c_nh_{2n+1}. *Philosophical magazine*, 18(3):34–35, 1877.

[318] Eranda Cela. Qaudratuc assignment problem library, 2007.

[319] A. M. Chalk, C. Wahlestedt, and E. L. L. Sonnhammer. Improved and automated prediction of effective siRNA. *Biochem. Biophys. Res. Commun.*, 319(1):264–74, Jun 2004.

[320] A. M. Chalk, R. E. Warfinge, P. Georgii-Hemming, and E. L. L. Sonnhammer. siRNAdb: a database of siRNA sequences. *Nucleic Acids Res.*, 33(Database issue):D131–D134, Jan 2005.

[321] Ian Chan, William Wells, Robert V Mulkern, Steven Haker, Jianqing Zhang, Kelly H Zou, Stephan E Maier, and Clare M C Tempany. Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier. *Med Phys*, 30(9):2390–8, Sep 2003.

[322] Kwokleung Chan, Te-Won Lee, Pamela A Sample, Michael H Goldbaum, Robert N Weinreb, and Terrence J Sejnowski. Comparison of machine learning and traditional classifiers in glaucoma diagnosis. *IEEE Trans Biomed Eng*, 49(9):963–74, Sep 2002.

[323] S. K. Chanda and J. S. Caldwell. Fulfilling the promise: drug discovery in the post-genomic era. *Drug Discov Today*, 8(4):168–174, Feb 2003.

[324] Ming-Wei Chang, Chih-Jen Lin, and Ruby Chiu-Hsing Weng. Analysis of switching dynamics with competing support vector machines. *IEEE Trans Neural Netw*, 15(3):720–7, May 2004.

[325] Ruey-Feng Chang, Wen-Jie Wu, Woo Kyung Moon, and Dar-Ren Chen. Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis. *Ultrasound Med Biol*, 29(5):679–86, May 2003.

[326] Ruey-Feng Chang, Wen-Jie Wu, Woo Kyung Moon, and Dar-Ren Chen. Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors. *Breast Cancer Res Treat*, 89(2):179–85, Jan 2005.

[327] Ruey-Feng Chang, Wen-Jie Wu, Woo Kyung Moon, Yi-Hong Chou, and Dar-Ren Chen. Support vector machines for diagnosis of breast tumors on US images. *Acad Radiol*, 10(2):189–97, Feb 2003.

[328] Olivier Chapelle and Zaid Harchaoui. A machine learning approach to conjoint analysis. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems 17*, pages 257–264. MIT Press, Cambridge, MA, 2005.

[329] I. Chavel. *Eigenvalues in Riemannian geometry.* Academic Press, Orlando, Fl., 1984.

[330] J. Chen, S. J. Swamidass, Y. Dou, J. Bruand, and P. Baldi. ChemDB: a public database of small molecules and related chemoinformatics resources. *Bioinformatics*, 21(22):4133–4139, Sep 2005.

[331] J.-Z. Chen, J. Wang, and X.-Q. Xie. Gpcr structure-based virtual screening approach for cb2 antagonist search. *J. Chem. Inf. Model.*, 47(4):1626–1637, 2007.

[332] Jianhui Chen, Lei Tang, Jun Liu, and Jieping Ye. A convex formulation for learning shared structures from multiple tasks. In *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*, pages 137–144, New York, NY, USA, 2009. ACM.

[333] Jiun-Hung Chen and Chu-Song Chen. Reducing SVM classification time using multiple mirror classifiers. *IEEE Trans Syst Man Cybern B Cybern*, 34(2):1173–83, Apr 2004.

[334] K.-C. Chen, T.-Y. Wang, H.-H. Tseng, C.-Y. F. Huang, and C.-Y. Kao. A stochastic differential equation model for quantifying transcriptional regulatory network in Saccharomyces cerevisiae. *Bioinformatics*, 21(12):2883–2890, Jun 2005.

[335] Q. Chen, Z. M. Zhou, Y. G. Qu, P. A. Heng, and D. S. Xia. Level set based auto segmentation of the tagged left ventricle MR images. *Stud Health Technol Inform*, 98:63–5, 2004.

[336] S. S. Chen, D. L. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1998.

[337] Sheng Chen, Xia Hong, and Chris J Harris. Sparse kernel density construction using orthogonal forward regression with leave-one-out test score and local regularization. *IEEE Trans Syst Man Cybern B Cybern*, 34(4):1708–17, Aug 2004.

[338] T. Chen, H. L. He, and G. M. Church. Modeling gene expression with differential equations. *Pac. Symp. Biocomput.*, pages 29–40, 1999.

[339] Wei Chen, Vera Kalscheuer, Andreas Tzschach, Corinna Menzel, Reinhard Ullmann, Marcel Holger Schulz, Fikret Erdogan, Na Li, Zofia Kijas, Ger Arkesteijn, Isidora Lopez Pajares, Margret Goetz-Sothmann, Uwe Heinrich, Imma Rost, Andreas Dufke, Ute Grasshoff, Birgitta Glaeser, Martin Vingron, and H. Hilger Ropers. Mapping translocation breakpoints by next-generation sequencing. *Genome Res*, 18(7):1143–1149, Jul 2008.

[340] X. Chen, A. Russinko III, and S. S. Young. Recursive Partitioning Analysis of a Large Structure-Activity Data Set Using Three-Dimensional Descriptors. *J Chem Inf Comput Sci*, 38:1054–1062, 1998.

[341] Y. Chen and D. Xu. Understanding protein dispensability through machine-learning analysis of high-throughput data. *Bioinformatics*, 21:575–581, Mar 2005.

[342] Y.C. Chen, Y.S. Lin, C.J. Lin, and J.K. Hwang. Prediction of the bonding states of cysteines using the support vector machines based on multiple feature vectors and cysteine state sequences. *Proteins*, 55(4):1036–1042, 2004.

[343] Betty Yee Man Cheng, Jaime G Carbonell, and Judith Klein-Seetharaman. Protein classification based on text document classification techniques. *Proteins*, 58(4):955–70, Mar 2005.

[344] R. Chenna, H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, and J. D. Thompson. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.*, 31(13):3497–3500, Jul 2003.

[345] Vadim Cherezov, Daniel M Rosenbaum, Michael A Hanson, Sren G F Rasmussen, Foon Sun Thian, Tong Sun Kobilka, Hee-Jung Choi, Peter Kuhn, William I Weis, Brian K Kobilka, and Raymond C Stevens. High-resolution crystal structure of an engineered human beta2-adrenergic g protein-coupled receptor. *Science*, 318(5854):1258–1265, Nov 2007.

[346] Vladimir Cherkassky and Yunqian Ma. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Netw*, 17(1):113–26, Jan 2004.

[347] Kelly Rae Chi. The year of sequencing. *Nature Methods*, 5:11–14, 2008.

[348] D. Y. Chiang, P. O. Brown, and M. B. Eisen. Visualizing associations between genome sequences and gene expression data using genome-mean expression profiles. *Bioinformatics*, 17:49S–55S, 2001.

[349] Derek Y Chiang, Gad Getz, David B Jaffe, Michael J T O'Kelly, Xiaojun Zhao, Scott L Carter, Carsten Russ, Chad Nusbaum, Matthew Meyerson, and Eric S Lander. High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat Methods*, 6(1):99–103, Jan 2009.

[350] S. F. Chin, A. E. Teschendorff, J. C. Marioni, Y. Wang, N. L. Barbosa-Morais, N. P. Thorne, J. L. Costa, S. E. Pinder, M. A. van de Wiel, A. R. Green, I. O. Ellis, P. L. Porter, S. Tavaré, J. D. Brenton, B. Ylstra, and C. Caldas. High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol.*, 8(10):R215, 2007.

[351] S.-F. Chin, Y. Wang, N. P. Thorne, A. E. Teschendorff, S. E. Pinder, M. Vias, A. Naderi, I. Roberts, N. L. Barbosa-Morais, M. J. Garcia, N. G. Iyer, T. Kranjac, J. F. R. Robertson, S. Aparicio, S. Tavare, I. Ellis, J. D. Brenton, and C. Caldas. Using array-comparative genomic hybridization to define molecular portraits of primary breast cancers. *Oncogene*, 26(13):1959–1970, September 2006.

[352] K.-C. Chou. Prediction of protein signal sequences and their cleavage sites. *Protein. Struct. Funct. Genet.*, 42:136–139, 2001.

[353] K.-C. Chou. Using subsite coupling to predict signal peptides. *Protein Eng.*, 14(2):75–79, 2001.

[354] K.-C. Chou and Y.-D. Cai. Using Functional Domain Composition and Support Vector Machines for Prediction of Protein Subcellular Location. *J. Biol. Chem.*, 277(48):45765–45769, 2002.

[355] M. L. Chow, E. J. Moler, and I. S. Mian. Identifying marker genes in transcription profiling data using a mixture of feature relevance experts. *Physiol. Genomics*, 5(2):99–111, Mar 2001.

[356] S. Chu, J. DeRisi, M. Eisen, J. Mulholland, D. Botstein, P.O. Brown, and I. Herskowitz. The Transcriptional Program of Sporulation in Budding Yeast. *Science*, 282:699–705, 1998.

[357] Wei Chu, Chong Jin Ong, and S. Sathiya Keerthi. An improved conjugate gradient scheme to the solution of least squares SVM. *IEEE Trans Neural Netw*, 16(2):498–501, Mar 2005.

[358] H.-Y. Chuang, E. Lee, Y.-T. Liu, D. Lee, and T. Ideker. Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.*, 3:140, 2007.

[359] F. R. K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series*. American Mathematical Society, Providence, 1997.

[360] Fan Chung, Linyuan Lu, and Van Vu. Spectra of random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA*, 100(11):6313–6318, May 2003.

[361] Kai-Min Chung, Wei-Chun Kao, Chia-Liang Sun, Li-Lun Wang, and Chih-Jen Lin. Radius margin bounds for support vector machines with the RBF kernel. *Neural Comput*, 15(11):2643–81, Nov 2003.

[362] G. Cianchetta, Y. Li, J. Kang, D. Rampe, A. Fravolini, G. Cruciani, and R. J. Vaz. Predictive models for hERG potassium channel blockers. *Bioorg. Med. Chem. Lett.*, 15(15):3637–3642, Aug 2005.

[363] Mary Cianfrocca and Lori J. Goldstein. Prognostic and predictive factors in early-stage breast cancer. *The Oncologist*, 9:606–616, 2004.

[364] B. S. Clarke and A. R. Barron. Jeffreys' prior is asymptotically least favorable under entropy risk. *J. Stat. Plann. Infer.*, 31(1):37–60, 1994.

[365] B.S. Clarke and A.R. Barron. Information-theoretic asymptotics of Bayes methods. *IEEE Trans. Inform. Theory*, 36(3):453–471, May 1990.

[366] W. S. Cleveland, E. Grosse, and W. M. Shyu. *Statistical Models in S*, chapter Local regression models. Wasworth & Brooks/Cole, 1992.

[367] Bradley P. Coe, Bauke Ylstra, Beatriz Carvalho, Gerrit A. Meijer, Calum MacAulay, and Wan L. Lam. Resolving the resolution of array CGH. *Genomics*, 89:647–653, 2007.

[368] P.S. Coelho, A. Kumar, and M. Snyder. Genome-wide mutant collections: toolboxes for functional genomics. *Curr. Opin. Microbiol.*, 3:309–315, 2000.

[369] C. Cogoni, J. T. Irelan, M. Schumacher, T. J. Schmidhauser, E. U. Selker, and G. Macino. Transgene silencing of the al-1 gene in vegetative cells of Neurospora is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. *EMBO J.*, 15(12):3153–3163, Jun 1996.

[370] Gilles Cohen, Mélanie Hilario, Hugo Sax, Stphane Hugonnet, Christian Pellegrini, and Antoine Geissbuhler. An application of one-class support vector machine to nosocomial infection detection. *Medinfo*, 11(Pt 1):716–20, 2004.

[371] William W. Cohen, Andrew McCallum, and Sam T. Roweis, editors. *Machine Learning, Proceedings of the Twenty-Fifth International Conference (ICML 2008), Helsinki, Finland, June 5-9, 2008*, volume 307 of *ACM International Conference Proceeding Series*. ACM, 2008.

[372] A. Coi, I. Massarelli, L. Murgia, M. Saraceno, V. Calderone, and A. M. Bianucci. Prediction of hERG potassium channel affinity by the CODESSA approach. *Bioorg. Med. Chem.*, 14(9):3153–3159, May 2006.

[373] Jason C Cole, Christopher W Murray, J. Willem M Nissink, Richard D Taylor, and Robin Taylor. Comparing protein-ligand docking programs is difficult. *Proteins*, 60(3):325–332, Aug 2005.

[374] D. Jackson Coleman, Ronald H Silverman, Mark J Rondeau, H. Culver Boldt, Harriet O Lloyd, Frederic L Lizzi, Thomas A Weingeist, Xue Chen, Sumalee Vangveeravong, and Robert Folberg. Noninvasive in vivo detection of prognostic indicators for high-risk uveal melanoma: ultrasound parameter imaging. *Ophthalmology*, 111(3):558–64, Mar 2004.

[375] Nigel Collier and Koichi Takeuchi. Comparison of character-level and part of speech features for name recognition in biomedical texts. *J Biomed Inform*, 37(6):423–35, Dec 2004.

[376] M. Collins and N. Duffy. Convolution Kernels for Natural Language. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Adv. Neural. Inform. Process Syst.*, volume 14, pages 625–632. MIT Press, 2001.

[377] Ronan Collobert, Samy Bengio, and Yoshua Bengio. A parallel mixture of SVMs for very large scale problems. *Neural Comput*, 14(5):1105–14, May 2002.

[378] Pierre Comon. Independent component analysis: a new concept? *Signal Processing*, 36(3):287–314, 1994.

[379] International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, Feb 2001.

[380] D. Conte, P. Foggia, C. Sansone, and M. Vento. Thirty years of graph matching in pattern recognition. *International journal of pattern recognition and artificial intelligence*, 18(3):265–298, 2004.

[381] C. Cooper, M. Harrison, M. Wilkins, and N. Packer. GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources. *Nucleic Acids Res.*, 29:332–335, 2001.

[382] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento. Performance evaluation of the vf graph matching algorithm. In *ICIAP '99: Proceedings of the 10th International Conference on Image Analysis and Processing*, page 1172, Washington, DC, USA, 1999. IEEE Computer Society.

[383] C. Cortes, P. Haffner, and M. Mohri. Rational Kernels. In Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, editors, *Advances in Neural Information Processing Systems 15*. MIT Press, 2003.

[384] B. Coupez and R. A. Lewis. Docking and scoring–theoretically easy, practically impossible? *Curr Med Chem*, 13(25):2995–3003, 2006.

[385] Timothee Cour, Praveen Srinivasan, and Jinbo Shi. Balanced graph matching. In *Advanced in Neural Information Processing Systems*, 2006.

[386] T.M. Cover and E. Ordentlich. Universal portfolios with side information. *IEEE Trans. Inform. Theory*, 42(2):348–363, Mar 1996.

[387] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley, New-York, 1990.

[388] J.K. Cowell and L. Hawthorn. The application of microarray technology to the analysis of the cancer genome. *Current Molecular Medicine*, 7:103–120, 2007.

[389] D. Cox and F. O'Sullivan. Asymptotic analysis of penalized likelihood and related estimators. *Ann. Stat.*, 18:1676–1695, 1990.

[390] David D Cox and Robert L Savoy. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*, 19(2 Pt 1):261–70, Jun 2003.

[391] F. Crick. Central dogma of molecular biology. *Nature*, 227:561–563, 1970.

[392] N. Cristianini and J. Shawe-Taylor. *An introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.

[393] N. Cristianini, J. Shawe-Taylor, and H. Lodhi. Latent semantic kernels. *J. Intell. Inform. Syst.*, 18(2-3):127–152, 2002.

[394] C. M. Croce. Oncogenes and cancer. *N. Engl. J. Med.*, 358(5):502–511, Jan 2008.

[395] M. Csorgo and P. Revesz. Strong Approximations of the Quantile Process. *Ann. Stat.*, 6(4):882–894, July 1978.

[396] F. Cucker and S. Smale. Best choices for regularization parameters in learning theory: on the bias-variance problem. *Foundations of Computational Mathematics*, 2(4):413–428, 2002.

[397] F. Cucker and S. Smale. On the mathematical foundations of learning. *Bull. Amer. Math. Soc*, 39:1–49, 2002.

[398] J. A. Cuff and G. J. Barton. Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Protein. Struct. Funct. Genet.*, 34:508–519, 1999.

[399] Q. Cui, T. Jiang, B. Liu, and S. Ma. Esub8: A novel tool to predict protein subcellular localizations in eukaryotic organisms. *BMC Bioinformatics*, 5(66):66, 2004.

[400] M. E. Curran, I. Splawski, K. W. Timothy, G. M. Vincent, E. D. Green, and M. T. Keating. A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome. *Cell*, 80(5):795–803, Mar 1995.

[401] Keira R. Curtis, Matej Oresic, and Antonio Vidal-Puig. Pathways to the analysis of microarray data. *Trends in Biotechnology*, 23(8):429–435, 2005.

[402] M. Cuturi, K. Fukumizu, and J.-P. Vert. Semigroup kernels on measures. *J. Mach. Learn. Res.*, 6:1169–1198, 2005.

[403] M. Cuturi and J.-P. Vert. A mutual information kernel for strings. In *Proceedings of IJCNN 2004*, pages 1904–1910, 2004.

[404] M. Cuturi and J.-P. Vert. The context-tree kernel for strings. *Neural Network.*, 18(4):1111–1123, 2005.

[405] M. Cuturi and J.-P. Vert. Semigroup kernels on finite sets. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Adv. Neural Inform. Process. Syst.*, volume 17, pages 329–336. MIT Press, Cambridge, MA, 2005.

[406] M. Cuturi, J. P. Vert, O. Birkenes, and T. Matsui. A kernel for time series based on global alignment. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)*, volume 2, pages II–413–II–416, 2007.

[407] J. P. Daily, D. Scanfeld, N. Pochet, K. Le Roch, D. Plouffe, M. Kamal, O. Sarr, S. Mboup, O. Ndir, D.j Wypi, K. Levasseur, E. Thomas, P. Tamayo, C. Dong, Y. Zhou, E. S. Lander, D. Ndiaye, D. Wirth, E. A. Winzeler, J. P. Mesirov, and A. Regev. Distinct physiological states of plasmodium falciparum in malaria-infected patients. *Nature*, 450(7172):1091–1095, Dec 2007.

[408] James A R Dalton and Richard M Jackson. An evaluation of automated homology modelling methods at low target template sequence similarity. *Bioinformatics*, 23(15):1901–1908, Aug 2007.

[409] Georges A Darbellay, Rebecca Duff, Jean-Marc Vesin, Paul-Andr Despland, Dirk W Droste, Carlos Molina, Joachim Serena, Roman Sztajzel, Patrick Ruchat, Theodoros Karapanayiotides, Afksendyios Kalangos, Julien Bogousslavsky, Erich B Ringelstein, and Grald Devuyst. Solid or gaseous circulating brain emboli: are they separable by transcranial ultrasound? *J Cereb Blood Flow Metab*, 24(8):860–8, Aug 2004.

[410] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

[411] H. Daume. Bayesian Multitask Learning with Latent Hierarchies. In *25th Conference on Uncertainty in Artificial Intelligence*, 2009.

[412] E. B. Davies, G. M. L. Gladwell, J. Leydold, and P. F. Stadler. Discrete Nodal Domain Theorems. *Lin. Alg. Appl.*, 336:51–60, 2001.

[413] Jonathan J. Davies, Ian M. Wilson, and Wan L. Lam. Array CGH technologies and their applications to cancer genomes. *Chromosome Research*, 13:237–48, 2005.

[414] J.R. Davies, R.M. Jackson, K.V. Mardia, and C.C. Taylor. The poisson index: a new probabilistic model for protein ligand binding site similarity. *Bioinformatics*, 23(22):3001–3008, Nov 2007.

[415] Matthew N Davies and Darren R Flower. Harnessing bioinformatics to discover new vaccines. *Drug Discov Today*, 12(9-10):389–395, May 2007.

[416] L. Davisson. Universal noiseless coding. *IEEE Trans. Inform. Theory*, 19(6):783– 795, Nov 1973.

[417] T. De Bie, L.-C. Tranchevent, L. M. M. van Oeffelen, and Y. Moreau. Kernel-based data fusion for gene prioritization. *Bioinformatics*, 23(13):i125–i132, Jul 2007.

[418] Morris H. De Groot. *Optimal statistical decisions / Morris H. De Groot.* McGraw-Hill, New York :,, 1970.

[419] Marina de Tommaso, Sebastiano Stramaglia, Jan Mathijs Schoffelen, Marco Guido, Giuseppe Libro, Luciana Losito, Vittorio Sciruicchio, Michele Sardaro, Mario Pellicoro, and Franco Michele Puca. Steady-state visual evoked potentials in the low frequency range in migraine: a study of habituation and variability phenomena. *Int J Psychophysiol*, 49(2):165–74, Aug 2003.

[420] Y. C. de Verdière. Multiplicités des valeurs propres Laplaciens discrets et Laplaciens continus. *Rendiconti di Matematica*, 13:433–460, 1993.

[421] Kalyanmoy Deb and A. Raji Reddy. Reliable classification of two-class cancer data using evolutionary algorithms. *Biosystems*, 72(1-2):111–29, Nov 2003.

[422] A.K. Debnath, R.L. Lopez de Compadre, G. Debnath, A.J. Schusterman, and C. Hansch. Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. correlation with molecular orbital energies and hydrophobicity. *Journal of Medicinal Chemistry*, 34(2):786–797, 1991.

[423] C. Debouck and P. N. Goodfellow. DNA microarrays in drug discovery and development. *Nat Genet*, 21(1 Suppl):48–50, Jan 1999.

[424] D. Decoste and B. Schölkopf. Training invariant support vector machines. *Mach. Learn.*, 46(1-3):161–190, 2002.

[425] S. Degroeve, B. De Baets, Y. Van de Peer, and P. Rouze. Feature subset selection for splice site prediction. *Bioinformatics*, 18(Suppl. 1):S75–S83, 2002.

[426] S. Degroeve, Y. Saeys, B. De Baets, P. Rouze, and Y. Van de Peer. SpliceMachine: predicting splice sites from high-dimensional local context representations. *Bioinformatics*, 21:1332–1338, 2005.

[427] J. Demšar. Statistical Comparisons of Classifiers over Multiple Data Sets. *J. Mach. Learn. Res.*, 7:1–30, 2006.

[428] Y. Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(8):800–810, Aug 2001.

[429] Jayne L Dennis and Karin A Oien. Hunting the primary: novel strategies for defining the origin of tumours. *J Pathol*, 205(2):236–47, Jan 2005.

[430] Meghana Deodhar and Joydeep Ghosh. A framework for simultaneous co-clustering and learning from complex data. In *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 250–259, New York, NY, USA, 2007. ACM.

[431] J. L. DeRisi, V. R. Iyer, and P. O. Brown. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, 278(5338):680–686, 1997.

[432] D. A. Deshpande and R. B. Penn. Targeting G protein-coupled receptor signaling in asthma. *Cell. Signal.*, 18(12):2105–2120, Dec 2006.

[433] M. Deshpande and G. Karypis. Automated Approaches for Classifying Structures. In *Proceedings of the 2nd Workshop on Data Mining in Bioinformatics (BIOKDD '02), 2002*, 2002.

[434] M. Deshpande and G. Karypis. Evaluation of Techniques for Classifying Biological Sequences. In *PAKDD '02: Proceedings of the 6th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, pages 417–431. Springer Verlag, 2002.

[435] M. Deshpande, M. Kuramochi, N. Wale, and G. Karypis. Frequent Substructure-Based Approaches for Classifying Chemical Compounds. *IEEE Transactions on Knowledge and Data Engineering*, 17(8):1036–1050, August 2005.

[436] X. Deupi, N. Dölker, M. L. Lòpez-Rodrìguez, M. Campillo, J. A. Ballesteros, and L. Pardo. Structural models of class a G protein-coupled receptors as a tool for drug design: insights on transmembrane bundle plasticity. *Curr. Top. Med. Chem.*, 7(10):991–998, 2007.

[437] J. Devillers. *Neural Networks in QSAR and Drug Design.* Academic Press, London, 1996.

[438] R. A. DeVore and G. G. Lorentz. *Constructive Approximation.* Springer Grundlehren der Mathematischen Wissenschaften. Springer Verlag, 1993.

[439] A. Devos, L. Lukas, J. A K Suykens, L. Vanhamme, A. R. Tate, F. A. Howe, C. Majs, A. Moreno-Torres, M. van der Graaf, C. Ars, and S. Van Huffel. Classification of brain tumours using short echo time 1H MR spectra. *J Magn Reson*, 170(1):164–75, Sep 2004.

[440] A. Devos, A. W. Simonetti, M. van der Graaf, L. Lukas, J. A K Suykens, L. Vanhamme, L. M C Buydens, A. Heerschap, and S. Van Huffel. The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification. *J Magn Reson*, 173(2):218–28, Apr 2005.

[441] L. Devroye. Automatic pattern recognition: a study of the probability of error. *IEEE Trans. Pattern Anal. Mach. Intell.*, 10(4):530–543, Jul 1988.

[442] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*, volume 31 of *Applications of Mathematics*. Springer, 1996.

[443] L. Devroye and G. Lugosi. *Combinatorial Methods in Density Estimation.* Springer Series in Statistics. Springer, 2000.

[444] Vikas Dhingra, Mukta Gupta, Tracy Andacht, and Zhen F Fu. New frontiers in proteomics research: a perspective. *Int J Pharm*, 299(1-2):1–18, Aug 2005.

[445] Raquel Dias and Walter Filgueira de Azevedo. Molecular docking algorithms. *Curr Drug Targets*, 9(12):1040–1047, Dec 2008.

[446] Casey Diekman, Wei He, Nagabhushana Prabhu, and Harvey Cramer. Hybrid methods for automated diagnosis of breast tumors. *Anal Quant Cytol Histol*, 25(4):183–90, Aug 2003.

[447] Frank Dieterle, Silvia Mller-Hagedorn, Hartmut M Liebich, and Gnter Gauglitz. Urinary nucleosides as potential tumor markers evaluated by learning vector quantization. *Artif. Intell. Med.*, 28(3):265–79, Jul 2003.

[448] T. G. Dietterich. Machine Learning for Sequential Data: A Review. In T. Caelli, editor, *Structural, Syntactic, and Statistical Pattern Recognition; Lecture Notes in Computer Science, Vol. 2396*, pages 15–30. Springer-Verlag, 2002.

[449] T.G. Dietterich, R.H. Lathrop, and T. Lozano-Perez. Solving the Multiple Instance Problem with Axis-Parallel Rectangles. *Artificial Intelligence*, 89(1-2):31–71, 1997.

[450] J. A. DiMasi, R. W. Hansen, and H. G. Grabowski. The price of innovation: new estimates of drug development costs. *J Health Econ*, 22(2):151–185, Mar 2003.

[451] C.H.Q. Ding and I. Dubchak. Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics*, 17:349–358, 2001.

[452] Chris Ding and Hanchuan Peng. Minimum redundancy feature selection from microarray gene expression data. *J Bioinform Comput Biol*, 3(2):185–205, Apr 2005.

[453] Y. Ding and C. E. Lawrence. A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.*, 31(24):7280–301, Dec 2003.

[454] Y. Ding, C. Chan Yu, and C. E. Lawrence. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.*, 32(Web Server issue):W135–W141, Jul 2004.

[455] Marko Djordjevic, Anirvan M Sengupta, and Boris I Shraiman. A biophysical approach to transcription factor binding site discovery. *Genome Res.*, 13(11):2381–90, Nov 2003.

[456] P.D. Dobson and A.J. Doig. Distinguishing enzyme structures from non-enzymes without alignments. *J. Mol. Biol.*, 330(4):771–783, 2003.

[457] P.D. Dobson and A.J. Doig. Predicting enzyme class from protein structure without alignments. *J. Mol. Biol.*, 345(1):187–199, Jan 2005.

[458] A. Doi, H. Matsuno, M. Nagasaki, and S. Miyano. Hybrid Petri net representation of gene regulatory network. In *Proceedings of the Pacific Symposium on Biocomputing*, volume 5, pages 341–352, 2000.

[459] I. Donaldson, J. Martin, B. de Bruijn, C. Wolting, V. Lay, B. Tuekam, S. Zhang, B. Baskin, G.D. Bader, K. Michalickova, T. Pawson, and C.W.V. Hogue. PreBIND and Textomy - mining the biomedical literature for protein-protein interactions using a support vector machine. *BMC Bioinformatics*, 4(1):11, Mar 2003.

[460] Hai-Long Dong and Yan-Fang Sui. Prediction of HLA-A2-restricted CTL epitope specific to HCC by SYFPEITHI combined with polynomial method. *World J Gastroenterol*, 11(2):208–211, Jan 2005.

[461] S. Doniger, T. Hofmann, and J. Yeh. Predicting CNS permeability of drug molecules: comparison of neural network and support vector machine algorithms. *J. Comput. Biol.*, 9(6):849–864, 2002.

[462] P. Dönnes and A. Elofsson. Prediction of MHC class I binding peptides, using SVMHC. *BMC Bioinformatics*, 3(1):25, Sep 2002.

[463] P. Dönnes and O. Kohlbacher. Integrated modeling of the major events in the MHC class I antigen processing pathway. *Protein Sci.*, 14:2132–2140, Jun 2005.

[464] Pierre Dönnes and Arne Elofsson. Prediction of MHC class I binding peptides, using SVMHC. *BMC Bioinformatics*, 3:25, Sep 2002.

[465] D. L. Donoho and C. Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proc. Natl. Acad. Sci. USA*, 100(10):5591–5596, 2003.

[466] David L. Donoho. De-noising by soft-thresholding. *IEEE Trans. IT*, 41(3):613–627, 1994.

[467] Denise L Doolan, Scott Southwood, Daniel A Freilich, John Sidney, Norma L Graber, Lori Shatney, Lolita Bebris, Laurence Florens, Carlota Dobano, Adam A Witney, Ettore Appella, Stephen L Hoffman, John R Yates, Daniel J Carucci, and Alessandro Sette. Identification of Plasmodium falciparum antigens by antigenic analysis of genomic and proteomic data. *Proc Natl Acad Sci U S A*, 100(17):9952–9957, Aug 2003.

[468] S. Doubet and P. Albersheim. CarbBank. *Glycobiology*, 2(6):505, Dec 1992.

[469] S. Doubet, K. Bock, D. Smith, A. Darvill, and P. Albersheim. The Complex Carbohydrate Structure Database. *Trends Biochem Sci*, 14(12):475–7, Dec 1989.

[470] Irini A Doytchinova, Pingping Guan, and Darren R Flower. Identifying human MHC supertypes using bioinformatic methods. *J Immunol*, 172(7):4314–4323, Apr 2004.

[471] Irini A Doytchinova, Valerie Walshe, Persephone Borrow, and Darren R Flower. Towards the chemometric dissection of peptide–hla-a*0201 binding affinity: comparison of local and global qsar models. *J Comput Aided Mol Des*, 19(3):203–212, Mar 2005.

[472] S. Dreiseitl, L. Ohno-Machado, H. Kittler, S. Vinterbo, H. Billhardt, and M. Binder. A comparison of machine learning methods for the diagnosis of pigmented skin lesions. *J Biomed Inform*, 34(1):28–36, Feb 2001.

[473] J. Drews. Drug Discovery: A Historical Perspective. *Science*, 287:1960–1964, March 2000.

[474] G. Dror, R. Sorek, and R. Shamir. Accurate identification of alternatively spliced exons using support vector machine. *Bioinformatics*, 21(7):897–901, Apr 2005.

[475] Anshul Dubey, Matthew J Realff, Jay H Lee, and Andreas S Bommarius. Support vector machines for learning to identify the critical positions of a protein. *J Theor Biol*, 234(3):351–61, Jun 2005.

[476] E. Dubus, I. Ijjaali, F. Petitet, and A. Michel. In Silico Classification of hERG Channel Blockers: a Knowledge-Based Strategy. *Chem. Med. Chem.*, 1(6):622–630, Jun 2006.

[477] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. Efficient projections onto the l1-ball for learning in high dimensions. In Andrew McCallum and Sam Roweis, editors, *Proceedings of the 25th Annual International Conference on Machine Learning (ICML 2008)*, pages 272–279. Omnipress, 2008.

[478] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2001.

[479] D. Dunson, Y. Xue, and L. Carin. The matrix stick-breaking process: Flexible bayes meta-analysis. *Journal of the American Statistical Association*, 103(481):317–327, March 2008.

[480] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998.

[481] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Ann. Stat.*, 32(2):407–499, 2004.

[482] W. J. Egan, K. M. Merz, and J. J. Baldwin. Prediction of drug absorption using multivariate statistics. *J. Med. Chem.*, 43(21):3867–3877, Oct 2000.

[483] L. M. Egolf and P. C. Jurs. Prediction of Boiling Points of Organic Hterocyclic Compounds Using Regression and Neural Networks Techniques. *J Chem Inf Comput Sci*, 33:616–635, 1993.

[484] Justis P Ehlers and J. William Harbour. NBS1 expression as a prognostic marker in uveal melanoma. *Clin. Cancer Res.*, 11(5):1849–53, Mar 2005.

[485] J. H. J. Einmahl and D. M. Mason. Generalized Quantile Process. *Ann. Stat.*, 20:1062–1078, June 1992.

[486] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA*, 95:14863–14868, Dec 1998.

[487] S. Ekins. In silico approaches to predicting drug metabolism, toxicology and beyond. *Biochem. Soc. Trans.*, 31(Pt 3):611–614, Jun 2003.

[488] S. Ekins, B. Boulanger, P. W. Swaan, and M. A. Z. Hupcey. Towards a new age of virtual ADME/TOX and multidimensional drug discovery. *J Comput Aided Mol Des*, 16(5-6):381–401, 2002.

[489] S. Ekins, W. J. Crumb, R. D. Sarazan, J. H. Wikel, and S. A. Wrighton. Three-dimensional quantitative structure-activity relationship for inhibition of human ether-a-go-go-related gene potassium channel. *J. Pharmacol. Exp. Ther.*, 301(2):427–434, May 2002.

[490] I. El-Naqa, Y. Yang, N. P. Galatsanos, R. M. Nishikawa, and M. N. Wernick. A similarity learning approach to content-based image retrieval: application to digital mammography. *IEEE Trans Med Imaging*, 23(10):1233–44, Oct 2004.

[491] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa. A support vector machine approach for detection of microcalcifications. *IEEE Trans Med Imaging*, 21(12):1552–63, Dec 2002.

[492] S. M. Elbashir, J. Harborth, W. Lendeckel, A. Yalcin, K. Weber, and T. Tuschl. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, 411(6836):494–498, May 2001.

[493] S. M. Elbashir, W. Lendeckel, and T. Tuschl. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.*, 15(2):188–200, Jan 2001.

[494] I O Ellis, M Galea, N Broughton, A Locker, R W Blamey, and C W Elston. Pathological prognostic factors in breast cancer. II. Histological type. Relationship with survival in a large study with long-term follow-up. *Histopathology*, 20:479–89, 1992.

[495] C W Elston and I O Ellis. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, 19:403–10, 1991.

[496] John Emsley. *The Elements (third edition).* Oxford University Press, 1998.

[497] B. E. Engelhardt, M. I. Jordan, K. E. Muratore, and S. E. Brenner. Protein Molecular Function Prediction by Bayesian Phylogenomics. *PLoS Comput. Biol.*, 1(5):e45, Oct 2005.

[498] D. Erhan, P.-J. L'heureux, S. Y. Yue, and Y. Bengio. Collaborative filtering on a family of biological targets. *J. Chem. Inf. Model.*, 46(2):626–635, 2006.

[499] D. Erös, G. Kéri, I. Kövesdi, C. Szántai-Kis, G. Mészáros, and L. Orfi. Comparison of predictive ability of water solubility QSPR models generated by MLR, PLS and ANN methods. *Mini Rev Med Chem*, 4(2):167–177, Feb 2004.

[500] E. Eskin, W.N. Grundy, and Y. Singer. Protein family classification using sparse Markov transducers. In *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology (ISMB 2000)*, pages 134–145, 2000.

[501] A. Esquela-Kerscher and F. J. Slack. Oncomirs - microRNAs with a role in cancer. *Nat. Rev. Cancer*, 6(4):259–269, Apr 2006.

[502] M. Esteller. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.*, 8(4):286–298, Apr 2007.

[503] M. Esteller. Epigenetics in cancer. *N. Engl. J. Med.*, 358(11):1148–1159, Mar 2008.

[504] A. Evers and T. Klabunde. Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor. *J. Med. Chem.*, 48(4):1088–1097, Feb 2005.

[505] T. Evgeniou, C. Micchelli, and M. Pontil. Learning multiple tasks with kernel methods. *J. Mach. Learn. Res.*, 6:615–637, 2005.

[506] T. Evgeniou, M. Pontil, and T. Poggio. Regularization Networks and Support Vector Machines. *Adv. Comput. Math.*, 13(1):1–50, 2000.

[507] Theodoros Evgeniou and Massimiliano Pontil. Regularized multi–task learning. In *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 109–117, New York, NY, USA, 2004. ACM.

[508] M. Fabbri, C. M. Croce, and G. A. Calin. MicroRNAs. *Cancer J.*, 14(1):1–6, 2008.

[509] Theres Fagerberg, Jean-Charles Cerottini, and Olivier Michielin. Structural prediction of peptides bound to MHC class I. *J Mol Biol*, 356(2):521–546, Feb 2006.

[510] J. J. Faith, B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J. J. Collins, and T. S. Gardner. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biol.*, 5(1):e8, Jan 2007.

[511] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *Comput. Comm. Rev.*, 29:251–262, 1999.

[512] A. Farago and G. Lugosi. Strong universal consistency of neural network classifiers. *IEEE Trans. Inform. Theory*, 39(4):1146–1151, Jul 1993.

[513] R. Farid, T. Day, R. A. Friesner, and R. A. Pearlstein. New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorg. Med. Chem.*, 14(9):3160–3173, May 2006.

[514] Olivier Faugeras, Geoffray Adde, Guillaume Charpiat, Christophe Chefd'hotel, Maureen Clerc, Thomas Deneux, Rachid Deriche, Gerardo Hermosillo, Renaud Keriven, Pierre Kornprobst, Jan Kybic, Christophe Lenglet, Lucero Lopez-Perez, Tho Papadopoulo, Jean-Philippe Pons, Florent Segonne, Bertrand Thirion, David Tschumperl, Thierry Viville, and Nicolas Wotawa. Variational, geometric, and statistical methods for modeling brain anatomy and function. *Neuroimage*, 23 Suppl 1:S46–55, 2004.

[515] T. Fawcett. ROC graphs: notes and practical considerations for data mining researchers. Technical Report 2003-4, HP Laboratories, Palo Alto, CA, USA, 2003.

[516] M. Fazel, H. Hindi, and S. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the 2001 American Control Conference*, volume 6, pages 4734–4739, 2001.

[517] M. Feder. Maximum entropy as a special case of the minimum description length criterion. *IEEE Trans. Inform. Theory*, 32(6):847 – 849, Nov 1986.

[518] M. Feder and N. Merhav. Relations between entropy and error probability. *IEEE Trans. Inform. Theory*, 40(1):259 – 266, Jan 1994.

[519] M. Feder and N. Merhav. Hierarchical universal coding. *IEEE Trans. Inform. Theory*, 42(5):1354–1364, Sep 1996.

[520] M. Feder, N. Merhav, and M. Gutman. Universal prediction of individual sequences. *IEEE Trans. Inform. Theory*, 38(4):1258–1270, Jul 1992.

[521] M. Feder and A.C. Singer. Universal data compression and linear prediction. *Data Compression Conference*, 1998.

[522] J. Felsenstein. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, 17:368–376, 1981.

[523] Kai-Yan Feng, Yu-Dong Cai, and Kuo-Chen Chou. Boosting classifier for predicting protein domain structural class. *Biochem Biophys Res Commun*, 334(1):213–7, Aug 2005.

[524] T. L. Ferea, D. Botstein, P. O. Brown, and R. F. Rosenzweig. Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc. Natl. Acad. Sci. USA*, 96(17):9721–9726, 1999.

[525] Miguel A Ferrer, Jess B Alonso, and Carlos M Travieso. Offline geometric parameters for automatic signature verification using fixed-point arithmetic. *IEEE Trans Pattern Anal Mach Intell*, 27(6):993–7, Jun 2005.

[526] S. Fields, Y. Kohara, and D. J. Lockhart. Functional genomics. *Proc. Natl. Acad. Sci. USA*, 96:8825–8826, August 1999.

[527] S. Fields and O. Song. A novel genetic system to detect protein-protein interactions. *Nature*, 340(6230):245–246, Jul 1989.

[528] A. Filatov, A. Gitis, and I. Kil. Graph-based handwritten digit string recognition. In *ICDAR '95: Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 2)*, page 845, Washington, DC, USA, 1995. IEEE Computer Society.

[529] W. Filipowicz, S. N. Bhattacharyya, and N. Sonenberg. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.*, 9(2):102–114, Feb 2008.

[530] S. Fine and K. Scheinberg. Efficient SVM training using low-rank kernel representations. *J. Mach. Learn. Res.*, 2:243–264, 2001.

[531] P. Finn, S. Muggleton, D. Page, and A. Srinivasan. Pharmacophore discovery using the inductive logic programming language Progol. *Machine Learning*, 30:241–270, 1998.

[532] A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, 391(6669):806–811, Feb 1998.

[533] J. Flannick, A. Novak, B.S. Srinivasan, H.H. McAdams, and S. Batzoglou. Graemlin: general and robust alignment of multiple large interaction networks. *Genome Res.*, 16(9):1169–1181, Sep 2006.

[534] A. F. Fliri, W. T. Loging, P. F. Thadeio, and R. A. Volkmann. Analysis of drug-induced effect patterns to link structure and side effects of medicines. *Nat. Chem. Biol.*, 1(7):389–397, Dec 2005.

[535] A. F. Fliri, W. T. Loging, P. F. Thadeio, and R. A. Volkmann. Biological spectra analysis: Linking biological activity profiles to molecular structure. *Proc. Natl. Acad. Sci. USA*, 102(2):261–266, Jan 2005.

[536] Anton F Fliri, William T Loging, Peter F Thadeio, and Robert A Volkmann. Biospectra analysis: model proteome characterizations for linking molecular structure and biological response. *J. Med. Chem.*, 48(22):6918–6925, Nov 2005.

[537] Laurence Florens, Michael P Washburn, J. Dale Raine, Robert M Anthony, Munira Grainger, J. David Haynes, J. Kathleen Moch, Nemone Muster, John B Sacci, David L Tabb, Adam A Witney, Dirk Wolters, Yimin Wu, Malcolm J Gardner, Anthony A Holder, Robert E Sinden, John R Yates, and Daniel J Carucci. A proteomic view of the Plasmodium falciparum life cycle. *Nature*, 419(6906):520–526, Oct 2002.

[538] D. R. Flower. On the properties of bit string-based measures of chemical similarity. *J Chem Inf Comput Sci*, 38:379–386, 1998.

[539] S. P. Fodor, J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu, and D. Solas. Light-directed, spatially addressable parallel chemical synthesis. *Science*, 251:767–773, 1991.

[540] J. H. Fong, A. E. Keating, and M. Singh. Predicting specificity in bZIP coiled-coil protein interactions. *Genome Biol.*, 5(R11), 2004.

[541] G. D. Forney. The viterbi algorithm. *Proceedings of the IEEE*, 61, 1973.

[542] Dean P. Foster and Edward I. George. The risk inflation criterion for multiple regression. *The Annals of Statistics*, 22(4):1947–1975, 1994.

[543] Simon Foucart and Ming-Jun Lai. Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$. *Applied and Computational Harmonic Analysis*, 26(3):395–407, May 2009.

[544] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3:95–110, 1956.

[545] M. J. Fraunholz. Systems biology in malaria research. *Trends Parasitol.*, 21(9):393–395, Sep 2005.

[546] B. B. Fredholm, T. Hökfelt, and G. Milligan. G-protein-coupled receptors: an update. *Acta Physiol.*, 190(1):3–7, May 2007.

[547] Jennifer L. Freeman, George H. Perry, Lars Feuk, Richard Redon, Steven A. McCarroll, David M. Altshuler, Hiroyuki Aburatani, Keith W. Jones, Chris Tyler-Smith, Matthew E. Hurles, Nigel P. Carter, Stephen W. Scherer, and Charles Lee. Copy number variation: new insights in genome diversity. *Genome Research*, 16:949–961, 2006.

[548] S. M. Freier, R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner. Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. USA*, 83(24):9373–7, Dec 1986.

[549] J. Freudenberg and P. Propping. A similarity-based method for genome-wide prediction of disease-relevant human genes. *Bioinformatics*, 18 Suppl 2:S110–S115, 2002.

[550] Y. Freund, Y. Mansour, and R. E. Schapire. Analysis of a Pseudo-Bayesian Prediction Method. In *Conference on Information Sciences and Systems, Princeton University, March 15-17*, 2000.

[551] E. Freyhult, P. Prusis, M. Lapinsh, J. E. S. Wikberg, V. Moulton, and M. G. Gustafsson. Unbiased descriptor and parameter selection confirms the potential of proteochemometric modelling. *BMC Bioinformatics*, 6:50, 2005.

[552] C. C. Friedel, K. H. V. Jahn, S. Sommer, S. Rudd, H. W. Mewes, and I. V. Tetko. Support vector machines for separation of mixed plant-pathogen EST collections based on codon usage. *Bioinformatics*, 21:1383–1388, 2005.

[553] J. Friedman. Some geometric aspects of graphs and their eigenfunctions. *Duke Math. J.*, 69:487–525, March 1993.

[554] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani. Pathwise coordinate optimization. *Ann. Appl. Statist.*, 1(1):302–332, 2007.

[555] J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, Jul 2008.

[556] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, pages kxm045+, December 2007.

[557] N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using Bayesian networks to analyze expression data. *J. Comput. Biol.*, 7(3-4):601–620, 2000.

[558] J. Frigola, J. Song, C. Stirzaker, R. A. Hinshelwood, M. A. Peinado, and S. J. Clark. Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. *Nat. Genet.*, 38(5):540–549, May 2006.

[559] T. M. Frimurer, T. Ulven, C. E. Elling, L.-O. Gerlach, E. Kostenis, and T. Högberg. A physicogenetic method to assign ligand-binding relationships between 7tm receptors. *Bioorg. Med. Chem. Lett.*, 15(16):3707–3712, Aug 2005.

[560] Martin C. Frith, Neil F. W. Saunders, Bostjan Kobe, and Timothy L. Bailey. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol*, 4(5):e1000071+, May 2008.

[561] B. Fritz, F. Schubert, G. Wrobel, C. Schwaenen, S. Wessendorf, M. Nessling, C. Korz, R. J. Rieker, K. Montgomery, R. Kucherlapati, G. Mechtersheimer, R. Eils, S. Joos, and P. Lichter. Microarray-based Copy Number and Expression Profiling in Dedifferentiated and Pleomorphic Liposarcoma. *Cancer Res.*, 62(11):2993–2998, 2002.

[562] H. Fröhlich, J. K. Wegner, F. Sieker, and A. Zell. Optimal assignment kernels for attributed molecular graphs. In *Proceedings of the 22nd international conference on Machine learning*, pages 225 – 232, New York, NY, USA, 2005. ACM Press.

[563] J. C. Fu, S. K. Lee, S. T C Wong, J. Y. Yeh, A. H. Wang, and H. K. Wu. Image segmentation feature selection and pattern classification for mammographic microcalcifications. *Comput Med Imaging Graph*, Jul 2005.

[564] W. Fu. Penalized regressions: the bridge versus the lasso. *Journal of Computational and Graphical Statistics*, 7:397–416, 1998.

[565] Wenjie Fu, Pradipta Ray, and Eric P. Xing. Discover: a feature-based discriminative method for motif search in complex genomes. *Bioinformatics (Oxford, England)*, 25(12):i321–329, June 2009.

[566] W. Fujibuchi, K. Sato, H. Ogata, S. Goto, and M. Kanehisa. KEGG and DBGET/LinkDB: Integration of biological relationships in divergenet molecular biology data. Knowledge Sharing Across Biological and Medical Knowledge-Based Systems WS-98-04, AAAI Press, 1998.

[567] K. Fukumizu, F. R. Bach, and A. Gretton. Statistical consistency of kernel canonical correlation analysis. *J. Mach. Learn. Res.*, 8:361–383, 2008.

[568] Katrin Fundel, Daniel Gttler, Ralf Zimmer, and Joannis Apostolakis. A simple approach for protein name identification: prospects and limits. *BMC Bioinformatics*, 6 Suppl 1:S15, 2005.

[569] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16(10):906–914, Oct 2000.

[570] C. Furlanello, M. Serafini, S. Merler, and G. Jurman. Entropy-based gene ranking without selection bias for the predictive classification of microarray data. *BMC Bioinformatics*, 4(54), 2003.

[571] M. N. Fuster and J. D. Esko. The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat. Rev. Cancer*, 5(7):526–42, Jul 2005.

[572] Roberto Fdez Galn, Silke Sachse, C. Giovanni Galizia, and Andreas V M Herz. Odor-driven attractor dynamics in the antennal lobe allow for simple and rapid olfactory pattern classification. *Neural Comput*, 16(5):999–1012, May 2004.

[573] Rajeev Gangal and Pankaj Sharma. Human pol II promoter prediction: time series descriptors and machine learning. *Nucleic Acids Res*, 33(4):1332–6, 2005.

[574] Juan M Garca-Gmez, Csar Vidal, Luis Mart-Bonmat, Joaqun Galant, Nicolas Sans, Montserrat Robles, and Francisco Casacuberta. Benign/malignant classifier of soft tissue tumors using MR imaging. *MAGMA*, 16(4):194–201, Mar 2004.

[575] Malcolm J Gardner, Neil Hall, Eula Fung, Owen White, Matthew Berriman, Richard W Hyman, Jane M Carlton, Arnab Pain, Karen E Nelson, Sharen Bowman, Ian T Paulsen, Keith James, Jonathan A Eisen, Kim Rutherford, Steven L Salzberg, Alister Craig, Sue Kyes, Man-Suen Chan, Vishvanath Nene, Shamira J Shallom, Bernard Suh, Jeremy Peterson, Sam Angiuoli, Mihaela Pertea, Jonathan Allen, Jeremy Selengut, Daniel Haft, Michael W Mather, Akhil B Vaidya, David M A Martin, Alan H Fairlamb, Martin J Fraunholz, David S Roos, Stuart A Ralph, Geoffrey I McFadden, Leda M Cummings, G. Mani Subramanian, Chris Mungall, J. Craig Venter, Daniel J Carucci, Stephen L Hoffman, Chris Newbold, Ronald W Davis, Claire M Fraser, and Bart Barrell. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, 419(6906):498–511, Oct 2002.

[576] T. S. Gardner, D. Bernardo, D. Lorenz, and J. J. Collins. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, 301(5629):102–105, Jul 2003.

[577] J. L. Gardy, M. R. Laird, F. Chen, S. Rey, C. J. Walsh, M. Ester, and F. S. L. Brinkman. PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics*, 21(5):617–623, Mar 2005.

[578] M. R. Garey and D. S. Johnson. *Computer and intractability: A guide to the theory of NP-completeness.* San Francisco, CA: W. H. Freeman, 1979.

[579] A. Garg, M. Bhasin, and G.P. Raghava. SVM-based method for subcellular localization of human proteins using amino acid compositions, their order and similarity search. *J. Biol. Chem.*, 280(15):14427–32, Apr 2005.

[580] Deon Garrett, David A Peterson, Charles W Anderson, and Michael H Thaut. Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *IEEE Trans Neural Syst Rehabil Eng*, 11(2):141–4, Jun 2003.

[581] T. Gärtner. Exponential and Geometric Kernels for Graphs. In NIPS Workshop on Unreal Data: Principles of Modeling Nonvectorial Data, 2002.

[582] T. Gärtner. A Survey of Kernels for Structured Data. *SIGKDD Explor. Newsl.*, 5(1):49–58, 2003.

[583] T. Gärtner, P. Flach, and S. Wrobel. On graph kernels: hardness results and efficient alternatives. In B. Schölkopf and M. Warmuth, editors, *Proceedings of the Sixteenth Annual Conference on Computational Learning Theory and the Seventh Annual Workshop on Kernel Machines*, volume 2777 of *Lecture Notes in Computer Science*, pages 129–143, Heidelberg, 2003. Springer.

[584] T. Gärtner, P.A. Flach, A. Kowalczyk, and A.J. Smola. Multi-Instance Kernels. In C. Sammut and A. Hoffmann, editors, *Proceedings of the Nineteenth International Conference on Machine Learning*, pages 179–186. Morgan Kaufmann, 2002.

[585] T. Gärtner, P. Glach, and S. Wrobel. On graph kernels: hardness results and efficient alternatives. In *Proceedings of COLT / Kernel workshop*, 2003.

[586] T. Gärtner, J.W. Lloyd, and P.A. Flach. Kernels and distances for structured data. *Mach. Learn.*, 57(3):205–232, 2004.

[587] A. P. Gasch, P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein, and P. O. Brown. Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. *Mol. Biol. Cell*, 11:4241–4257, Dec 2000.

[588] A.P. Gasch, M. Huang, S. Metzner, D. Botstein, S.J. Elledge, and P.O. Brown. Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Mol. Biol. Cell*, 12(10):2987–3003, 2001.

[589] J. Gasteiger and T. Engel, editors. *Chemoinformatics : a Textbook*. Wiley, New York, NY, USA, 2003.

[590] Johann Gasteiger, Jens Sadowski, Jan Schuur, Paul Selzer, Larissa Steinhauer, and Valentin Steinhauer. Chemical information in 3d space. *Journal of Chemical Information and Computer Sciences*, 36(5):1030–1037, 1996.

[591] I. Gat-Viks, A. Tanay, and R. Shamir. Modeling and analysis of heterogeneous regulation in biological networks. *J Comput Biol*, 11(6):1034–49, 2004.

[592] S. Gaudan, H. Kirsch, and D. Rebholz-Schuhmann. Resolving abbreviations to their senses in Medline. *Bioinformatics*, Jul 2005.

[593] Anne-Claude Gavin, Markus Bsche, Roland Krause, Paola Grandi, Martina Marzioch, Andreas Bauer, Jrg Schultz, Jens M Rick, Anne-Marie Michon, Cristina-Maria Cruciat, Marita Remor, Christian Hfert, Malgorzata Schelder, Miro Brajenovic, Heinz Ruffner, Alejandro Merino, Karin Klein, Manuela Hudak, David Dickson, Tatjana Rudi, Volker Gnau, Angela Bauch, Sonja Bastuck, Bettina Huhse, Christina Leutwein, Marie-Anne Heurtier, Richard R Copley, Angela Edelmann, Erich Querfurth, Vladimir Rybin, Gerard Drewes, Manfred Raida, Tewis Bouwmeester, Peer Bork, Bertrand Seraphin, Bernhard Kuster, Gitte Neubauer, and Giulio Superti-Furga. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415(6868):141–7, Jan 2002.

[594] Xijin Ge, Shuichi Tsutsumi, Hiroyuki Aburatani, and Shuichi Iwata. Reducing false positives in molecular pattern recognition. *Genome Inform Ser Workshop Genome Inform*, 14:34–43, 2003.

[595] T. Van Gestel, J. A K Suykens, G. Lanckriet, A. Lambrechts, B. De Moor, and J. Vandewalle. Bayesian framework for least-squares support vector machine classifiers, gaussian processes, and kernel Fisher discriminant analysis. *Neural Comput*, 14(5):1115–47, May 2002.

[596] U. Gether. Uncovering molecular mechanisms involved in activation of g protein-coupled receptors. *Endocr Rev*, 21(1):90–113, Feb 2000.

[597] Debashis Ghosh and Arul M Chinnaiyan. Classification and Selection of Biomarkers in Genomic Data Using LASSO. *J Biomed Biotechnol*, 2005(2):147–54, 2005.

[598] V. Gillet, P. Willett, and J. Bradshaw. Similarity searching using reduced graphs. *J Chem Inf Comput Sci*, 43:338–345, 2003.

[599] Santhosh Girirajan, Lin Chen, Tina Graves, Tomas Marques-Bonet, Mario Ventura, Catrina Fronick, Lucinda Fulton, Mariano Rocchi, Robert S Fulton, Richard K Wilson, Elaine R Mardis, and Evan E Eichler. Sequencing human-gibbon breakpoints of synteny reveals mosaic new insertions at rearrangement sites. *Genome Res.*, 19(2):178–190, Feb 2009.

[600] Girosi. An Equivalence Between Sparse Approximation and Support Vector Machines. *Neural Comput*, 10(6):1455–80, Jul 1998.

[601] F. Girosi, M. Jones, and T. Poggio. Regularization Theory and Neural Networks Architectures. *Neural Comput.*, 7(2):219–269, 1995.

[602] F. Glaser, R. J. Morris, R. J. Najmanovich, R. A. Laskowski, and J. M. Thornton. A method for localizing ligand binding pockets in protein structures. *Proteins*, 62(2):479–488, February 2006.

[603] Dimitris Glotsos, Panagiota Spyridonos, Dionisis Cavouras, Panagiota Ravazoula, Petroula-Arampantoni Dadioti, and George Nikiforidis. Automated segmentation of routinely hematoxylin-eosin-stained microscopic images by combining support vector machine clustering and active contour models. *Anal Quant Cytol Histol*, 26(6):331–40, Dec 2004.

[604] Dimitris Glotsos, Panagiota Spyridonos, Panagiotis Petalas, Dionisis Cavouras, Panagiota Ravazoula, Petroula-Arampatoni Dadioti, Ioanna Lekka, and George Nikiforidis. Computer-based malignancy grading of astrocytomas employing a support vector machine classifier, the WHO grading system and the regular hematoxylin-eosin diagnostic staining procedure. *Anal Quant Cytol Histol*, 26(2):77–83, Apr 2004.

[605] Dimitris Glotsos, Jussi Tohka, Panagiota Ravazoula, Dionisis Cavouras, and George Nikiforidis. Automated diagnosis of brain tumours astrocytomas using probabilistic neural network clustering and support vector machines. *Int J Neural Syst*, 15(1-2):1–11, 2005.

[606] A. Goffeau, B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.D. Hoheisel, C. Jacq, M. Johnston, E.J. Louis, H.W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. Life with 6000 genes. *Science*, 274:546–567, October 1996.

[607] N.D. Gold and R.M. Jackson. Fold independent structural comparisons of protein-ligand binding sites for exploring functional relationships. *J Mol Biol*, 355(5):1112–1124, Feb 2006.

[608] N.D. Gold and R.M. Jackson. Sitesbase: a database for structure-based protein-ligand binding site comparisons. *Nucleic Acids Res*, 34:D231–D234, Jan 2006.

[609] S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(4):377–388, April 1996.

[610] Michael H Goldbaum, Pamela A Sample, Kwokleung Chan, Julia Williams, Te-Won Lee, Eytan Blumenthal, Christopher A Girkin, Linda M Zangwill, Christopher Bowd, Terrence Sejnowski, and Robert N Weinreb. Comparing machine learning classifiers for diagnosing glaucoma from standard automated perimetry. *Invest Ophthalmol Vis Sci*, 43(1):162–9, Jan 2002.

[611] L. Goldfarb, O. Golubitsky, and D. Korkin. What is a structural representation? Technical report, University of New Brunswick, 2000. Technical report TR00-137.

[612] Polina Golland, W. Eric L Grimson, Martha E Shenton, and Ron Kikinis. Detection and analysis of statistical differences in anatomical shape. *Med Image Anal*, 9(1):69–86, Feb 2005.

[613] Gene H. Golub and Charles F. Van Loan. *Matrix computations (3rd ed.)*. Johns Hopkins University Press, Baltimore, MD, USA, 1996.

[614] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science*, 286:531–537, 1999.

[615] S. M. Gomez, W. S. Noble, and A. Rzhetsky. Learning to predict protein-protein interactions from protein sequences. *Bioinformatics*, 19(15):1875–1881, 2003.

[616] D. Gong and J. E. Ferrell. Picking a winner: new mechanistic insights into the design of effective siRNAs. *Trends Biotechnol.*, 22(9):451–4, Sep 2004.

[617] A.C. Good and W.G. Richards. Rapid Evaluation of Molecular Shape Similarity Using Gaussian Functions. *J Chem Inf Comput Sci*, 33:112–116, 1993.

[618] L. Gordon, A. Y. Chervonenkis, A. J. Gammerman, I. A. Shahmuradov, and V. V. Solovyev. Sequence alignment kernel for recognition of promoter regions. *Bioinformatics*, 19(15):1964–1971, 2003.

[619] S. Goto, T. Nishioka, and M. Kanehisa. LIGAND: chemical database for enzyme reactions. *Bioinformatics*, 14:591–599, 1998.

[620] S. Goto, Y. Okuno, M. Hattori, T. Nishioka, and M. Kanehisa. LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.*, 30:402–404, 2002.

[621] Johannes Graumann, Leslie A Dunipace, Jae Hong Seol, W. Hayes McDonald, John R Yates, Barbara J Wold, and Raymond J Deshaies. Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. *Mol Cell Proteomics*, 3(3):226–37, Mar 2004.

[622] Paul E. Green and V. Srinivasan. Conjoint analysis in consumer research: Issues and outlook. *The Journal of Consumer Research*, 5(2):103–123, 1978.

[623] E. Greenshtein and Y. Ritov. Persistence in high-dimensional linear predictor selection and the virtue of overparametrization. *Bernoulli*, 10(6):971–988, 2004.

[624] M. Gribskov, R. Lüthy, , and D. Eisenberg. Profile Analysis. *Methods in Enzymology*, 183:146–159, 1990.

[625] M. Gribskov and N. L. Robinson. Use of receiver operating characteristic (ROC) analysis to evaluate sequence matching. *Comput. Chem.*, 20(1):25–33, 1996.

[626] D. R. Groebe and O. C. Uhlenbeck. Characterization of RNA hairpin loop stability. *Nucleic Acids Res.*, 16(24):11725–35, Dec 1988.

[627] C. Gross, M. Kelleher, V.R. Iyer, P.O. Brown, and D.R. Winge. Identification of the copper regulon in Saccharomyces cerevisiae by DNA microarrays. *J. Biol. Chem.*, 275(41):32310–32316, 2000.

[628] W. N. Grundy. Family-based Homology Detection via Pairwise Sequence Comparison. In *Proceedings of the Second Annual International Conference on Computational Molecular Biology, March 22-25*, pages 94–100, 1998.

[629] W. Guba. Chemogenomics strategies for g-protein coupled receptor hit finding. *Ernst Schering Res Found Workshop*, 58:21–29, 2006.

[630] N. Guelzim, S. Bottani, P. Bourgine, and F. Képès. Topological and causal structure of the yeast transcriptional regulatory network. *Nat. Genet.*, 31:60–63, 2002.

[631] Y. Guermeur. Combining Discriminant Models with New Multi-Class SVMs. *Pattern Anal. Appl.*, 5(2):168–179, 2002.

[632] Y. Guermeur, A. Lifschitz, and R. Vert. A kernel for protein secondary structure prediction. In B. Schölkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 193–206. MIT Press, 2004.

[633] Y. Guermeur, G. Pollastri, A. Elisseeff, D. Zelus, H. Paugam-Moisy, and P. Baldi. Combining protein secondary structure prediction models with ensemble methods of optimal complexity. *Neurocomputing*, 56:305–327, 2004.

[634] K. Gulukota, J. Sidney, A. Sette, and C. DeLisi. Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *J Mol Biol*, 267(5):1258–1267, Apr 1997.

[635] O. F. Güner. *Pharmacophore Perception, Development, and Use in Drug Design*, volume 2 of *IUL Biotechnology Series*. International University Line, 2000.

[636] Guodong Guo and Charles R Dyer. Learning from examples in the small sample case: face expression recognition. *IEEE Trans Syst Man Cybern B Cybern*, 35(3):477–88, Jun 2005.

[637] Hong Guo, Lindsay B Jack, and Asoke K Nandi. Feature generation using genetic programming with application to fault classification. *IEEE Trans Syst Man Cybern B Cybern*, 35(1):89–99, Feb 2005.

[638] J. Guo, H. Chen, Z. Sun, and Y. Lin. A novel method for protein secondary structure prediction using dual-layer SVM and profiles. *Proteins*, 54(4):738–743, 2004.

[639] Ting Guo, Yanxin Shi, and Zhirong Sun. A novel statistical ligand-binding site predictor: application to ATP-binding sites. *Protein Eng Des Sel*, 18(2):65–70, Feb 2005.

[640] T. Gururaja, W. Li, W.S. Noble, D.G. Payan, and D.C. Anderson. Multiple functional categories of proteins identified in an in vitro cellular ubiquitin affinity extract using shotgun peptide sequencing. *J Proteome Res*, 2(394-404):394–404, 2003.

[641] B. Gusterson. Do 'basal-like' breast cancers really exist? *Nat. Rev. Cancer*, 9(2):128–134, Feb 2009.

[642] B. A. Gusterson, D. T. Ross, V. J. Heath, and T. Stein. Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. *Breast Cancer Res.*, 7(4):143–148, 2005.

[643] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, editors. *Feature Extraction, Foundations and Applications.* Springer, 2006.

[644] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Mach. Learn.*, 46(1/3):389–422, Jan 2002.

[645] L. Gyorfi, G. Lugosi, and G. Morvai. A simple randomized algorithm for sequential prediction of ergodic time series. *IEEE Trans. Inform. Theory*, 47(5):2642 – 2650, Nov 1999.

[646] L. Gyorfi, I. Pali, and E.C. Van der Meulen. There is no universal source code for an infinite source alphabet. *IEEE Trans. Inform. Theory*, 40(1):267–271, Jan 1994.

[647] S. B. Gktrk, C. Tomasi, B. Acar, C. F. Beaulieu, D. S. Paik, R. B. Jeffrey, J. Yee, and S. Napel. A statistical 3-D pattern processing method for computer-aided detection of polyps in CT colonography. *IEEE Trans Med Imaging*, 20(12):1251–60, Dec 2001.

[648] Bernard Haasdonk. Feature space interpretation of SVMs with indefinite kernels. *IEEE Trans Pattern Anal Mach Intell*, 27(4):482–92, Apr 2005.

[649] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, 13:49–52, 1902.

[650] J. Hadamard. *Lectures on Cauchy's Problem: In Linear Partial Differential Equations.* Dover Publications, 1923.

[651] Torsten Haferlach, Alexander Kohlmann, Susanne Schnittger, Martin Dugas, Wolfgang Hiddemann, Wolfgang Kern, and Claudia Schoch. AML M3 and AML M3 variant each have a distinct gene expression signature but also share patterns different from other genetically defined AML subtypes. *Genes Chromosomes Cancer*, 43(2):113–27, Jun 2005.

[652] Torsten Haferlach, Alexander Kohlmann, Susanne Schnittger, Martin Dugas, Wolfgang Hiddemann, Wolfgang Kern, and Claudia Schoch. A global approach to the diagnosis of leukemia using gene expression profiling. *Blood*, 106(4):1189–1198, Aug 2005.

[653] B. Haibe-Kains, C. Desmedt, F. Piette, M. Buyse, F. Cardoso, L. Van't Veer, M. Piccart, G. Bontempi, and C. Sotiriou. Comparison of prognostic gene expression signatures for breast cancer. *BMC Genomics*, 9:394, 2008.

[654] James A Haigh, Barry T Pickup, J. Andrew Grant, and Anthony Nicholls. Small molecule shape-fingerprints. *J Chem Inf Model*, 45(3):673–684, 2005.

[655] J. Hakenberg, S. Schmeier, A. Kowald, E. Klipp, and U. Leser. Finding kinetic parameters using text mining. *OMICS*, 8(2):131–152, 2004.

[656] Jrg Hakenberg, Steffen Bickel, Conrad Plake, Ulf Brefeld, Hagen Zahn, Lukas Faulstich, Ulf Leser, and Tobias Scheffer. Systematic feature evaluation for gene name recognition. *BMC Bioinformatics*, 6 Suppl 1:S9, 2005.

[657] B. Haley and P. D. Zamore. Kinetic analysis of the RNAi enzyme complex. *Nat. Struct. Mol. Biol.*, 11(7):599–606, Jul 2004.

[658] J. Hall. Unravelling the general properties of siRNAs: strength in numbers and lessons from the past. *Nat. Rev. Genet.*, 5(7):552–7, Jul 2004.

[659] Neil Hall, Marianna Karras, J. Dale Raine, Jane M Carlton, Taco W A Kooij, Matthew Berriman, Laurence Florens, Christoph S Janssen, Arnab Pain, Georges K Christophides, Keith James, Kim Rutherford, Barbara Harris, David Harris, Carol Churcher, Michael A Quail, Doug Ormond, Jon Doggett, Holly E Trueman, Jacqui Mendoza, Shelby L Bidwell, Marie-Adele Rajandream, Daniel J Carucci, John R Yates, Fotis C Kafatos, Chris J Janse, Bart Barrell, C. Michael R Turner, Andrew P Waters, and Robert E Sinden. A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. *Science*, 307(5706):82–86, Jan 2005.

[660] I. Halperin, B. Ma, H. Wolfson, and R. Nussinov. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, 47(4):409–443, Jun 2002.

[661] Peter Hammond, Tim J Hutton, Judith E Allanson, Linda E Campbell, Raoul C M Hennekam, Sean Holden, Michael A Patton, Adam Shaw, I. Karen Temple, Matthew Trotter, Kieran C Murphy, and Robin M Winter. 3D analysis of facial morphology. *Am J Med Genet A*, 126(4):339–48, May 2004.

[662] L.Y. Han, C.Z. Cai, Z.L. Ji, Z.W. Cao, J. Cui, and Y.Z. Chen. Predicting functional family of novel enzymes irrespective of sequence similarity: a statistical learning approach. *Nucl. Acids Res.*, 32(21):6437–6444, 2004.

[663] L.Y. Han, C.Z. Cai, Z.L. Ji, and Y.Z. Chen. Prediction of functional class of novel viral proteins by a statistical learning method irrespective of sequence similarity. *Virology*, 331(1):136–143, 2005.

[664] L.Y. Han, C.Z. Cai, S.L. Lo, M.C. Chung, and Y.Z. Chen. Prediction of RNA-binding proteins from primary sequence by a support vector machine approach. *RNA*, 10(3):355–368, 2004.

[665] Sangjo Han, Byung-Chul Lee, Seung Taek Yu, Chan-Seok Jeong, Soyoung Lee, and Dongsup Kim. Fold recognition by combining profile-profile alignment and support vector machine. *Bioinformatics*, 21(11):2667–73, Jun 2005.

[666] D. Hanahan and R. A. Weinberg. The hallmarks of cancer. *Cell*, 100:57–70, 2000.

[667] S. Hanash. Integrated global profiling of cancer. *Nat. Rev. Cancer*, 4(8):638–644, 2004.

[668] D. Hanisch, A. Zien, R. Zimmer, and T. Lengauer. Co-clustering of biological networks and gene expression data. *Bioinformatics*, 2002.

[669] M. Hann and R. Green. Chemoinformatics–a new name for an old problem? *Curr Opin Chem Biol*, 3(4):379–383, Aug 1999.

[670] G. J. Hannon and J. J. Rossi. Unlocking the potential of the human genome with RNA interference. *Nature*, 431(7006):371–8, Sep 2004.

[671] C. Hansch and T. Fujita. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc*, 86:1616–1626, 1964.

[672] C. Hansch, J. E. Quinlan, and G. L. Lawrence. Linear free-energy relationship between partition coefficients and the aqueous solubility of organic liquids. *J. Org. Chem.*, 33:347 – 350, 1968.

[673] J. Harborth, S. M. Elbashir, K. Vandenburgh, H. Manninga, S. A. Scaringe, K. Weber, and T. Tuschl. Sequence, chemical, and structural variation of small interfering RNAs and short hairpin RNAs and the effect on mammalian gene silencing. *Antisense Nucleic Acid. Drug. Dev.*, 13(2):83–105, Apr 2003.

[674] Z. Harchaoui. *Méthodes à noyaux pour la détection.* PhD thesis, Telecom ParisTech, 2008.

[675] Z. Harchaoui and F. Bach. Image classification with segmentation graph kernels. In *2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007)*, pages 1–8. IEEE Computer Society, 2007.

[676] Zaid Harchaoui and Céline Levy-Leduc. Catching change-points with lasso. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 617–624. MIT Press, Cambridge, MA, 2008.

[677] A.J. Hartemink, D.K. Gifford, T.S. Jaakkola, and R.A. Young. Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. In Russ B. Altman, A. Keith Dunker, Lawrence Hunter, Kevin Lauerdale, and Teri E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing 2002*, pages 422–433. World Scientific, 2002.

[678] J. A. Hartigan. Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.*, 82(397):267–270, 1987.

[679] J. A. Hartigan and M. A. Wong. A K-means clustering algorithm. *Applied Statistics*, 28:100–108, 1979.

[680] T. Hastie and R. Tibshirani. *Generalized Additive Models.* Chapman and Hall, London, UK, 1999.

[681] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference, and prediction.* Springer, 2001.

[682] D. Haussler. A general minimax result for relative entropy. *IEEE Trans. Inform. Theory*, 43(4):1276–1280, Jul 1997.

[683] D. Haussler. Convolution Kernels on Discrete Structures. Technical Report UCSC-CRL-99-10, UC Santa Cruz, 1999.

[684] W. Haverkamp, G. Breithardt, A. J. Camm, M. J. Janse, M. R. Rosen, C. Antzelevitch, D. Escande, M. Franz, M. Malik, A. Moss, and R. Shah. The potential for QT prolongation and proarrhythmia by non-antiarrhythmic drugs: clinical and regulatory implications. Report on a policy conference of the European Society of Cardiology. *Eur. Heart J.*, 21(15):1216–1231, Aug 2000.

[685] D.M. Hawkins, S.S. Young, and A. Rusinko. Analysis of a large structure-activity data set using recursive partitioning. *Quantitative Structure-Activity Relationships*, 16:296–302, 1997.

[686] D. Heckerman. A tutorial on learning with Bayesian networks. In M. Jordan, editor, *Learning in graphical models*, pages 301–354. MIT Press, Cambridge, MA, USA, 1999.

[687] D. Heckerman, D. M. Chickering, C. Meek, R. Rounthwaite, and C. Kadie. Dependency networks for inference, collaborative filtering, and data visualization. *J. Mach. Learn. Res.*, 1:49–75, 2000.

[688] D. Heckerman, D. Kadie, and J. Listgarten. Leveraging information across HLA alleles/supertypes improves epitope prediction. *J. Comput. Biol.*, 14(6):736–746, 2007.

[689] C. Helma, T. Cramer, S. Kramer, and L. De Raedt. Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds. *J. Chem. Inf. Comput. Sci.*, 44(4):1402–11, 2004.

[690] D. P. Helmbold and R. E. Schapire. Predicting Nearly As Well As the Best Pruning of a Decision Tree. *Machine Learning*, 27(1):51–68, 1997.

[691] Martin Hendrix and Christopher Kallus. Phosphodiesterase inhibitors: A chemogenomic view. In *Chemogenomics in Drug Discovery*, chapter 9, pages 243–288. Wiley-VCH, 2005.

[692] S. Henikoff and J. G. Henikoff. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA*, 89(22):10915–10919, Nov 1992.

[693] Y. Hershkovits and J. Ziv. On fixed-database universal data compression with limited memory. *IEEE Trans. Inform. Theory*, 43(6):1966–1976, Nov 1997.

[694] T. Hertz and C. Yanover. PepDist: a new framework for protein-peptide binding prediction based on learning peptide distance functions. *BMC Bioinformatics*, 7 Suppl 1:S3, 2006.

[695] Tomer Hertz and Chen Yanover. Identifying hla supertypes by learning distance functions. *Bioinformatics*, 23(2):e148–e155, Jan 2007.

[696] Tom Heskes. Empirical bayes for learning to learn. In *ICML '00: Proceedings of the Seventeenth International Conference on Machine Learning*, pages 367–374, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.

[697] S. J. Hill. G-protein-coupled receptors: past, present and future. *Br. J. Pharmacol.*, 147 Suppl 1:S27–S37, Jan 2006.

[698] O. Hindsgaul. Carbohydrate chemistry. Sugars out in the open. *Nature*, 399(6737):644–5, Jun 1999.

[699] Y. Hizukuri, Y. Yamanishi, O. Nakamura, F. Yagi, S. Goto, and M. Kanehisa. Extraction of leukemia specific glycan motifs in humans by computational glycomics. *Carbohydr. Res.*, 340(14):2270–8, Oct 2005.

[700] Yoshiyuki Hizukuri, Yoshihiro Yamanishi, Kosuke Hashimoto, and Minoru Kanehisa. Extraction of species-specific glycan substructures. *Genome Inform Ser Workshop Genome Inform*, 15(1):69–81, 2004.

[701] Yuen Ho, Albrecht Gruhler, Adrian Heilbut, Gary D Bader, Lynda Moore, Sally-Lin Adams, Anna Millar, Paul Taylor, Keiryn Bennett, Kelly Boutilier, Lingyun Yang, Cheryl Wolting, Ian Donaldson, Sren Schandorff, Juanita Shewnarane, Mai Vo, Joanne Taggart, Marilyn Goudreault, Brenda Muskat, Cris Alfarano, Danielle Dewar, Zhen Lin, Katerina Michalickova, Andrew R Willems, Holly Sassi, Peter A Nielsen, Karina J Rasmussen, Jens R Andersen, Lene E Johansen, Lykke H Hansen, Hans Jespersen, Alexandre Podtelejnikov, Eva Nielsen, Janne Crawford, Vibeke Poulsen, Birgitte D Srensen, Jesper Matthiesen, Ronald C Hendrickson, Frank Gleeson, Tony Pawson, Michael F Moran, Daniel Durocher, Matthias Mann, Christopher W V Hogue, Daniel Figeys, and Mike Tyers. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature*, 415(6868):180–3, Jan 2002.

[702] S. Hochreiter and K. Obermayer. Gene selection for microarray data. In B. Schlkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 319–355. MIT Press, 2004.

[703] Neil G Hockstein, Erica R Thaler, Drew Torigian, Wallace T Miller, Olivia Deffenderfer, and C. William Hanson. Diagnosis of pneumonia with an electronic nose: correlation of vapor signature with chest computed tomography scan findings. *Laryngoscope*, 114(10):1701–5, Oct 2004.

[704] A. E. Hoerl. Application of ridge regression analysis to regression problems. *Chemical Engineering Progress*, 58:54–59, 1962.

[705] A. E. Hoerl and R. W. Kennard. Ridge regression : biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[706] A. E. Hoerl and R. W. Kennard. Citation classic - ridge regression : biased estimation for nonorthogonal problems. *CC/Eng. Tech. Appl. Sci.*, 35:18–18, 1982.

[707] Oliver Hofmann and Dietmar Schomburg. Concept-based annotation of enzyme classes. *Bioinformatics*, 21(9):2059–66, May 2005.

[708] Jrg D. Hoheisel. Microarray technology: beyond transcript profiling and genotype analysis. *Nature Reviews Genetics*, 7:200–210, 2006.

[709] T. Holen, M. Amarzguioui, M. T. Wiiger, E. Babaie, and H. Prydz. Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor. *Nucleic Acids Res.*, 30(8):1757–1766, Apr 2002.

[710] J. D. Holliday and P. Willett. Using a genetic algorithm to identify common structural features in sets of ligands. *J. Mol. Graph. Model.*, 15(4):221–232, Aug 1997.

[711] M. C. Honeyman, V. Brusic, N. L. Stone, and L. C. Harrison. Neural network-based prediction of candidate T-cell epitopes. *Nat. Biotechnol.*, 16(10):966–969, Oct 1998.

[712] A. L. Hopkins and C. R. Groom. The druggable genome. *Nat. Rev. Drug Discov.*, 1(9):727–730, Sep 2002.

[713] Fereydoun Hormozdiari, Can Alkan, Evan E Eichler, and S. Cenk Sahinalp. Combinatorial algorithms for structural variation detection in high-throughput sequenced genomes. *Genome Res*, 19(7):1270–1278, Jul 2009.

[714] David Horn, Gideon Dror, and Brigitte Quenet. Dynamic proximity of spatio-temporal sequences. *IEEE Trans Neural Netw*, 15(5):1002–8, Sep 2004.

[715] F. Horn, E. Bettler, L. Oliveira, F. Campagne, F. E. Cohen, and G. Vriend. GPCRDB information system for G protein-coupled receptors. *Nucl. Acids Res.*, 31(1):294–297, 2003.

[716] J. J. Hornberg, F. J. Bruggeman, H. V. Westerhoff, and J. Lankelma. Cancer: a systems biology disease. *Biosystems*, 83(2-3):81–90, 2006.

[717] D. Horvath and C. Jeandenans. Neighborhood behavior of in silico structural spaces with respect to in vitro activity spaces–a benchmark for neighborhood behavior assessment of different in silico similarity metrics. *J. Chem. Inf. Comput. Sci.*, 43(2):691–698, 2003.

[718] T. Horváth, T. Gärtner, and S. Wrobel. Cyclic pattern kernels for predictive graph mining. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 158–167, New York, NY, USA, 2004. ACM Press.

[719] H. Hotelling. Relation between two sets of variates. *Biometrika*, 28:322–377, 1936.

[720] Y. Hou, W. Hsu, M. L. Lee, and C. Bystroff. Efficient remote homology detection using local structure. *Bioinformatics*, 19(17):2294–2301, 2003.

[721] Y. Hou, W. Hsu, M. L. Lee, and C. Bystroff. Remote homolog detection using local sequence-structure correlations. *Proteins*, 57(3):518–530, 2004.

[722] A. C. Hsieh, R. Bo, J. Manola, F. Vazquez, O. Bare, A. Khvorova, S. Scaringe, and W. R. Sellers. A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. *Nucleic Acids Res.*, 32(3):893–901, 2004.

[723] C. Hu, X. Li, and J. Liang. Developing optimal non-linear scoring function for protein design. *Bioinformatics*, 20(17):3080–3098, 2004.

[724] H.J. Hu, Y. Pan, R. Harrison, and P.C. Tai. Improved protein secondary structure prediction using support vector machine with a new encoding scheme and an advanced tertiary classifier. *IEEE Trans. Nanobioscience*, 3(4):265–271, 2004.

[725] J. Hua, Z. Xiong, J. Lowey, E. Suh, and E. R. Dougherty. Optimal number of features as a function of sample size for various classification rules. *Bioinformatics*, 21(8):1509–1515, Apr 2005. To appear.

[726] S. Hua and Z. Sun. A Novel Method of Protein Secondary Structure Prediction with High Segment Overlap Measure: Support Vector Machine Approach. *J. Mol. Biol.*, 308(2):397–407, April 2001.

[727] S. Hua and Z. Sun. Support vector machine approach for protein subcellular localization prediction. *Bioinformatics*, 17(8):721–728, 2001.

[728] Si hua Peng, Long jiang Fan, Xiao ning Peng, Shu lin Zhuang, Wei Du, and Liang biao Chen. Splicing-site recognition of rice (Oryza sativa L.) DNA sequences by support vector machines. *J Zhejiang Univ Sci*, 4(5):573–7, 2003.

[729] Lian hua Zhang, Guan hua Zhang, Jie Zhang, and Ying cai Bai. Intrusion detection using rough set classification. *J Zhejiang Univ Sci*, 5(9):1076–86, Sep 2004.

[730] J. Huan, W. Wang, A. Washington, J. Prins, R. Shah, and A. Tropsha. Accurate classification of protein structural families using coherent subgraph analysis. In *Proceedings of the Pacific Symposium on Biocomputing 2002*, pages 411–422, 2004.

[731] Bingding Huang and Michael Schroeder. Ligsitecsc: predicting ligand binding sites using the connolly surface and degree of conservation. *BMC Struct Biol*, 6:19, 2006.

[732] E. Huang, S. Ishida, J. Pittman, H. Dressman, A. Bild, M. Kloos, M. D'Amico, R. G. Pestell, M. West, and J. R. Nevins. Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nat Genet*, 34(2):226–30, 2003.

[733] Jing Huang and Feng Shi. Support vector machines for predicting apoptosis proteins types. *Acta Biotheor.*, 53(1):39–47, 2005.

[734] Junzhou Huang and Tong Zhang. The benefit of group sparsity, 2009.

[735] K. Huang and R.F. Murphy. Boosting accuracy of automated classification of fluorescence microscope images for location proteomics. *BMC Bioinformatics*, 5(78):78, 2004.

[736] N. Huang, H. Chen, and Z. Sun. CTKPred: an SVM-based method for the prediction and classification of the cytokine superfamily. *Protein Eng. Des. Sel.*, Jun 2005.

[737] Shao-Wei Huang and Jenn-Kang Hwang. Computation of conformational entropy from protein sequences using the machine-learning method–application to the study of the relationship between structural conservation and local structural stability. *Proteins*, 59(4):802–9, Jun 2005.

[738] T. M. Huang and V. Kecman. Gene extraction for cancer diagnosis by support vector machines-An improvement. *Artif. Intell. Med.*, Jul 2005.

[739] Yu-Len Huang and Dar-Ren Chen. Support vector machines in sonography: application to decision making in the diagnosis of breast cancer. *Clin Imaging*, 29(3):179–84, 2005.

[740] C.A. Hudis. Trastuzumab–mechanism of action and use in clinical practice. *N. Engl. J. Med.*, 357(1):39–51, Jul 2007.

[741] M. Hue. Semi-supervised learning for protein structure prediction. Master's thesis, Ecole des Mines de Paris, 2004.

[742] D. Huesken, J. Lange, C. Mickanin, J. Weiler, F. Asselbergs, J. Warner, B. Meloon, S. Engel, A. Rosenberg, D. Cohen, M. Labow, M. Reinhardt, F. Natt, and J. Hall. Design of a genome-wide siRNA library using an artificial neural network. *Nat. Biotechnol.*, 23(8):995–1001, Aug 2005.

[743] R. Hughey and A. Krogh. Hidden Markov models for sequence analysis: Extension and analysis of the basic method. *CABIOS*, 12(2):95–107, 1996.

[744] W.-K. Huh, J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman, and E. K. O'Shea. Global analysis of protein localization in budding yeast. *Nature*, 425(6959):686–691, Oct 2003.

[745] W. Humphrey, A. Dalke, and K. Schulten. VMD: visual molecular dynamics. *J. Mol. Graph.*, 14(1):33–8, 27–8, Feb 1996.

[746] P. Hupé, N. Stransky, J.-P. Thiery, F. Radvanyi, and E. Barillot. Analysis of array CGH data: from signal ratio to gain and loss of dna regions. *Bioinformatics*, 20(18):3413–3422, Dec 2004.

[747] K. Huppi, S. E. Martin, and N. J. Caplen. Defining and assaying RNAi in mammalian cells. *Mol. Cell*, 17(1):1–10, Jan 2005.

[748] B. Hutter, C. Schaab, S. Albrecht, M. Borgmann, N. A. Brunner, C. Freiberg, K. Ziegelbauer, C. O. Rock, I. Ivanov, and H. Loferer. Prediction of Mechanisms of Action of Antibacterial Compounds by Gene Expression Profiling. *Antimicrob. Agents Chemother.*, 48(8):2838–2844, Aug 2004.

[749] A John Iafrate, Lars Feuk, Miguel N Rivera, Marc L Listewnik, Patricia K Donahoe, Ying Qi, Stephen W Scherer, and Charles Lee. Detection of large-scale variation in the human genome. *Nature Genetics*, 36:949–951, 2004.

[750] A. John Iafrate, Lars Feuk, Miguel N Rivera, Marc L Listewnik, Patricia K Donahoe, Ying Qi, Stephen W Scherer, and Charles Lee. Detection of large-scale variation in the human genome. *Nat. Genet.*, 36(9):949–951, Sep 2004.

[751] A. Ifantis and S. Papadimitriou. The nonlinear predictability of the electrotelluric field variations data analyzed with support vector machines as an earthquake precursor. *Int J Neural Syst*, 13(5):315–32, Oct 2003.

[752] Norio Iizuka, Masaaki Oka, Hisafumi Yamada-Okabe, Minekatsu Nishida, Yoshitaka Maeda, Naohide Mori, Takashi Takao, Takao Tamesa, Akira Tangoku, Hisahiro Tabuchi, Kenji Hamada, Hironobu Nakayama, Hideo Ishitsuka, Takanobu Miyamoto, Akira Hirabayashi, Shunji Uchimura, and Yoshihiko Hamamoto. Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection. *Lancet*, 361(9361):923–9, Mar 2003.

[753] Kazushi Ikeda and Tsutomu Aoishi. An asymptotic statistical analysis of support vector machines with soft margins. *Neural Netw*, 18(3):251–9, Apr 2005.

[754] Naoya Ikeda and Takashi Uozumi. [Tongue diagnosis support system]. *Hokkaido Igaku Zasshi*, 80(3):269–77, May 2005.

[755] S. Imoto, T. Goto, and S. Miyano. Estimation of genetic networks and functional structures between genes by using Bayesian networks and nonparametric regression. *Pac. Symp. Biocomput.*, pages 175–186, 2002.

[756] S. Imoto, S. Kim, T. Goto, S. Miyano, S. Aburatani, K. Tashiro, and S. Kuhara. Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network. *J. Bioinform. Comput. Biol.*, 1(2):231–252, Jul 2003.

[757] S. Imoto, K. Sunyong, T. Goto, S. Aburatani, K. Tashiro, S. Kuhara, and S. Miyano. Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network. *Proc. IEEE Comput. Soc. Bioinform. Conf.*, 1:219–227, 2002.

[758] A. Inokuchi, T. Washio, and H. Motoda. Complete mining of frequent patterns from graphs: mining graph data. *Mach. Learn.*, 50(3):321–354, 2003.

[759] International Union of Biochemistry and Molecular Biology. *Enzyme Nomenclature 1992*. Academic Press, San Diego, California, United States, August 1992.

[760] Rafael A. Irizarry, Bridget Hobbs, Francois Collin, Yasmin D. Beazer-Barclay, Kristen J. Antonellis, Uwe Scherf, and Terence P. Speed. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4:249–264, 2003.

[761] J. J. Irwin and B. K. Shoichet. ZINC–a free database of commercially available compounds for virtual screening. *J Chem Inf Model*, 45(1):177–82, 2005.

[762] A. S. Ishkanian, C. A. Malloff, S. K. Watson, R. J. DeLeeuw, B. Chi, B. P. Coe, A. Snijders, D. G. Albertson, D. Pinkel, M. A. Marra, V. Ling, C. MacAulay, and W. L. Lam. A tiling resolution DNA microarray with complete coverage of the human genome. *Nat. Genet.*, 36(3):299–303, Mar 2004.

[763] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA*, 98(8):4569–4574, 2001.

[764] T. Ito, K. Tashiro, S. Muta, R. Ozawa, T. Chiba, M. Nishizawa, K. Yamamoto, S. Kuhara, and Y. Sakaki. Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl. Acad. Sci. USA*, 93(3):1143–1147, 2000.

[765] O. Ivanciuc. Applications of support vector machines in chemistry. In K. B. Lipkowitz and T. R. Cundari, editors, *Reviews in Computational Chemistry*, volume 23, pages 291–400, Weiheim, 2007. Wiley-VCH.

[766] V.V. Ivanov. *The theory of approximate methods and their application to the numerical solution of singular integral equations.* Nordhoff International, Leiden, 1976.

[767] Natraj Iyer, Subramaniam Jayanti, Kuiyang Lou, Yagnanarayanan Kalyanaraman, and Karthik Ramani. Three-dimensional shape searching: state-of-the-art review and future trends. *Computer-Aided Design*, 37(5):509–530, April 2005.

[768] T. Jaakkola, M. Diekhans, and D. Haussler. A Discriminative Framework for Detecting Remote Protein Homologies. *J. Comput. Biol.*, 7(1,2):95–114, 2000.

[769] T. S. Jaakkola, M. Diekhans, and D. Haussler. Using the Fisher Kernel Method to Detect Remote Protein Homologies. In *Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology*, pages 149–158. AAAI Press, 1999.

[770] T. S. Jaakkola and D. Haussler. Exploiting generative models in discriminative classifiers. In *Proc. of Tenth Conference on Advances in Neural Information Processing Systems*, 1999.

[771] T. S. Jaakkola and D. Haussler. Probabilistic kernel regression models. In *Proceedings of the 1999 Conference on AI and Statistics.* Morgan Kaufmann, 1999.

[772] Tommi Jaakkola, Marina Meila, and Tony Jebara. Maximum entropy discrimination. In *Adv. Neural Inform. Process. Syst.*, volume 12. MIT Press, Cambridge, MA, 1999.

[773] A. Jablonka and M. J. Lamb. The changing concept of epigenetics. *Ann N Y Acad Sci*, 981:82–96, Dec 2002.

[774] A. L. Jackson, S. R. Bartz, J. Schelter, S. V. Kobayashi, J. Burchard, M. Mao, B. Li, G. Cavet, and P. S. Linsley. Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.*, 21(6):635–7, Jun 2003.

[775] A. L. Jackson and P. S. Linsley. Noise amidst the silence: off-target effects of siRNAs? *Trends Genet.*, 20(11):521–4, Nov 2004.

[776] L. Jacob, F. Bach, and J.-P. Vert. Clustered multi-task learning: A convex formulation. In *Advances in Neural Information Processing Systems 21*, pages 745–752. MIT Press, 2009.

[777] L. Jacob, B. Hoffmann, B. Stoven, and J.-P. Vert. Virtual screening of GPCRs: an *in silico* chemogenomics approach. *BMC Bioinformatics*, 9:363, 2008.

[778] L. Jacob, B. Hoffmann, B. Stoven, and J.-P. Vert. Virtual screening of GPCRs: an *in silico* chemogenomics approach. Technical Report 0801.4301, Arxiv, 2008.

[779] L. Jacob, G. Obozinski, and J.-P. Vert. Group lasso with overlaps and graph lasso. In *ICML'09 Proceedings of the 26th international conference on Machine learning*, 2009. To appear.

[780] L. Jacob and J.-P. Vert. Epitope prediction improved by multitask support vector machines. Technical Report arXiv:q-bio/0702008v1, arXiv, 2006.

[781] L. Jacob and J.-P. Vert. Kernel methods for in silico chemogenomics. Technical Report 0709.3931v1, arXiv, 2007.

[782] L. Jacob and J.-P. Vert. Efficient peptide-MHC-I binding prediction for alleles with few known binders. *Bioinformatics*, 24(3):358–366, Feb 2008.

[783] L. Jacob and J.-P. Vert. Protein-ligand interaction prediction: an improved chemogenomics approach. *Bioinformatics*, 24(19):2149–2156, 2008.

[784] E. Jacoby, editor. *Chemogenomics: Knowledge-based Approaches to Drug Discovery.* Imperial College Press, 2006.

[785] Edgar Jacoby, Rochdi Bouhelal, Marc Gerspacher, and Klaus Seuwen. The 7 tm g-protein-coupled receptor target family. *ChemMedChem*, 1(8):761–782, Aug 2006.

[786] Rudolf Jaenisch and Adrian Bird. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*, 33 Suppl:245–254, Mar 2003.

[787] Martin Jambon, Anne Imberty, Gilbert Delage, and Christophe Geourjon. A new bioinformatic approach to detect common 3d sites in protein structures. *Proteins*, 52(2):137–145, Aug 2003.

[788] C. Jamieson, E. M. Moir, Z. Rankovic, and G. Wishart. Medicinal chemistry of hERG optimizations: Highlights and hang-ups. *J. Med. Chem.*, 49(17):5029–5046, Aug 2006.

[789] Isabelle Janoueix-Lerosey, Philippe Hup, Zofia Maciorowski, Philippe La Rosa, Gudrun Schleiermacher, Gaelle Pierron, Stephane Liva, Emmanuel Barillot, and Olivier Delattre. Preferential occurrence of chromosome breakpoints within early replicating regions in neuroblastoma. *Cell Cycle*, 4:1842–1846, 2005.

[790] R. Jansen, H. Yu, D. Greenbaum, Y. Kluger, N.J. Krogan, S. Chung, A. Emili, M. Snyder, J.F. Greenblatt, and M. Gerstein. A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science*, 302(5644):449–453, 2003.

[791] S. E. Jaroch and H. Weinmann, editors. *Chemical Genomics: Small Molecule Probes to Study Cellular Function.* Ernst Schering Research Foundation Workshop. Springer, Berlin, 2006.

[792] Barbara Jarzab, Malgorzata Wiench, Krzysztof Fujarewicz, Krzysztof Simek, Michal Jarzab, Malgorzata Oczko-Wojciechowska, Jan Wloch, Agnieszka Czarniecka, Ewa Chmielik, Dariusz Lange, Agnieszka Pawlaczek, Sylwia Szpak, Elzbieta Gubala, and Andrzej Swierniak. Gene expression profile of papillary thyroid cancer: sources of variability and diagnostic implications. *Cancer Res.*, 65(4):1587–97, Feb 2005.

[793] T. Jebara, R. Kondor, and A. Howard. Probability Product Kernels. *J. Mach. Learn. Res.*, 5:819–844, 2004.

[794] Tony Jebara. Multi-task feature and kernel selection for svms. In *ICML '04: Proceedings of the twenty-first international conference on Machine learning*, page 55, New York, NY, USA, 2004. ACM.

[795] R. Jenatton, J.-Y. Audibert, and F. Bach. Structured Variable Selection with Sparsity-Inducing Norms. Research report, WILLOW - INRIA Rocquencourt - INRIA - Ecole Normale Supérieure de Paris - Ecole Nationale des Ponts et Chaussées - CNRS : UMR8548 - Imagine - Université Paris-Est, 2009.

[796] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411:41–42, 2001.

[797] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.

[798] Anna K Jerebko, James D Malley, Marek Franaszek, and Ronald M Summers. Support vector machines committee classification method for computer-aided polyp detection in CT colonography. *Acad Radiol*, 12(4):479–86, Apr 2005.

[799] P. Jia, T. Shi, Y. Cai, and Y. Li. Demonstration of two novel methods for predicting functional siRNA efficiency. *BMC Bioinformatics*, 7:271, 2006.

[800] S. Jiang-Ning, L. Wei-Jiang, and X. Wen-Bo. Cooperativity of the oxidization of cysteines in globular proteins. *J. Theor. Biol.*, 231(1):85–95, 2004.

[801] T. Joachims. Making large-Scale SVM Learning Practical. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, pages 169–184. MIT Press, 1999.

[802] T. Joachims. *Learning to Classify Text Using Support Vector Machines.* Kluwer Academic Publishers, 2002.

[803] D. S. Johnson, A. Mortazavi, R. M. Myers, and B. Wold. Genome-wide mapping of in vivo protein-dna interactions. *Science*, 316(5830):1497–1502, Jun 2007.

[804] M. A. Johnson and G. M. Maggiora, editors. *Concepts and Applications of Molecular Similarity.* Wiley, 1990.

[805] Sam A Johnson and Tony Hunter. Kinomics: methods for deciphering the kinome. *Nature Methods*, 2:17–25, 2005.

[806] N. Jojic, M. Reyes-Gomez, D. Heckerman, C. Kadie, and O. Schueler-Furman. Learning MHC I–peptide binding. *Bioinformatics*, 22(14):e227–e235, Jul 2006.

[807] I.T. Jolliffe. *Principal component analysis.* Springer-Verlag, New-York, 1996.

[808] C. Jones, E. Ford, C. Gillett, K. Ryder, S. Merrett, J. S. Reis-Filho, L. G. Fulford, A. Hanby, and S. R. Lakhani. Molecular cytogenetic identification of subgroups of grade iii invasive ductal breast carcinomas with different clinical outcomes. *Clin. Cancer Res.*, 10(18):5988–5997, 2004.

[809] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*, 267(3):727–748, Apr 1997.

[810] P. A. Jones and S. B. Baylin. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.*, 3(6):415–428, Jun 2002.

[811] Peter A Jones. Dna methylation and cancer. *Oncogene*, 21(35):5358–5360, Aug 2002.

[812] K. Jong, E. Marchiori, G. Meijer, A. V. D. Vaart, and B. Ylstra. Breakpoint identification and smoothing of array comparative genomic hybridization data. *Bioinformatics*, 20(18):3636–3637, Dec 2004.

[813] Michael Jordan, editor. *Learning in Graphical Models.* The MIT Press, 2001.

[814] R. A. Jorgensen. Sense cosuppression in plants: Past, present, and future. In G. J. Hannon, editor, *RNAi: A guide to gene silencing.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2003.

[815] W. L. Jorgensen. The many roles of computation in drug discovery. *Science*, 303(5665):1813–1818, Mar 2004.

[816] R. N. Jorissen and M. K. Gilson. Virtual screening of molecular databases using a support vector machine. *J Chem Inf Model*, 45(3):549–61, 2005.

[817] G. Joshi-Tope, M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt, B. de Bono, B. Jassal, G. R. Gopinath, G. R. Wu, L. Matthews, S. Lewis, E. Birney, and L. Stein. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res*, 33(Database issue):D428–32, 2005. 1362-4962 (Electronic) Journal Article.

[818] A. Juditsky and A. Nemirovski. Functional Aggregation for Nonparametric Estimation. *Ann. Stat.*, 28(3):681–712, June 2000.

[819] Svava Osk Jnsdttir, Flemming Steen Jrgensen, and Sren Brunak. Prediction methods and databases within chemoinformatics: emphasis on drugs and drug candidates. *Bioinformatics*, 21(10):2145–2160, May 2005.

[820] Abdullah Kahraman, Richard J Morris, Roman A Laskowski, and Janet M Thornton. Shape variation in protein binding pockets and their ligands. *J Mol Biol*, 368(1):283–301, Apr 2007.

[821] I. Kalatzis, D. Pappas, N. Piliouras, and D. Cavouras. Support vector machines based analysis of brain SPECT images for determining cerebral abnormalities in asymptomatic diabetic patients. *Med Inform Internet Med*, 28(3):221–30, Sep 2003.

[822] I. Kalatzis, N. Piliouras, E. Ventouras, C. C. Papageorgiou, A. D. Rabavilas, and D. Cavouras. Design and implementation of an SVM-based computer classification system for discriminating depressive patients from healthy controls using the P600 component of ERP signals. *Comput Methods Programs Biomed*, 75(1):11–22, Jul 2004.

[823] A. Kallioniemi. CGH microarrays and cancer. *Current Opinion in Biotechnology*, 18:1–5, 2007.

[824] A Kallioniemi, O-P Kallioniemi, D Sudar, D Rutovitz, J W Gray, F M Waldman, and D Pinkel. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, 258:818–821, 1992.

[825] F. Kamangar, G. M. Dores, and W. F. Anderson. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J. Clin. Oncol.*, 24(14):2137–2150, May 2006.

[826] R. S. Kamath, A. G. Fraser, Y. Dong, G. Poulin, R. Durbin, M. Gotta, A. Kanapin, N. Le Bot, S. Moreno, M. Sohrmann, D. P. Welchman, P Zipperlen, and J. Ahringer. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature*, 421(6920):231–237, Jan 2003.

[827] J. Kandola, J. Shawe-Taylor, and N. Cristianini. On the application of diffusion kernel to text data. Technical report, Neurocolt, 2002. NeuroCOLT Technical Report NC-TR-02-122.

[828] J. Kandola, J. Shawe-Taylor, and N. Cristianini. Learning Semantic Similarity. In Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, editors, *Advances in Neural Information Processing Systems 15*. MIT Press, 2003.

[829] M. Kanehisa. A database for post-genome analysis. *Trends Genet.*, 13:375–376, 1997.

[830] M. Kanehisa. Prediction of higher order functional networks from genomic data. *Pharmacogenomics*, 2(4):373–385, 2001.

[831] M. Kanehisa, S. Goto, S. Kawashima, and A. Nakaya. The KEGG databases at GenomeNet. *Nucleic Acids Res.*, 30:42–46, 2002.

[832] M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, and M. Hattori. The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, 32(Database issue):D277–80, Jan 2004.

[833] Matthias Kaper, Peter Meinicke, Ulf Grossekathoefer, Thomas Lingner, and Helge Ritter. BCI Competition 2003–Data set IIb: support vector machines for the P300 speller paradigm. *IEEE Trans Biomed Eng*, 51(6):1073–6, Jun 2004.

[834] Izet M Kapetanovic, Simon Rosenfeld, and Grant Izmirlian. Overview of commonly used bioinformatics methods and their applications. *Ann N Y Acad Sci*, 1020:10–21, May 2004.

[835] R. Karchin, K. Karplus, and D. Haussler. Classifying G-protein coupled receptors with support vector machines. *Bioinformatics*, 18:147–159, 2002.

[836] R. Karchin, L. Kelly, and A. Sali. Improving functional annotation of non-synonomous SNPs with information theory. *Pac Symp Biocomput*, pages 397–408, 2005.

[837] Y. Karklin, R. F. Meraz, and S.R. Holbrook. Classification of non-coding RNA using graph representations of secondary structure. *Pac. Symp. Biocomput.*, pages 4–15, 2005.

[838] Yan Karklin, Richard F Meraz, and Stephen R Holbrook. Classification of non-coding RNA using graph representations of secondary structure. *Pac Symp Biocomput*, pages 4–15, 2005.

[839] Gerhard Hessler Karl-Heinz Baringhaus. A chemical genomics approach for ion channel modulators. In *Chemogenomics in Drug Discovery*, chapter 8, pages 221–242. Wiley-VCH, 2005.

[840] P. D. Karp, C. A. Ouzounis, C. Moore-Kochlacs, L. Goldovsky, P. Kaipa, D. Ahren, S. Tsoka, N. Darzentas, V. Kunin, and N. Lopez-Bigas. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res*, 33(19):6083–9, 2005.

[841] K. Karplus, C. Barrett, and R. Hughey. Hidden Markov Models for Detecting Remote Protein Homologies. *Bioinformatics*, 14(10):846–856, 1998.

[842] H. Kashima, K. Tsuda, and A. Inokuchi. Marginalized Kernels between Labeled Graphs. In T. Faucett and N. Mishra, editors, *Proceedings of the Twentieth International Conference on Machine Learning*, pages 321–328, New York, NY, USA, 2003. AAAI Press.

[843] H. Kashima, K. Tsuda, and A. Inokuchi. Kernels for graphs. In B. Schölkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 155–170. MIT Press, The MIT Press, Cambridge, Massachussetts, 2004.

[844] T. Kato. Operator dynamics in molecular biology. Technical report, I.H.E.S., 2001. Technical report IHES/M/01/41.

[845] T. Kato, K. Tsuda, and K. Asai. Selective integration of multiple biological data for supervised network inference. *Bioinformatics*, 21(10):2488–2495, May 2005.

[846] Michael Kazhdan, Thomas Funkhouser, and Szymon Rusinkiewicz. Rotation invariant spherical harmonic representation of 3d shape descriptors. In *SGP '03: Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 156–164, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association.

[847] S. S. Keerthi and S. K. Shevade. SMO algorithm for least-squares SVM formulations. *Neural Comput*, 15(2):487–507, Feb 2003.

[848] S. Sathiya Keerthi and Chih-Jen Lin. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Comput*, 15(7):1667–89, Jul 2003.

[849] E. Kellenberger, J. Rodrigo, P. Muller, and D. Rognan. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins*, 57(2):225–242, Nov 2004.

[850] B.P. Kelley, R. Sharan, R.M. Karp, T. Sittler, D.E. Root, B.R. Stockwell, and T. Ideker. Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proc. Natl. Acad. Sci. USA*, 100(20):11394–11399, Sep 2003.

[851] B.P. Kelley, B. Yuan, F. Lewitter, R. Sharan, B.R. Stockwell, and T. Ideker. PathBLAST: a tool for alignment of protein interaction networks. *Nucleic Acids Res.*, 32(Web Server issue):W83–W88, Jul 2004.

[852] R. Kelley and T. Ideker. Systematic interpretation of genetic interactions using protein networks. *Nat. Biotechnol.*, 23(5):561–566, May 2005.

[853] Yakov Keselman, Ali Shokoufandeh, M. Fatih Demirci, and Sven Dickinson. Many-to-many graph matching via metric embedding. In *In CVPR), pages I850 I857 vol.1*, pages 850–857, 2003.

[854] G. M. Keserü. Prediction of hERG potassium channel affinity by traditional and hologram qSAR methods. *Bioorg. Med. Chem. Lett.*, 13(16):2773–2775, Aug 2003.

[855] Shahid M Khan, Blandine Franke-Fayard, Gunnar R Mair, Edwin Lasonder, Chris J Janse, Matthias Mann, and Andrew P Waters. Proteome analysis of separated male and female gametocytes reveals novel sex-specific Plasmodium biology. *Cell*, 121(5):675–687, Jun 2005.

[856] P. Kharchenko, L. Chen, Y. Freund, D. Vitkup, and G. M. Church. Identifying metabolic enzymes with multiple types of association evidence. *BMC Bioinformatics*, 7:177, 2006.

[857] P. Kharchenko, D. Vitkup, and G. M. Church. Filling gaps in a metabolic network using expression information. *Bioinformatics*, 20 Suppl 1:I178–I185, Aug 2004.

[858] A. Khvorova, A. Reynolds, and S.D. Jayasena. Functional siRNAs and miRNAs exhibit strand bias. *Cell*, 115(2):209–216, Oct 2003.

[859] J. Kieffer. A unified approach to weak universal source coding. *IEEE Trans. Inform. Theory*, 24(6):674–682, Nov 1978.

[860] Eddo Kim, Amir Goren, and Gil Ast. Insights into the connection between cancer and alternative splicing. *Trends in Genetics*, 24:7–10, 2008.

[861] H. Kim and H. Park. Protein secondary structure prediction based on an improved support vector machines approach. *Protein Eng.*, 16(8):553–560, Aug 2003.

[862] H. Kim and H. Park. Prediction of protein relative solvent accessibility with support vector machines and long-range interaction 3D local descriptor. *Proteins*, 54(3):557–562, Feb 2004.

[863] J. Kim, P.L. Krapivsky, B. Kahng, and S. Redner. Evolving protein interaction networks. E-print cond-mat/0203167, 2001.

[864] J. H. Kim, J. Lee, B. Oh, K. Kimm, and I. Koh. Prediction of phosphorylation sites using SVMs. *Bioinformatics*, 20(17):3179–3184, 2004.

[865] K. H. Kim, S. W. Bang, and S. R. Kim. Emotion recognition system using short-term monitoring of physiological signals. *Med Biol Eng Comput*, 42(3):419–27, May 2004.

[866] Sang-Woon Kim and B. John Oommen. Enhancing prototype reduction schemes with recursion: a method applicable for "large" data sets. *IEEE Trans Syst Man Cybern B Cybern*, 34(3):1384–97, Jun 2004.

[867] Tae-Kyun Kim and Josef Kittler. Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. *IEEE Trans Pattern Anal Mach Intell*, 27(3):318–27, Mar 2005.

[868] G. S. Kimeldorf and G. Wahba. Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.*, 33:82–95, 1971.

[869] T. Kin, K. Tsuda, and K. Asai. Marginalized kernels for RNA sequence data analysis. In R.H. Lathtop, K. Nakai, S. Miyano, T. Takagi, and M. Kanehisa, editors, *Genome Informatics 2002*, pages 112–122. Universal Academic Press, 2002.

[870] R. D. King, S. Muggleton, R. A. Lewis, and M. J. Sternberg. Drug design by machine learning: the use of inductive logic programming to model the structure-activity relationships of trimethoprim analogues binding to dihydrofolate reductase. *Proc. Natl. Acad. Sci. USA*, 89(23):11322–11326, Dec 1992.

[871] R. D. King, S. H. Muggleton, A. Srinivasan, and M. J. Sternberg. Structure-activity relationships derived by machine learning: the use of atoms and their bond connectivities to predict mutagenicity by inductive logic programming. *Proc. Natl. Acad. Sci. USA*, 93(1):438–442, Jan 1996.

[872] M. Kirby and L. Sirovich. Application of the Karhunen-Loève procedure for the characterization of human faces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(1):103–108, 1990.

[873] H. Kitano. *Foundations of Systems Biology.* MIT Press, 2001.

[874] H. Kitano. Computational systems biology. *Nature*, 420:206–210, 2002.

[875] H. Kitano. Cancer as a robust system: implications for anticancer therapy. *Nat. Rev. Cancer*, 4:227–235, 2004.

[876] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*, 3(11):935–949, Nov 2004.

[877] T. Klabunde. Chemogenomics approaches to ligand design. In *Ligand Design for G Protein-coupled Receptors*, chapter 7, pages 115–135. Wiley-VCH, Great Britain, 2006.

[878] T. Klabunde. Chemogenomic approaches to drug discovery: similar receptors bind similar ligands. *Br. J. Pharmacol.*, 152:5–7, May 2007.

[879] T. Klabunde and R. Jäger. Chemogenomics approaches to g-protein coupled receptor lead finding. *Ernst Schering Res Found Workshop*, 58:31–46, 2006.

[880] G. Klebe. Recent developments in structure-based drug design. *J Mol Med*, 78(5):269–281, 2000.

[881] K. Knight and W. Fu. Asymptotics for lasso-type estimators. *Ann. Stat.*, 28(5):1356–1378, 2000.

[882] Alfred G. Knudson. Mutation and cancer: Statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences*, 68:820–823, 1971.

[883] B. K. Kobilka. G protein coupled receptor structure and activation. *Biochim. Biophys. Acta*, 1768(4):794–807, Apr 2007.

[884] A. Kohlmann, C. Schoch, S. Schnittger, M. Dugas, W. Hiddemann, W. Kern, and T. Haferlach. Pediatric acute lymphoblastic leukemia (ALL) gene expression signatures classify an independent cohort of adult ALL patients. *Leukemia*, 18(1):63–71, 2004.

[885] A. Koike and T. Takagi. Prediction of protein-protein interaction sites using support vector machines. *Protein Eng. Des. Sel.*, 17(2):165–173, Feb 2004.

[886] V. Koltchinskii. Localized Rademacher complexities. Manuscript, september 2003.

[887] D. Komura, H. Nakamura, S. Tsutsumi, H. Aburatani, and S. Ihara. Multidimensional support vector machines for visualization of gene expression data. *Bioinformatics*, 21(4):439–444, Feb 2005.

[888] R. Kondor and J.-P. Vert. Diffusion kernels. In B. Schölkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 171–192. MIT Press, 2004.

[889] R. I. Kondor and J. Lafferty. Diffusion Kernels on Graphs and Other Discrete Input. In *ICML 2002*, 2002.

[890] Risi Kondor and Karsten M. Borgwardt. The skew spectrum of graphs. In *ICML '08: Proceedings of the 25th international conference on Machine learning*, pages 496–503, New York, NY, USA, 2008. ACM.

[891] Risi Kondor and Tony Jebara. A kernel between sets of vectors. In *In International Conference on Machine Learning (ICML)*, 2003.

[892] Risi Kondor, Nino Shervashidze, and Karsten M. Borgwardt. The graphlet spectrum. In *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*, pages 529–536, New York, NY, USA, 2009. ACM.

[893] J. Korbel, A. Abyzov, X. Mu, N. Carriero, P. Cayting, Z. Zhang, Z. Snyder, and M. Gerstein. PEMer: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data. *Genome Biol.*, 10(2):R23, Feb 2009.

[894] Jan O Korbel, Alexander Eckehart Urban, Jason P Affourtit, Brian Godwin, Fabian Grubert, Jan Fredrik Simons, Philip M Kim, Dean Palejev, Nicholas J Carriero, Lei Du, Bruce E Taillon, Zhoutao Chen, Andrea Tanzer, A. C Eugenia Saunders, Jianxiang Chi, Fengtang Yang, Nigel P Carter, Matthew E Hurles, Sherman M Weissman, Timothy T Harkins, Mark B Gerstein, Michael Egholm, and Michael Snyder. Paired-end mapping reveals extensive structural variation in the human genome. *Science*, 318(5849):420–426, Oct 2007.

[895] Bette Korber, Montiago LaBute, and Karina Yusim. Immunoinformatics comes of age. *PLoS Comput Biol*, 2(6):e71, Jun 2006.

[896] K. Korn and E. Krausz. Cell-based high-content screening of small-molecule libraries. *Curr. Opin. Chem. Biol.*, 11(5):503–510, Oct 2007.

[897] Zsofia Kote-Jarai, Richard D Williams, Nicola Cattini, Maria Copeland, Ian Giddings, Richard Wooster, Robert H tePoele, Paul Workman, Barry Gusterson, John Peacock, Gerald Gui, Colin Campbell, and Ros Eeles. Gene expression profiling after radiation-induced DNA damage is strongly predictive of BRCA1 mutation carrier status. *Clin. Cancer Res.*, 10(3):958–63, Feb 2004.

[898] Zhenzhen Kou, Liang Ji, and Xuegong Zhang. Karyotyping of comparative genomic hybridization human metaphases by using support vector machines. *Cytometry*, 47(1):17–23, Jan 2002.

[899] Assia Kovatcheva, Alexander Golbraikh, Scott Oloff, Yun-De Xiao, Weifan Zheng, Peter Wolschann, Gerhard Buchbauer, and Alexander Tropsha. Combinatorial QSAR of ambergris fragrance compounds. *J Chem Inf Comput Sci*, 44(2):582–95, 2004.

[900] M. Koyutürk, Y. Kim, U. Topkara, S. Subramaniam, W. Szpankowski, and A. Grama. Pairwise alignment of protein interaction networks. *J. Comput. Biol.*, 13(2):182–199, Mar 2006.

[901] S. Kramer and L. De Raedt. Feature Construction with Version Spaces for Biochemical Applications. In C.E. Brodley and A. Pohoreckyj Danyluk, editors, *Proceedings of the Eighteenth International Conference on Machine Learning*, pages 258–265. Morgan Kaufmann, 2001.

[902] S. Kramer, E. Frank, and C. Helma. Fragment generation and support vector machines for inducing SARs. *SAR QSAR Environ Res*, 13(5):509–23, Jul 2002.

[903] N. A. Kratochwil, P. Malherbe, L. Lindemann, M. Ebeling, M. C. Hoener, A. Mühlemann, R. H. P. Porter, M. Stahl, and P. R. Gerber. An automated system for the analysis of G protein-coupled receptor transmembrane binding pockets: alignment, receptor-based pharmacophores, and their application. *J. Chem. Inf. Model.*, 45(5):1324–1336, 2005.

[904] R. E. Krichevskiy. Laplace's law of succession and universal encoding. *IEEE Trans. Inform. Theory*, 44(1):296–303, Jan 1998.

[905] R. Krichevsky and V. Trofimov. The performance of universal encoding. *IEEE Trans. Inform. Theory*, 27(2):199–207, Mar 1981.

[906] V. G. Krishnan and D. R. Westhead. A comparative study of machine-learning methods to predict the effects of single nucleotide polymorphisms on protein function. *Bioinformatics*, 19(17):2199–2209, 2003.

[907] B. Krishnapuram, L. Carin, and A. Hartemink. Gene expression analysis: joint feature selection and classifier design. In B. Schlkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 299–317. MIT Press, 2004.

[908] B. Krishnapuram, L. Carin, and A. Hartemink. Joint Classifier and Feature Optimization for Comprehensive Cancer Diagnosis Using Gene Expression Data. *J. Comput. Biol.*, 11(2-3):227–242, 2004.

[909] Balaji Krishnapuram, Alexander J Hartemink, Lawrence Carin, and Mrio A T Figueiredo. A bayesian approach to joint feature selection and classifier design. *IEEE Trans Pattern Anal Mach Intell*, 26(9):1105–11, Sep 2004.

[910] K. Kristiansen, S. G. Dahl, and O. Edvardsen. A database of mutants and effects of site-directed mutagenesis experiments on G protein-coupled receptors. *Proteins*, 26(1):81–94, Sep 1996.

[911] Romano T Kroemer. Structure-based drug design: docking and scoring. *Curr Protein Pept Sci*, 8(4):312–328, Aug 2007.

[912] A. Krogh, M. Brown, I. Mian, K. Sjolander, and D. Haussler. Hidden Markov models in computational biology: Applications to protein modeling. *J. Mol. Biol.*, 235:1501–1531, 1994.

[913] M. Krull, S. Pistor, N. Voss, A. Kel, I. Reuter, D. Kronenberg, H. Michael, K. Schwarzer, A. Potapov, C. Choi, O. Kel-Margoulis, and E. Wingender. TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Res*, 34(Database issue):D546–51, 2006.

[914] R. Kuang, E. Ie, K. Wang, K. Wang, M. Siddiqi, Y. Freund, and C. Leslie. Profile-based string kernels for remote homology detection and motif extraction. *Proc IEEE Comput Syst Bioinform Conf*, pages 152–160, 2004.

[915] R. Kuang, E. Ie, K. Wang, K. Wang, M. Siddiqi, Y. Freund, and C. Leslie. Profile-based string kernels for remote homology detection and motif extraction. *J. Bioinform. Comput. Biol.*, 3(3):527–550, Jun 2005.

[916] R. Kuang, C. S. Leslie, and A.-S. Yang. Protein backbone angle prediction with machine learning approaches. *Bioinformatics*, 20(10):1612–1621, 2004.

[917] H. Kubinyi. Comparative Molecular Field Analysis. In J. Gasteiger, editor, *Handbook of Chemoinformatics. From Data to Knowledge, Volume 4*, pages 1555–1574. Wiley-VCH, Weinheim, 2003.

[918] H. Kubinyi. Chemogenomics in drug discovery. *Ernst Schering Res Found Workshop*, 58:1–19, 2006.

[919] H. Kubinyi, G. Müller, R. Mannhold, and G. Folkers, editors. *Chemogenomics in Drug Discovery: A Medicinal Chemistry Perspective*. Methods and Principles in Medicinal Chemistry. Wiley-VCH, New York, 2004.

[920] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research*, 2:83–97, 1955.

[921] K. M. Kuhn, J. L. DeRisi, P. O. Brown, and P. Sarnow. Global and specific translational regulation in the genomic response of Saccharomyces cerevisiae to a rapid transfer from a fermentable to a nonfermentable carbon source. *Mol. Cell. Biol.*, 21(3):916–927, 2001.

[922] W. Kuich and A. Salomaa. Semirings, Automata, Languages. In *EATCS Monographs on Computer Science*, volume 5. Springer-Verlag, 1986.

[923] Pavel P. Kuksa, Pai-Hsi Huang, and Vladimir Pavlovic. Scalable algorithms for string kernels with inexact matching. In Daphne Koller, Dale Schuurmans, Yoshua Bengio, and L'eon Bottou, editors, *NIPS*, pages 881–888. MIT Press, 2008.

[924] S. Kumagai. An implicit function theorem: Comment. *Journal of Optimization Theory and Applications*, 31:285–288, Jun 1980.

[925] M. Kumar, M. Bhasin, N. K. Natt, and G. P. S. Raghava. BhairPred: prediction of beta-hairpins in a protein from multiple alignment information using ANN and SVM techniques. *Nucleic Acids Res*, 33(Web Server issue):W154–9, Jul 2005.

[926] Ludmila I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.

[927] Stephen LaConte, Stephen Strother, Vladimir Cherkassky, Jon Anderson, and Xiaoping Hu. Support vector machines for temporal classification of block design fMRI data. *Neuroimage*, 26(2):317–29, Jun 2005.

[928] S. Lacoste-Julien. An introduction to Max-Margin Markov Networks. UC Berkeley cs281a project report, December 2003.

[929] Douglas J LaCount, Marissa Vignali, Rakesh Chettier, Amit Phansalkar, Russell Bell, Jay R Hesselberth, Lori W Schoenfeld, Irene Ota, Sudhir Sahasrabudhe, Cornelia Kurschner, Stanley Fields, and Robert E Hughes. A protein interaction network of the malaria parasite Plasmodium falciparum. *Nature*, 438(7064):103–107, Nov 2005.

[930] J. Lafferty, A. McCallum, and F. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proc. 18th International Conf. on Machine Learning*, pages 282–289. Morgan Kaufmann, San Francisco, CA, 2001.

[931] P.L. Lai and C. Fyfe. Kernel and nonlinear canonical correlation analysis. *Int. J. Neural Syst.*, 10(5):365–377, 2000.

[932] Thomas Navin Lal, Michael Schrder, Thilo Hinterberger, Jason Weston, Martin Bogdan, Niels Birbaumer, and Bernhard Schlkopf. Support vector channel selection in BCI. *IEEE Trans Biomed Eng*, 51(6):1003–10, Jun 2004.

[933] G. R. G. Lanckriet, T. De Bie, N. Cristianini, M. I. Jordan, and W. S. Noble. A statistical framework for genomic data fusion. *Bioinformatics*, 20(16):2626–2635, 2004.

[934] G.R. Lanckriet, M. Deng, N. Cristianini, M.I. Jordan, and W.S. Noble. Kernel-based data fusion and its application to protein function prediction in yeast. In *Proceedings of the Pacific Symposium on Biocomputing*, pages 300–311, 2004.

[935] G.R.G. Lanckriet, N. Cristianini, P. Bartlett, L. El Ghaoui, and M.I. Jordan. Learning the Kernel Matrix with Semidefinite Programming. *J. Mach. Learn. Res.*, 5:27–72, 2004.

[936] G.R.G. Lanckriet, N. Cristianini, M.I. Jordan, and W.S. Noble. Kernel-based integration of genomic data using semidefinite programming. In B. Schlkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 231–259. MIT Press, 2004.

[937] S. R. Land and J. H. Friedman. Variable fusion: A new adaptive signal regression method. Technical Report 656, Department of Statistics, Carnegie Mellon University Pittsburgh, 1997.

[938] Zhiqiang Lao, Dinggang Shen, Zhong Xue, Bilge Karacali, Susan M Resnick, and Christos Davatzikos. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *Neuroimage*, 21(1):46–57, Jan 2004.

[939] M. Lapinsh, P. Prusis, A. Gutcaits, T Lundstedt, and J. E. S. Wikberg. Development of proteo-chemometrics: A novel technology of use for analysis of drug-receptor interactions. *Biochem. Biophys. Acta*, 1525:180–190, 2001.

[940] M. Lapinsh, P. Prusis, S. Uhlén, and J. E. S. Wikberg. Improved approach for proteochemometrics modeling: application to organic compound–amine G protein-coupled receptor interactions. *Bioinformatics*, 21(23):4289–4296, Dec 2005.

[941] Mette Voldby Larsen, Claus Lundegaard, Kasper Lamberth, Søren Buus, Søren Brunak, Ole Lund, and Morten Nielsen. An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur J Immunol*, 35(8):2295–2303, Aug 2005.

[942] Edwin Lasonder, Yasushi Ishihama, Jens S Andersen, Adriaan M W Vermunt, Arnab Pain, Robert W Sauerwein, Wijnand M C Eling, Neil Hall, Andrew P Waters, Hendrik G Stunnenberg, and Matthias Mann. Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. *Nature*, 419(6906):537–542, Oct 2002.

[943] Andrs Lass and Emanuele Trucco. Vessel enhancement in digital X-ray angiographic sequences by temporal statistical learning. *Comput Med Imaging Graph*, 29(5):343–55, Jul 2005.

[944] Guillaume Launay and Thomas Simonson. Homology modelling of protein-protein complexes: a simple method and its possibilities and limitations. *BMC Bioinformatics*, 9:427, 2008.

[945] Alasdair T. R. Laurie and Richard M. Jackson. Q-sitefinder: an energy-based method for the prediction of protein–ligand binding sites. *Bioinformatics*, 21(9):1908–1916, 2005.

[946] S. Lauritzen. *Graphical Models.* Oxford, 1996.

[947] J. S. Lazo and P. Wipf. Combinatorial chemistry and contemporary pharmacology. *J Pharmacol Exp Ther*, 293(3):705–709, Jun 2000.

[948] M. Le Novère. MELTING, computing the melting temperature of nucleic acid duplex. *Bioinformatics*, 17(12):1226–7, Dec 2001.

[949] K. G. Le Roch, J. R. Johnson, H. Ahiboh, D.-W. D. Chung, J. Prudhomme, D. Plouffe, K. Henson, Y. Zhou, W. Witola, J. R. Yates, C. Ben Mamoun, E. A. Winzeler, and H. Vial. A systematic approach to understand the mechanism of action of the bisthiazolium compound T4 on the human malaria parasite, Plasmodium falciparum. *BMC Genomics*, 9:513, 2008.

[950] K. G. Le Roch, Y. Zhou, P. L. Blair, M. Grainger, J. K. Moch, J. D. Haynes, P. De la Vega, A. A. Holder, S. Batalov, D. J. Carucci, and E. A. Winzeler. Discovery of Gene Function by Expression Profiling of the Malaria Parasite Life Cycle. *Science*, 301(5639):1503–1508, 2004.

[951] A. R. Leach and V. J. Gillet. *An introduction to chemoinformatics.* Kluwer Academic Publishers, 2003.

[952] Andrew R Leach, Brian K Shoichet, and Catherine E Peishoff. Prediction of protein-ligand interactions. docking and scoring: successes and gaps. *J Med Chem*, 49(20):5851–5855, Oct 2006.

[953] Dongkwon Lee, Sang Wook Choi, Myengsoo Kim, Jin Hyun Park, Moonkyu Kim, Jungchul Kim, and In-Beum Lee. Discovery of differentially expressed genes related to histological subtype of hepatocellular carcinoma. *Biotechnol Prog.*, 19(3):1011–5, 2003.

[954] Jaewook Lee and Daewon Lee. An improved cluster labeling method for support vector clustering. *IEEE Trans Pattern Anal Mach Intell*, 27(3):461–4, Mar 2005.

[955] Martin M S Lee, S. Sathiya Keerthi, Chong Jin Ong, and Dennis DeCoste. An efficient method for computing leave-one-out error in support vector machines with Gaussian kernels. *IEEE Trans Neural Netw*, 15(3):750–7, May 2004.

[956] Raymond S. T. Lee and James N. K. Liu. An oscillatory elastic graph matching model for recognition of offline handwritten chinese characters. In *KES*, pages 284–287, 1999.

[957] Y. Lee and C.-K. Lee. Classification of multiple cancer types by multicategory support vector machines using gene expression data. *Bioinformatics*, 19(9):1132–1139, 2003.

[958] R. J. Lefkowitz, J.-P. Sun, and A. K. Shukla. A crystal clear view of the beta2-adrenergic receptor. *Nat. Biotechnol.*, 26(2):189–191, Feb 2008.

[959] F. Leighton and R. Rivest. Estimating a probability using finite memory. *IEEE Trans. Inform. Theory*, 32(6):733– 742, Nov 1986.

[960] Claude Lemarechal, Claudia Sagastizábal, Echal, Claudia Sagastiz Abal, and Pii S. Practical aspects of the moreau-yosida regularization: Theoretical preliminaries. *SIAM Journal on Optimization*, 7:367–385, 1997.

[961] C. Lemmen and T. Lengauer. Computational methods for the structural alignment of molecules. *J. Comput. Aided. Mol. Des.*, 14(3):215–232, Mar 2000.

[962] Chenlei Leng, Chenlei Leng, Yi Lin, Yi Lin, Grace Wahba, and Grace Wahba. A note on the lasso and related procedures in model selection. *Statistica Sinica*, 16(4):1273,1284, 2004.

[963] C. Leslie, E. Eskin, and W.S. Noble. The spectrum kernel: a string kernel for SVM protein classification. In Russ B. Altman, A. Keith Dunker, Lawrence Hunter, Kevin Lauerdale, and Teri E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing 2002*, pages 564–575, Singapore, 2002. World Scientific.

[964] C. Leslie, E. Eskin, J. Weston, and W.S. Noble. Mismatch String Kernels for SVM Protein Classification. In Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, editors, *Advances in Neural Information Processing Systems 15*. MIT Press, 2003.

[965] C. Leslie and R. Kuang. Fast string kernels using inexact matching for protein sequences. *J. Mach. Learn. Res.*, 5:1435–1455, 2004.

[966] C. S. Leslie, E. Eskin, A. Cohen, J. Weston, and W. S. Noble. Mismatch string kernels for discriminative protein classification. *Bioinformatics*, 20(4):467–476, 2004.

[967] D. Lett, M. Hsing, and F. Pio. Interaction profile-based protein classification of death domain. *BMC Bioinformatics*, 5(75), 2004.

[968] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *Int. J. Comput. Vision*, 43(1):29–44, 2001.

[969] H. A. Levine. Review of : Solutions of ill posed problems. *Bull. Amer. Math. Soc.*, 1:521–524, 1979.

[970] P. J. L'Heureux, J. Carreau, Y. Bengio, O. Delalleau, and S. Y. Yue. Locally linear embedding for dimensionality reduction in QSAR. *J. Comput. Aided Mol. Des.*, 18(7-9):475–82, 2004.

[971] H. Li, C. Ung, C. Yap, Y. Xue, Z. Li, Z. Cao, and Y. Chen. Prediction of genotoxicity of chemical compounds by statistical learning methods. *Chem. Res. Toxicol.*, 18(6):1071–1080, Jun 2005.

[972] Jinyan Li, Huiqing Liu, James R Downing, Allen Eng-Juh Yeoh, and Limsoon Wong. Simple rules underlying gene expression profiles of more than six subtypes of acute lymphoblastic leukemia (ALL) patients. *Bioinformatics*, 19(1):71–8, Jan 2003.

[973] L. Li, W. Jiang, X. Li, K.L. Moser, Z. Guo, L. Du, Q. Wang, E.J. Topol, Q. Wang, and S. Rao. A robust hybrid between genetic algorithm and support vector machine for extracting an optimal feature gene subset. *Genomics*, 85(1):16–23, 2005.

[974] L. Li, H. Tang, Z. Wu, J. Gong, M. Gruidl, J. Zou, M. Tockman, and R.A. Clark. Data mining techniques for cancer detection using serum proteomic profiling. *Artif. Intell. Med.*, 32(2):71–83, 2004.

[975] Shutao Li, James Tin-Yau Kwok, Ivor Wai-Hung Tsang, and Yaonan Wang. Fusing images with different focuses using support vector machines. *IEEE Trans Neural Netw*, 15(6):1555–61, Nov 2004.

[976] T. Li, C. Zhang, and M. Ogihara. A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics*, 20(15):2429–2437, 2004.

[977] H. Liang and Z. Lin. Detection of delayed gastric emptying from electrogastrograms with support vector machine. *IEEE Trans Biomed Eng*, 48(5):601–4, May 2001.

[978] L. Liao and W. S. Noble. Combining pairwise sequence similarity and support vector machines for remote protein homology detection. In *Proceedings of the Sixth International Conference on Computational Molecular Biology*, 2002.

[979] L. Liao and W.S. Noble. Combining Pairwise Sequence Similarity and Support Vector Machines for Detecting Remote Protein Evolutionary and Structural Relationships. *J. Comput. Biol.*, 10(6):857–868, 2003.

[980] D. A. Liberles, A. Thorén, G. von Heijne, and A. Elofsson. The use of phylogenetic profiles for gene predictions. *Curr. Genom.*, 2002. To appear.

[981] K. Lin, Y. Kuang, J. S. Joseph, and P. R. Kolatkar. Conserved codon composition of ribosomal protein coding genes in Escherichia coli, Mycobacterium tuberculosis and Saccharomyces cerevisiae: lessons from supervised machine learning in functional genomics. *Nucl. Acids Res.*, 30(11):2599–2607, 2002.

[982] S. H. S. Lin and O. Civelli. Orphan G protein-coupled receptors: targets for new therapeutic interventions. *Ann. Med.*, 36(3):204–214, 2004.

[983] Tzu-Chao Lin and Pao-Ta Yu. Adaptive two-pass median filter based on support vector machines for image restoration. *Neural Comput*, 16(2):332–53, Feb 2004.

[984] WuMei Lin, Xin Yuan, Powing Yuen, William I Wei, Jonathan Sham, PengCheng Shi, and Jianan Qu. Classification of in vivo autofluorescence spectra using support vector machines. *J Biomed Opt*, 9(1):180–6, 2004.

[985] Y. Lin. Support vector machines and the Bayes rule in classification. *Data Mining and Knowledge Discovery*, 6(3):259–275, 2002.

[986] P. Lind and T. Maltseva. Support vector machines for the estimation of aqueous solubility. *J Chem Inf Comput Sci*, 43(6):1855–9, 2003.

[987] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug. Deliv. Rev*, 46(1-3):3–26, Mar 2001.

[988] J. Listgarten, S. Damaraju, B. Poulin, L. Cook, J. Dufour, A. Driga, J. Mackey, D. Wishart, R. Greiner, and B. Zanke. Predictive Models for Breast Cancer Susceptibility from Multiple Single Nucleotide Polymorphisms. *Clin. Cancer Res.*, 10(8):2725–2737, 2004.

[989] Chengjun Liu. Gabor-based kernel PCA with fractional power polynomial models for face recognition. *IEEE Trans Pattern Anal Mach Intell*, 26(5):572–81, May 2004.

[990] H. Liu, J. Li, and L. Wong. A Comparative Study on Feature Selection and Classification Methods Using Gene Expression Profiles and Proteomic Patterns. In R. Lathrop, K. Nakai, S. Miyano, T. Takagi, and M. Kanehisa, editors, *Genome Informatics 2002*, volume 12, Tokyo, 2002. Universal Academy Press.

[991] H. X. Liu, C. X. Xue, R. S. Zhang, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan. Quantitative prediction of logk of peptides in high-performance liquid chromatography based on molecular descriptors by using the heuristic method and support vector machine. *J Chem Inf Comput Sci*, 44(6):1979–86, 2004.

[992] H. X. Liu, R. S. Zhang, F. Luan, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan. Diagnosing breast cancer based on support vector machines. *J. Chem. Inf. Comput. Sci.*, 43(3):900–7, 2003.

[993] H. X. Liu, R. S. Zhang, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan. QSAR study of ethyl 2-[(3-methyl-2,5-dioxo(3-pyrrolinyl))amino]-4-(trifluoromethyl) pyrimidine-5-carboxylate: an inhibitor of AP-1 and NF-kappa B mediated gene expression based on support vector machines. *J Chem Inf Comput Sci*, 43(4):1288–96, 2003.

[994] H. X. Liu, R. S. Zhang, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan. Prediction of the isoelectric point of an amino acid based on GA-PLS and SVMs. *J Chem Inf Comput Sci*, 44(1):161–7, 2004.

[995] H. X. Liu, R. S. Zhang, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan. QSAR and classification models of a novel series of COX-2 selective inhibitors: 1,5-diarylimidazoles based on support vector machines. *J Comput Aided Mol Des*, 18(6):389–99, Jun 2004.

[996] Han Liu, John Lafferty, and Larry Wasserman. Nonparametric regression and classification with joint sparsity constraints. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 969–976. MIT Press, 2009.

[997] Huiqing Liu, Hao Han, Jinyan Li, and Limsoon Wong. An in-silico method for prediction of polyadenylation signals in human sequences. *Genome Inform Ser Workshop Genome Inform*, 14:84–93, 2003.

[998] Huiqing Liu, Hao Han, Jinyan Li, and Limsoon Wong. Using amino acid patterns to accurately predict translation initiation sites. *In Silico Biol.*, 4(3):255–69, 2004.

[999] Huiqing Liu, Jinyan Li, and Limsoon Wong. Use of extreme patient samples for outcome prediction from gene expression data. *Bioinformatics*, Jun 2005.

[1000] Jane Jijun Liu, Gene Cutler, Wuxiong Li, Zheng Pan, Sihua Peng, Tim Hoey, Liangbiao Chen, and Xuefeng Bruce Ling. Multiclass cancer classification and biomarker discovery using GA-based algorithms. *Bioinformatics*, 21(11):2691–7, Jun 2005.

[1001] Jian Liu, Shu Zheng, Jie kai Yu, Xue bin Yu, Wei guo Liu, Jian min Zhang, and Xun Hu. [Establishment of diagnostic model of cerebrospinal protein fingerprint pattern for glioma and its clinical application.]. *Zhejiang Da Xue Xue Bao Yi Xue Ban*, 34(2):141–7, Mar 2005.

[1002] Xiaomei Liu, Lawrence O Hall, and Kevin W Bowyer. Comments on "a parallel mixture of SVMs for very large scale problems". *Neural Comput*, 16(7):1345–51, Jul 2004.

[1003] Y. Liu. Active learning with support vector machine applied to gene expression data for cancer classification. *J. Chem. Inf. Comput. Sci.*, 44(6):1936–1941, 2004.

[1004] Y. Liu. A comparative study on feature selection methods for drug discovery. *J Chem Inf Comput Sci*, 44(5):1823–8, 2004.

[1005] Zhenqiu Liu, Dechang Chen, and Halima Bensmail. Gene expression data classification with kernel principal component analysis. *J Biomed Biotechnol*, 2005(2):155–9, 2005.

[1006] M. Llinás, Z. Bozdech, E. D. Wong, A. T. Adai, and J. L. DeRisi. Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains. *Nucleic Acids Res*, 34(4):1166–1173, 2006.

[1007] Siaw Ling Lo, Cong Zhong Cai, Yu Zong Chen, and Maxey C M Chung. Effect of training datasets on support vector machine prediction of protein-protein interactions. *Proteomics*, 5(4):876–84, Mar 2005.

[1008] Miguel Sousa Lobo, Lobo I, Lieyen Vandenberghe, Herv Lebret, and Stephen Boyd. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284:193–228, November 1998.

[1009] H. Lodhi, C. Saunders, J. Shawe-Taylor, N. Cristianini, and C.je n'ai pas vraiment d'lments de rponse. Watkins. Text classification using string kernels. *J. Mach. Learn. Res.*, 2:419–444, 2002.

[1010] H. Lodhi, J. Shawe-Taylor, N. Cristianini, and C. J. C. H. Watkins. Text Classification using String Kernels. In *Adv. Neural Inform. Process. Syst.*, pages 563–569, 2000.

[1011] B. Logan, P. Moreno, B. Suzek, Z. Weng, and S. Kasif. A Study of Remote Homology Detection. Technical Report CRL 2001/05, Compaq Cambridge Research laboratory, June 2001.

[1012] G. Loosli. Simplesvm toolbox. Available at http://asi.insa-rouen.fr/ gloosli/simpleSVM.html, 2006.

[1013] N. López-Bigas and C. A. Ouzounis. Genome-wide identification of genes likely to be involved in human genetic disease. *Nucleic Acids Res.*, 32(10):3108–3114, 2004.

[1014] K. Lounici. Sup-norm convergence rate and sign concentration property of lasso and dantzig estimators. *Electron. J. Statist.*, 2:90–102, 2008.

[1015] Karim Lounici, Massimiliano Pontil, Alexandre B. Tsybakov, and Sara van de Geer. Taking advantage of sparsity in multi-task learning. In *Proceedings of COLT*, 2009.

[1016] C. Lu, T. Van Gestel, J.A. Suykens, S. Van Huffel, I. Vergote, and D. Timmerman. Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines. *Artif. Intell. Med.*, 28(3):281–306, 2003.

[1017] Wei-Zhen Lu and Wen-Jian Wang. Potential assessment of the "support vector machine" method in forecasting ambient air pollutant trends. *Chemosphere*, 59(5):693–701, Apr 2005.

[1018] Y.J. Lu, D. Williamson, R. Wang, B. Summersgill, S. Rodriguez, S. Rogers, K. Pritchard-Jones, C. Campbell, and J. Shipley. Expression profiling targeting chromosomes for tumor classification and prediction of clinical behavior. *Genes Chromosomes Cancer*, 38(3):207–214, 2003.

[1019] Feng Luan, Ruisheng Zhang, Chunyan Zhao, Xiaojun Yao, Mancang Liu, Zhide Hu, and Botao Fan. Classification of the carcinogenicity of N-nitroso compounds based on support vector machines and linear discriminant analysis. *Chem Res Toxicol*, 18(2):198–203, Feb 2005.

[1020] G. Lugosi. On concentration-of-measure inequalities. Seminar notes, 1998.

[1021] G. Lugosi. Concentration-of-measure inequalities. Lecture notes, January 2003.

[1022] G. Lugosi and A. Nobel. Adaptive Model Selection Using Empirical Complexities. *Ann. Stat.*, 27(6):1830–1864, December 1999.

[1023] G. Lugosi and N. Vayatis. On the Bayes-risk consistency of regularized boosting methods. *Ann. Stat.*, 32:30–55, 2004.

[1024] G. Lugosi and K. Zeger. Nonparametric estimation via empirical risk minimization. *IEEE Trans. Inform. Theory*, 41(3):677–687, May 1995.

[1025] G. Lugosi and K. Zeger. Concept learning using complexity regularization. *IEEE Trans. Inform. Theory*, 42(1):48–54, Jan 1996.

[1026] L. Lukas, A. Devos, J. A K Suykens, L. Vanhamme, F. A. Howe, C. Majs, A. Moreno-Torres, M. Van der Graaf, A. R. Tate, C. Ars, and S. Van Huffel. Brain tumor classification based on long echo proton MRS signals. *Artif. Intell. Med.*, 31(1):73–89, May 2004.

[1027] Bin Luo and Edwin R. Hancock. Alignment and correspondence using singular value decomposition. In *Proceedings of the Joint IAPR International Workshops on Advances in Pattern Recognition*, pages 226–235, London, UK, 2000. Springer-Verlag.

[1028] K. Q. Luo and D. C. Chang. The gene-silencing efficiency of siRNA is strongly dependent on the local structure of mRNA at the targeted region. *Biochem. Biophys. Res. Commun.*, 318(1):303–10, May 2004.

[1029] Tong Luo, Kurt Kramer, Dmitry B Goldgof, Lawrence O Hall, Scott Samson, Andrew Remsen, and Thomas Hopkins. Recognizing plankton images from the shadow image particle profiling evaluation recorder. *IEEE Trans Syst Man Cybern B Cybern*, 34(4):1753–62, Aug 2004.

[1030] P. E. Lnning. Breast cancer prognostication and prediction: are we making progress? *Annals of Oncology*, 18:viii3–viii7, 2007.

[1031] J-B. Ma, Y.-R. Yuan, G.. Meister, Y. Pei, T. Tuschl, and Patel D.J. Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus PIWI protein. *Nature*, 434:666–670, 2005.

[1032] J.B. Ma, K. Ye, and D.J. Patel. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature*, 429(6989):318–322, May 2004.

[1033] Z. Ma, S. Gong, H. Richard, D. L. Tucker, T. Conway, and J. W. Foster. GadE (YhiE) activates glutamate decarboxylase-dependent acid resistance in Escherichia coli K-12. *Mol. Microbiol.*, 49(5):1309–1320, Sep 2003.

[1034] E. Maby, R. Le Bouquin Jeanns, C. Ligeois-Chauvel, B. Gourevitch, and G. Faucon. Analysis of auditory evoked potential parameters in the presence of radiofrequency fields using a support vector machines method. *Med Biol Eng Comput*, 42(4):562–8, Jul 2004.

[1035] Roberto F Machado, Daniel Laskowski, Olivia Deffenderfer, Timothy Burch, Shuo Zheng, Peter J Mazzone, Tarek Mekhail, Constance Jennings, James K Stoller, Jacqueline Pyle, Jennifer Duncan, Raed A Dweik, and Serpil C Erzurum. Detection of lung cancer by sensor array analyses of exhaled breath. *Am J Respir Crit Care Med*, 171(11):1286–91, Jun 2005.

[1036] I. G. Maglogiannis and E. P. Zafiropoulos. Characterization of digital medical images utilizing support vector machines. *BMC Med. Informat. Decis. Making*, 4(4), 2004.

[1037] P. Mahé, L. Ralaivola, V. Stoven, and J.-P. Vert. The pharmacophore kernel for virtual screening with support vector machines. Technical Report Technical Report HAL:ccsd-00020066, Ecole des Mines de Paris, march 2006.

[1038] P. Mahé, L. Ralaivola, V. Stoven, and J.-P. Vert. The pharmacophore kernel for virtual screening with support vector machines. *J. Chem. Inf. Model.*, 46(5):2003–2014, 2006.

[1039] P. Mahé, N. Ueda, T. Akutsu, J.-L. Perret, and J.-P. Vert. Extensions of marginalized graph kernels. In R. Greiner and D. Schuurmans, editors, *Proceedings of the Twenty-First International Conference on Machine Learning (ICML 2004)*, pages 552–559. ACM Press, 2004.

[1040] P. Mahé, N. Ueda, T. Akutsu, J.-L. Perret, and J.-P. Vert. Graph kernels for molecular structure-activity relationship analysis with support vector machines. *J. Chem. Inf. Model.*, 45(4):939–51, 2005.

[1041] P. Mahé and J.-P. Vert. Graph kernels based on tree patterns for molecules. Technical Report ccsd-00095488, HAL, September 2006.

[1042] P. Mahé and J. P. Vert. Graph kernels based on tree patterns for molecules. *Mach. Learn.*, 75(1):3–35, 2009.

[1043] W. H. Majoros, L. Pertea, and S. L. Salzberg. Efficient implementation of a generalized pair hidden Markov model for comparative gene finding. *Bioinformatics*, 21(9):1782–1788, May 2005.

[1044] S. K. Majumder, N. Ghosh, and P. K. Gupta. Support vector machine for optical diagnosis of cancer. *J Biomed Opt*, 10(2):024034, 2005.

[1045] Shovan K Majumder, Nirmalya Ghosh, and Pradeep K Gupta. Relevance vector machine for optical diagnosis of cancer. *Lasers Surg Med*, 36(4):323–33, Apr 2005.

[1046] S. G. Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397–3415, 1993.

[1047] C. L. Mallows. Some comments on c_p. *Technometrics*, 15:661–675, 1973.

[1048] H. Mamitsuka. Predicting peptides that bind to MHC molecules using supervised learning of hidden Markov models. *Proteins*, 33(4):460–474, Dec 1998.

[1049] E. Mammen and A. Tsybakov. Smooth discrimination analysis. *Ann. Stat.*, 27(6):1808–1829, 1999.

[1050] M.Z. Man, G. Dyson, K. Johnson, and B. Liao. Evaluating methods for classifying expression data. *J. Biopharm. Stat.*, 14(4):1065–1084, 2004.

[1051] D.T. Manallack and D.J. Livingstone. Neural networks in drug-discovery: have they lived up with their promise? *Eur. J. Med. Chem.*, 34:195–208, 1999.

[1052] C. Manly, S. Louise-May, and J. Hammer. The impact of informatics and computational chemistry on synthesis and screening. *Drug Discov. Today*, 6(21):1101–1110, Nov 2001.

[1053] K. Z. Mao. Feature subset selection for support vector machines through discriminative function pruning analysis. *IEEE Trans Syst Man Cybern B Cybern*, 34(1):60–7, Feb 2004.

[1054] Yong Mao, Xiaobo Zhou, Daoying Pi, Youxian Sun, and Stephen T C Wong. Multiclass cancer classification by using fuzzy support vector machine and binary decision tree with gene selection. *J Biomed Biotechnol*, 2005(2):160–71, 2005.

[1055] I. Marchal, G. Golfier, O. Dugas, and M. Majed. Bioinformatics in glycobiology. *Biochimie*, 85(1-2):75–81, 2003.

[1056] E. M. Marcotte, M. Pellegrini, M. J. Thompson, T. O. Yeates, and D. Eisenberg. A combined algorithm for genome-wide prediction of protein function. *Nature*, 402:83–86, November 1999.

[1057] E.M. Marcotte, M. Pellegrini, H.-L. Ng, D.W. Rice, T.O. Yeates, and D. Eisenberg. Detecting Protein Function and Protein-Protein Interactions from Genome Sequences. *Science*, 285:751–753, 1999.

[1058] Elaine R. Mardis. The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, 24:133–141, 2008.

[1059] Elaine R Mardis. Next-generation dna sequencing methods. *Annu Rev Genomics Hum Genet*, 9:387–402, 2008.

[1060] A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. Dalla Favera, and A. Califano. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, 7 Suppl 1:S7, 2006.

[1061] F. Markowetz, L. Edler, and M. Vingron. Support Vector Machines for Protein Fold Class Prediction. *Biometrical Journal*, 45(3):377–389, 2003.

[1062] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, March 1952.

[1063] Stephen Marsland, Jonathan Shapiro, and Ulrich Nehmzow. A self-organising network that grows when required. *Neural Netw*, 15(8-9):1041–58, 2002.

[1064] S. Martin, D. Roe, and J.-L. Faulon. Predicting protein-protein interactions using signature products. *Bioinformatics*, 21(2):218–226, Jan 2005.

[1065] T. C. Martin, J. Moecks, A. Belooussov, S. Cawthraw, B. Dolenko, M. Eiden, J. Von Frese, W. Kohler, J. Schmitt, R. Somorjai, T. Udelhoven, S. Verzakov, and W. Petrich. Classification of signatures of Bovine Spongiform Encephalopathy in serum using infrared spectroscopy. *Analyst*, 129(10):897–901, Oct 2004.

[1066] Y. C. Martin. A bioavailability score. *J. Med. Chem.*, 48(9):3164–3170, May 2005.

[1067] Y. C. Martin, J. B. Holland, C. H. Jarboe, and N. Plotnikoff. Discriminant analysis of the relationship between physical properties and the inhibition of monoamine oxidase by aminotetralins and aminoindans. *J Med Chem*, 17(4):409–413, Apr 1974.

[1068] Ann-Marie Martoglio, James W Miskin, Stephen K Smith, and David J C MacKay. A decomposition model to track gene expression signatures: preview on observer-independent classification of ovarian cancer. *Bioinformatics*, 18(12):1617–24, Dec 2002.

[1069] S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks. *Science*, 296:910–913, 2002.

[1070] J. S. Mason, I. Morize, P. R. Menard, D. L. Cheney, C. Hulme, and R. F. Labaudiniere. New 4-point pharmacophore method for molecular similarity and diversity applications: overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J Med Chem*, 42(17):3251–3264, Aug 1999.

[1071] P. Massart. Some applications of concentration inequalities to statistics. *Ann. Fac. Sc. Toulouse*, IX(2):245–303, 2000.

[1072] M. T. Matache and V. Matache. Hilbert spaces induced by Toeplitz covariance kernels. In *Lecture Notes in Control and Information Sciences*, volume 280, pages 319–334. Springer, Jan 2002.

[1073] Alvaro Mateos, Joaqun Dopazo, Ronald Jansen, Yuhai Tu, Mark Gerstein, and Gustavo Stolovitzky. Systematic learning of gene functional classes from DNA array expression data by using multilayer perceptrons. *Genome Res.*, 12(11):1703–15, Nov 2002.

[1074] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, 288(5):911–40, May 1999.

[1075] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*, 288(5):911–940, May 1999.

[1076] A. Matsuda, J.-P. Vert, H. Saigo, N. Ueda, H. Toh, and T. Akutsu. A novel representation of protein sequences for prediction of subcellular location using support vector machines. *Protein Sci.*, 14(11):2804–2813, 2005.

[1077] H. Matter and T. Pötter. Comparing 3D pharmacophore triplets and 2D fingerprints for selecting diverse compound subsets. *J. Chem. Inf. Comput. Sci.*, 39(6):1211–1225, 1999.

[1078] T. Mattfeldt, H.W. Gottfried, H. Wolter, V. Schmidt, H.A. Kestler, and J. Mayer. Classification of prostatic carcinoma with artificial neural networks using comparative genomic hybridization and quantitative stereological data. *Pathol. Res. Pract.*, 199(12):773–784, 2003.

[1079] T. Mattfeldt, H. A. Kestler, and H. P. Sinn. Prediction of the axillary lymph node status in mammary cancer on the basis of clinicopathological data and flow cytometry. *Med Biol Eng Comput*, 42(6):733–9, Nov 2004.

[1080] Torsten Mattfeldt, Danilo Trijic, Hans-Werner Gottfried, and Hans A Kestler. Classification of incidental carcinoma of the prostate using learning vector quantization and support vector machines. *Cell Oncol*, 26(1-2):45–55, 2004.

[1081] John S. Mattick. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *BioEssays*, 25:930–939, 2003.

[1082] John S. Mattick and Igor V. Makunin. Non-coding RNA. *Human Molecular Genetics*, 15:R17–R29, 2006.

[1083] Michael Mavroforakis, Harris Georgiou, Nikos Dimitropoulos, Dionisis Cavouras, and Sergios Theodoridis. Significance analysis of qualitative mammographic features, using linear classifiers, neural networks and support vector machines. *Eur J Radiol*, 54(1):80–9, Apr 2005.

[1084] Torsten Mayr, Christian Igel, Gregor Liebsch, Ingo Klimant, and Otto S Wolfbeis. Cross-reactive metal ion sensor array in a micro titer plate format. *Anal Chem*, 75(17):4389–96, Sep 2003.

[1085] J. D. McAuliffe, L. Pachter, and M. I. Jordan. Multiple-sequence functional annotation and the generalized hidden Markov phylogeny. *Bioinformatics*, 20(12):1850–1860, Aug 2004.

[1086] Mark R. Mcgann, Harold R. Almond, Anthony Nicholls, Andrew J. Grant, and Frank K. Brown. Gaussian docking functions. *Biopolymers*, 68(1):76–90, 2003.

[1087] L. F. McGinnis. Implementation and testing of a primal-dual algorithm for the assignment problem. *Operations Research*, 31(2):277–291, 1983.

[1088] M. J. McGregor and S. M. Muskal. Pharmacophore fingerprinting. 1. Application to QSAR and focused library design. *J Chem Inf Comput Sci*, 39(3):569–574, 1999.

[1089] M.J. McGregor and V. Pallai. Clustering of Large Databases of Compounds: Using the mdl "Keys" as Structural Descriptors. *J Chem Inf Comput Sci*, 37:443–448, 1997.

[1090] Kevin Judd McKernan, Heather E Peckham, Gina L Costa, Stephen F McLaughlin, Yutao Fu, Eric F Tsung, Christopher R Clouser, Cisyla Duncan, Jeffrey K Ichikawa, Clarence C Lee, Zheng Zhang, Swati S Ranade, Eileen T Dimalanta, Fiona C Hyland, Tanya D Sokolsky, Lei Zhang, Andrew Sheridan, Haoning Fu, Cynthia L Hendrickson, Bin Li, Lev Kotler, Jeremy R Stuart, Joel A Malek, Jonathan M Manning, Alena A Antipova, Damon S Perez, Michael P Moore, Kathleen C Hayashibara, Michael R Lyons, Robert E Beaudoin, Brittany E Coleman, Michael W Laptewicz, Adam E Sannicandro, Michael D Rhodes, Rajesh K Gottimukkala, Shan Yang, Vineet Bafna, Ali Bashir, Andrew MacBride, Can Alkan, Jeffrey M Kidd, Evan E Eichler, Martin G Reese, Francisco M De La Vega, and Alan P Blanchard. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res*, 19(9):1527–1541, Sep 2009.

[1091] Larry McKnight and Padmini Srinivasan. Categorization of sentence types in medical abstracts. *AMIA Annu Symp Proc*, pages 440–4, 2003.

[1092] M. T. McManus and P. A. Sharp. Gene silencing in mammals by small interfering RNAs. *Nat. Rev. Genet.*, 3(10):737–747, Oct 2002.

[1093] A. McMichael and T. Hanke. The quest for an AIDS vaccine: is the CD8+ T-cell approach feasible? *Nat. Rev. Immunol.*, 2(4):283–291, Apr 2002.

[1094] Dong mei Qin, Zhan yi Hu, and Yong heng Zhao. Automated classification of celestial spectra based on support vector machines. *Guang Pu Xue Yu Guang Pu Fen Xi*, 24(4):507–11, Apr 2004.

[1095] L. Meier, S. van de Geer, and P. Bühlmann. The group lasso for logistic regression. *J. R. Stat. Soc. Ser. B*, 70(1):53–71, 2008.

[1096] P. Meinicke, M. Tech, B. Morgenstern, and R. Merkl. Oligo kernels for datamining on biological sequences: a case study on prokaryotic translation initiation sites. *BMC Bioinformatics*, 5(169), 2004.

[1097] N. Meinshausen and P. Bühlmann. High dimensional graphs and variable selection with the lasso. *Ann. Stat.*, 34:1436–1462, 2006.

[1098] N. Meinshausen, G. Rocha, and B. Yu. Discussion: A tale of three cousins: Lasso, l2boosting and dantzig. *ANNALS OF STATISTICS*, 35:2373, 2007.

[1099] Nicolai Meinshausen and Peter Buehlmann. Stability selection, May 2009.

[1100] S.I. Meireles, A.F. Carvalho, R. Hirata, A.L. Montagnini, W.K. Martins, F.B. Runza, B.S. Stolf, L. Termini, C.E. Neto, R.L. Silva, F.A. Soares, E.J. Neves, and L.F. Reis. Differentially expressed genes in gastric tumors identified by cDNA array. *Cancer Lett.*, 190(2):199–211, Feb 2003.

[1101] G. Meister and T. Tuschl. Mechanisms of gene silencing by double-stranded RNA. *Nature*, 431(7006):343–9, Sep 2004.

[1102] Craig C. Mello and Darryl Conte Jr. Revealing the world of RNA interference. *Nature*, 43:338–342, 2004.

[1103] S. Menchetti, F. Costa, and P. Frasconi. Weighted decomposition kernels. In L. De Raedt and S. Wrobel, editors, *Proceedings of the Twenty-Second International Conference on Machine Learning (ICML 2005)*, pages 585–592. ACM Press, 2005.

[1104] S. Mendelson. Geometric parameters in Learning Theory. Lecture notes, 2003.

[1105] O. Mercereau-Puijalon. Malaria research in the post-genomic era. *J. Soc. Biol.*, 198(3):193–197, 2004.

[1106] G. Mercier, N. Berthault, J. Mary, J. Peyre, A. Antoniadis, J.-P. Comet, A. Cornuejols, C. Froidevaux, and M. Dutreix. Biological detection of low radiation doses by combining results of two microarray analysis methods. *Nucleic Acids Res.*, 32(1):e12, 2004.

[1107] G. Mercier, Y. Denis, P. Marc, L. Picard, and M. Dutreix. Transcriptional induction of repair genes during slowing of replication in irradiated Saccharomyces cerevisiae. *Mutat. Res.*, 487(3-4):157–172, Dec 2001.

[1108] N. Merhav and M. Feder. Universal schemes for sequential decision from individual data sequences. *IEEE Trans. Inform. Theory*, 39(4):1280–1292, Jul 1993.

[1109] N. Merhav and M. Feder. A strong version of the redundancy-capacity theorem of universal. *IEEE Trans. Inform. Theory*, 41(3):714–722, May 1995.

[1110] N. Merhav and M. Feder. Universal prediction. *IEEE Trans. Inform. Theory*, 44(6):2124–2147, Oct 1998.

[1111] N. Merhav, M. Feder, and M. Gutman. Some properties of sequential predictors for binary Markov sources. *IEEE Trans. Inform. Theory*, 39(3):887–892, May 1993.

[1112] Christian Merkwirth, Harald Mauser, Tanja Schulz-Gasch, Olivier Roche, Martin Stahl, and Thomas Lengauer. Ensemble methods for classification in cheminformatics. *J Chem Inf Comput Sci*, 44(6):1971–8, 2004.

[1113] E. Meron and M. Feder. Finite-memory universal prediction of individual sequences. *IEEE Trans. Inform. Theory*, 50(7):1506–1523, Jul 2004.

[1114] Jordi Mestres. Computational chemogenomics approaches to systematic knowledge-based drug discovery. *Curr Opin Drug Discov Devel*, 7(3):304–313, May 2004.

[1115] H.W. Mewes, D. Frishman, U. Güldener, G. Mannhaupt, K. Mayer, M. Mokrejs, B. Morgenstern, M. Münsterkoetter, S. Rudd, and B. Weil. MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.*, 30(1):31–34, 2002.

[1116] Charles A Micchelli and Massimiliano Pontil. On learning vector-valued functions. *Neural Comput*, 17(1):177–204, Jan 2005.

[1117] M. Middendorf, E. Ziv, C. Adams, J. Hom, R. Koytcheff, C. Levovitz, G. Woods, L. Chen, and C. Wiggins. Discriminative topological features reveal biological network mechanisms. *BMC Bioinformatics*, 5(181), 2004.

[1118] S. Miertus, G. Fassina, and P.F. Seneci. Concepts of Combinatorial Chemistry and Combinatorial Technologies. *Chemické Listy*, 94:1104–1110, 2000.

[1119] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.R. Müller. Fisher discriminant analysis with kernels. In Y.-H. Hu, J. Larsen, E. Wilson, and S. Douglas, editors, *Neural Networks for Signal Processing IX*, pages 41–48. IEEE, 1999.

[1120] Sven Mika and Burkhard Rost. NLProt: extracting protein names and sequences from papers. *Nucleic Acids Res*, 32(Web Server issue):W634–7, Jul 2004.

[1121] Sven Mika and Burkhard Rost. Protein names precisely peeled off free text. *Bioinformatics*, 20(Suppl. 1):i241–i247, 2004.

[1122] M. Milik, D. Sauer, A. P. Brunmark, L. Yuan, A. Vitiello, M. R. Jackson, P. A. Peterson, J. Skolnick, and C. A. Glass. Application of an artificial neural network to predict specific class I MHC binding peptide sequences. *Nat. Biotechnol.*, 16(8):753–756, Aug 1998.

[1123] J.W. Miller, R. Goodman, and P. Smyth. On loss functions which minimize to conditional expected values and posterior probabilities. *IEEE Trans. Inform. Theory*, 39(4):1404–1408, Jul 1993.

[1124] J.W. Milnor. *Topology from the Differentiable Viewpoint.* Univ. Press of Virginia, 1969.

[1125] T. Mirzadegan, G. Benkö, S. Filipek, and K. Palczewski. Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin. *Biochemistry*, 42(10):2759–2767, Mar 2003.

[1126] F. Mitelman, B. Johansson, and F. Mertens. The impact of translocations and gene fusions on cancer causation. *Nature Reviews Cancer*, 7:233–245, 2007.

[1127] M. A. Miteva, W. H. Lee, M. O. Montes, and B. O. Villoutreix. Fast structure-based virtual ligand screening combining FRED, DOCK, and Surflex. *J Med Chem*, 48(19):6012–6022, Sep 2005.

[1128] Pabitra Mitra, C. A. Murthy, and Sankar K Pal. A probabilistic active support vector learning algorithm. *IEEE Trans Pattern Anal Mach Intell*, 26(3):413–8, Mar 2004.

[1129] Tomohiro Mitsumori, Sevrani Fation, Masaki Murata, Kouichi Doi, and Hirohumi Doi. Gene/protein name recognition based on support vector machine using dictionary as features. *BMC Bioinformatics*, 6 Suppl 1:S8, 2005.

[1130] V. Mittal. Improving the efficiency of RNA interference in mammals. *Nat. Rev. Genet.*, 5(5):355–65, May 2004.

[1131] F. Miwakeichi, R. Ramirez-Padron, P. A. Valdes-Sosa, and T. Ozaki. A comparison of non-linear non-parametric models for epilepsy data. *Comput. Biol. Med.*, 31(1):41–57, Jan 2001.

[1132] F. Model, P. Adorjan, A. Olek, and C. Piepenbrock. Feature selection for DNA methylation based cancer classification. *Bioinformatics*, 17(Supp. 1):S157–S164, 2001.

[1133] D.S. Modha and E. Masry. Minimum complexity regression estimation with weakly dependent observations. *IEEE Trans. Inform. Theory*, 42(6):2133–2145, Nov 1996.

[1134] D.S. Modha and E. Masry. Memory-universal prediction of stationary random processes. *IEEE Trans. Inform. Theory*, 44(1):117–133, Jan 1998.

[1135] S. S. Mohamed, M. M A Salama, M. Kamel, E. F. El-Saadany, K. Rizkalla, and J. Chin. Prostate cancer multi-feature analysis using trans-rectal ultrasound images. *Phys Med Biol*, 50(15):N175–85, Aug 2005.

[1136] B. Mohar. The Laplacian spectrum of graphs. In Y. Alavi, G. Chartrand, O. Ollermann, and A. Schwenk, editors, *Graph theory, combinatorics, and applications*, pages 871–898, New-York, 1991. John Wiley and Sons, Inc.

[1137] B. Mohar. Some applications of Laplace eigenvalues of graphs. In G. Hahn and G. Sabidussi, editors, *Graph Symmetry: Algebraic Methods and Applications*, volume 497 of *NATO ASI Series C*, pages 227–275. Kluwer, Dordrecht, 1997.

[1138] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, and C. R. Corbeil. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br J Pharmacol*, 153 Suppl 1:S7–26, Mar 2008.

[1139] E. J. Moler, M. L. Chow, and I. S. Mian. Analysis of molecular profile data using generative and discriminative methods. *Physiol. Genomics*, 4(2):109–126, Dec 2000.

[1140] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Struct. Algorithm.*, 6:161–179, 1995.

[1141] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combinator. Probab. Comput.*, 7:295–305, 1998.

[1142] F. Mordelet and J.-P. Vert. Sirene: Supervised inference of regulatory networks. *Bioinformatics*, 24(16):i76–i82, 2008.

[1143] G. Moreau and P. Broto. Autocorrelation of molecular structures: Application to SAR studies. *Nouv. J. Chim.*, 757:764, 1980.

[1144] H.L. Morgan. The Generation of Unique Machine Description for Chemical Structures - A Technique Developed at Chemical Abstracts Service. *J Chem Doc*, 5:107–113, 1965.

[1145] R. J. Morris, R.J. Najmanovich, A. Kahraman, and J.M. Thornton. Real spherical harmonic expansion coefficients as 3d shape descriptors for protein binding pocket and ligand comparisons. *Bioinformatics*, 21(10):2347–2355, May 2005.

[1146] A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, 5(7):621–628, Jul 2008.

[1147] G. Morvai, S.J. Yakowitz, and P. Algoet. Weakly convergent nonparametric forecasting of stationary time series. *IEEE Trans. Inform. Theory*, 43(2):483–498, Mar 1997.

[1148] S. Mukherjee, P. Tamayo, J. P. Mesirov, D. Slonim, A. Verri, and T. Poggio. Support vector machine classification of microarray data. Technical Report 182, C.B.L.C., 1998. A.I. Memo 1677.

[1149] K.-R. Müller, G. Rätsch, S. Sonnenburg, S. Mika, M. Grimm, and N. Heinrich. Classifying 'drug-likeness' with Kernel-based learning methods. *J Chem Inf Model*, 45(2):249–53, 2005.

[1150] K. Murphy and S. Mian. Modelling gene expression data using dynamic Bayesian networks. Technical report, Computer Science Division, University of California, Berkeley, CA., 1999.

[1151] Kevin Murphy, Antonio Torralba, and William T.F. Freeman. Using the forest to see the trees: a graphical model relating features, objects and scenes. In *Adv. Neural Inform. Process. Syst.*, Vancouver, BC, 2003. MIT Press.

[1152] A. G. Murzin, S. E. Brenner, T. Hubbard, and C. Chothia. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, 247:536–540, 1995.

[1153] E. Myasnikova, A. Samsonova, M. Samsonova, and J. Reinitz. Support vector regression applied to the determination of the developmental age of a Drosophila embryo from its segmentation gene expression patterns. *Bioinformatics*, 18(Suppl. 1):S87–S95, 2002.

[1154] Rajesh Nair and Burkhard Rost. Mimicking cellular sorting improves prediction of subcellular localization. *J Mol Biol*, 348(1):85–100, Apr 2005.

[1155] R. Najmanovich, N. Kurbatova, and J. Thornton. Detection of 3d atomic similarities and their use in the discrimination of small molecule protein-binding sites. *Bioinformatics*, 24(16):i105–i111, Aug 2008.

[1156] Yusuke Nakamura. ATM: the p53 booster. *Nature Medicine*, 4:1231–1232, 1998.

[1157] A. Nakaya, S. Goto, and M. Kanehisa. Extraction of correlated gene clusters by multiple graph comparison. In *Genome Informatics 2001*, pages 44–53. Universal Academy Press, Tokyo, Japan, 2001.

[1158] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24(2):227–234, 1995.

[1159] Georges Natsoulis, Laurent El Ghaoui, Gert R G Lanckriet, Alexander M Tolley, Fabrice Leroy, Shane Dunlea, Barrett P Eynon, Cecelia I Pearson, Stuart Tugendreich, and Kurt Jarnagin. Classification of a large microarray data set: algorithm comparison and analysis of drug signatures. *Genome Res.*, 15(5):724–36, May 2005.

[1160] N.K. Natt, H. Kaur, and G.P. Raghava. Prediction of transmembrane regions of beta-barrel proteins using ANN- and SVM-based methods. *Proteins*, 56(1):11–18, 2004.

[1161] Tim W Nattkemper, Bert Arnrich, Oliver Lichte, Wiebke Timm, Andreas Degenhard, Linda Pointon, Carmel Hayes, Martin O Leach, and The UK MARIBS Breast Screening Study. Evaluation of radiological features for breast tumour classification in clinical screening with machine learning methods. *Artif. Intell. Med.*, 34(2):129–39, Jun 2005.

[1162] Nature Publishing Group. DNA Technologies - Milestones timeline. *Nature Milestones*, 2007. http://www.nature.com/milestones/miledna/timeline.html.

[1163] S.B. Needleman and C.D. Wunsch. A general method applicable to the search for similarities in the amino acid sequences of two proteins. *J. Mol. Biol.*, 48:443–453, 1970.

[1164] Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 455(7216):1061–1068, Oct 2008.

[1165] M. Neuhaus and H. Bunke. Edit distance-based kernel functions for structural pattern classification. *Pattern Recognition*, 39(10):1852–1863, Oct 2006.

[1166] M. Neuhaus and H. Bunke. *Bridging the Gap Between Graph Edit Distance and Kernel Machines.* World Scientific, September 2007.

[1167] Michel Neuhaus, Kaspar Riesen, and Horst Bunke. Fast suboptimal algorithms for the computation of graph edit distance. In Dit-Yan Yeung, James T. Kwok, Ana L. N. Fred, Fabio Roli, and Dick de Ridder, editors, *SSPR/SPR*, volume 4109 of *Lecture Notes in Computer Science*, pages 163–172. Springer, 2006.

[1168] M. E. J. Newman. Random graphs as models of networks. In S. Bornholdt and H. G. Schuster, editors, *Handbook of Graphs and Networks.* Wiley-VCH, Berlin, 2002. To appear.

[1169] Mark E. J. Newman, Steven H. Strogatz, and Duncan J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64:26118, 2001.

[1170] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems 14*, pages 849–856. MIT Press, 2001.

[1171] M. N. Nguyen and J. C. Rajapakse. Two-stage multi-class support vector machines to protein secondary structure prediction. *Pac Symp Biocomput*, pages 346–57, 2005.

[1172] Minh N Nguyen and Jagath C Rajapakse. Multi-class support vector machines for protein secondary structure prediction. *Genome Inform Ser Workshop Genome Inform*, 14:218–27, 2003.

[1173] Minh N Nguyen and Jagath C Rajapakse. Prediction of protein relative solvent accessibility with a two-stage SVM approach. *Proteins*, 59(1):30–7, Apr 2005.

[1174] A. Nicholls. Oechem, version 1.3.4, openeye scientific software. website, 2005.

[1175] H. Nielsen, J. Engelbrecht, S. Brunak, and G. von Heijne. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.*, 10(1):1–6, 1997.

[1176] M. Nielsen, C. Lundegaard, P. Worning, C. S. Hvid, K. Lamberth, S. Buus, S. Brunak, and O. Lund. Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. *Bioinformatics*, 20(9):1388–1397, Jun 2004.

[1177] M. Nielsen, C. Lundegaard, P. Worning, S. L. Lauemøller, K. Lamberth, S. Buus, S. Brunak, and O. Lund. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.*, 12(5):1007–1017, May 2003.

[1178] J. Nilsson, T. Fioretos, M. Hglund, and M. Fontes. Approximate geodesic distances reveal biologically relevant structures in microarray data. *Bioinformatics*, 20(6):874–80, Apr 2004.

[1179] W. S. Noble. Support vector machine applications in computational biology. In B. Schlkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 71–92. MIT Press, 2004.

[1180] G. Obozinski, B. Taskar, and M. Jordan. Joint covariate selection and joint subspace selection for multiple classification problems. *Statistics and Computing*, 2009. To appear.

[1181] G. Obozinski, B. Taskar, and M. I. Jordan. Multi-task feature selection. Technical report, arXiv, 2007.

[1182] G. Obozinski, M. J. Wainwright, and M. I. Jordan. Union support recovery in high-dimensional multivariate regression. Technical Report 0808.0711v1, arXiv, August 2008.

[1183] Rebekah K O'Donnell, Michael Kupferman, S. Jack Wei, Sunil Singhal, Randal Weber, Bert O'Malley, Yi Cheng, Mary Putt, Michael Feldman, Barry Ziober, and Ruth J Muschel. Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. *Oncogene*, 24(7):1244–51, Feb 2005.

[1184] Cath O'Driscoll. A Virtual Space Odyssey. Nature Horizon : Charting Chemical Space, 2004.

[1185] R. A. O'Flanagan, G. Paillard, R. Lavery, and A. M. Sengupta. Non-additivity in protein-DNA binding. *Bioinformatics*, 21(10):2254–63, May 2005.

[1186] Nobuo Ogawa, Joseph DeRisi, and Patrick O. Brown. New Components of a System for Phosphate Accumulation and Polyphosphate Metabolism in Saccharomyces cerevisiae Revealed by Genomic Expression Analysis. *Mol. Biol. Cell*, 11:4309–4321, Dec 2000.

[1187] R. C. O'Hagan, C. W. Brennan, A. Strahs, X. Zhang, K. Kannan, M. Donovan, C. Cauwels, N. E. Sharpless, W. H. Wong, and L. Chin. Array comparative genome hybridization for tumor classification and gene discovery in mouse models of malignant melanoma. *Cancer Res.*, 63(17):5352–5356, Sep 2003.

[1188] T. Okada, M. Sugihara, A.-N. Bondar, M. Elstner, P. Entel, and V. Buss. The retinal conformation and its environment in rhodopsin in light of a new 2.2 a crystal structure. *J. Mol. Biol.*, 342(2):571–583, Sep 2004.

[1189] S. Okamoto, Y. Yamanishi, S. Ehira, S. Kawashima, K. Tonomura, and M. Kanehisa. Prediction of nitrogen metabolism-related genes in anabaena by kernel-based network analysis. *Proteomics*, 7(6):900–909, Mar 2007.

[1190] Y. Okuno, A. Tamon, H. Yabuuchi, S. Niijima, Y. Minowa, K. Tonomura, R. Kunimoto, and C. Feng. GLIDA: GPCR ligand database for chemical genomics drug discovery database and tools update. *Nucleic Acids Res.*, 36(Database issue):D907–D912, Nov 2007.

[1191] Y. Okuno, J. Yang, K. Taneishi, H. Yabuuchi, and G. Tsujimoto. GLIDA: GPCR-ligand database for chemical genomic drug discovery. *Nucleic Acids Res.*, 34(Database issue):D673–D677, Jan 2006.

[1192] Scott Oloff, Shuxing Zhang, Nagamani Sukumar, Curt Breneman, and Alexander Tropsha. Chemometric analysis of ligand receptor complementarity: identifying complementary ligands based on receptor information (colibri). *J Chem Inf Model*, 46(2):844–851, 2006.

[1193] Byron P Olson, Jennie Si, Jing Hu, and Jiping He. Closed-loop cortical control of direction using support vector machines. *IEEE Trans Neural Syst Rehabil Eng*, 13(1):72–80, Mar 2005.

[1194] M. Opper and R. Urbanczik. Universal learning curves of support vector machines. *Phys Rev Lett*, 86(19):4410–3, May 2001.

[1195] M. Opper and O. Winther. Gaussian processes for classification: mean-field algorithms. *Neural Comput*, 12(11):2655–84, Nov 2000.

[1196] Celestino Ordez, Javier Taboada, Fernando Bastante, Jose Mara Matas, and Angel Manuel Felicsimo. Learning machines applied to potential forest distribution. *Environ Manage*, 35(1):109–20, Jan 2005.

[1197] D.S. Ornstein and B. Weiss. Entropy and data compression schemes. *IEEE Trans. Inform. Theory*, 39(1):78–83, Jan 1993.

[1198] M. R. Osborne, Brett Presnell, and B.A. Turlach. A new approach to variable selection in least squares problems. *IMA Journal of Numerical Analysis*, 20:389–404, 1999.

[1199] Michael R. Osborne, Brett Presnell, and Berwin A. Turlach. On the lasso and its dual. *Journal of Computational and Graphical Statistics*, 9:319–337, 1999.

[1200] Stanislaw Osowski, Linh Tran Hoai, and Tomasz Markiewicz. Support vector machine-based expert system for reliable heartbeat recognition. *IEEE Trans Biomed Eng*, 51(4):582–9, Apr 2004.

[1201] R. Overbeek, N. Larsen, G. D. Pusch, M. D'Souza, E. Jr. Selkov, N. Kyrpides, M. Fonstein, N. Maltsev, and E. Selkov. WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Res.*, 28:123–125, 2000.

[1202] M. Overhoff, M. Alken, R. K. Far, M. Lemaitre, B. Lebleu, G. Sczakiel, and I. Robbins. Local RNA target structure influences siRNA efficacy: a systematic global analysis. *J. Mol. Biol.*, 348(4):871–881, May 2005.

[1203] Yen-Jen Oyang, Shien-Ching Hwang, Yu-Yen Ou, Chien-Yu Chen, and Zhi-Wei Chen. Data classification with radial basis function networks based on a novel kernel density estimation algorithm. *IEEE Trans Neural Netw*, 16(1):225–36, Jan 2005.

[1204] Tapio Pahikkala, Filip Ginter, Jorma Boberg, Jouni Jarvinen, and Tapio Salakoski. Contextual weighting for Support Vector Machines in literature mining: an application to gene versus protein name disambiguation. *BMC Bioinformatics*, 6(1):157, Jun 2005.

[1205] S. I. Pai, Y.-Y. Lin, B. Macaes, A. Meneshian, HungC.-F., and T.-C. Wu. Prospects of RNA interference therapy for cancer. *Gene Ther.*, 13(6):464–477, Mar 2006.

[1206] Gonzalo Pajares and Jess M de la Cruz. On combining support vector machines and simulated annealing in stereovision matching. *IEEE Trans Syst Man Cybern B Cybern*, 34(4):1646–57, Aug 2004.

[1207] Gregory M Palmer, Changfang Zhu, Tara M Breslin, Fushen Xu, Kennedy W Gilchrist, and Nirmala Ramanujam. Comparison of multiexcitation fluorescence and diffuse reflectance spectroscopy for the diagnosis of breast cancer (March 2003). *IEEE Trans Biomed Eng*, 50(11):1233–42, Nov 2003.

[1208] Fei Pan, Baoying Wang, Xin Hu, and William Perrizo. Comprehensive vertical sample-based KNN/LSVM classification for gene expression analysis. *J Biomed Inform*, 37(4):240–8, Aug 2004.

[1209] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. Technical Report HKUST-CS08-08, Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, China, November 2008.

[1210] A. Pandey and M. Mann. Proteomics to study genes and genomes. *Nature*, 405:837–846, 2000.

[1211] Shaoning Pang, Daijin Kim, and Sung Yang Bang. Face membership authentication using SVM classification tree generated by membership-based LLE data partition. *IEEE Trans Neural Netw*, 16(2):436–46, Mar 2005.

[1212] Panagiotis Papadakis, Ioannis Pratikakis, Stavros Perantonis, and Theoharis Theoharis. Efficient 3d shape matching and retrieval using a concrete radialized spherical projection representation. *Pattern Recogn.*, 40(9):2437–2452, 2007.

[1213] A. Papadopoulos, D. I. Fotiadis, and A. Likas. Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines. *Artif. Intell. Med.*, 34(2):141–50, Jun 2005.

[1214] P.M. Pardalos, L. S. Pitsoulis, and M. G. C. Resende. A parallel grasp implementation for the quadratic assignment problem. In *Parallel Algorithms for Irregularly Structured Problems Irregular94*, pages 111–130. Kluwer Academic Publishers, 1995.

[1215] Peter Parham. *The Immune System*. Garland Science Publishing, 2004.

[1216] K.-J. Park and M. Kanehisa. Prediction of protein subcellular locations by support vector machines using compositions of amino acids and amino acid pairs. *Bioinformatics*, 19(13):1656–1663, 2003.

[1217] K. C. Parker, M. A. Bednarek, and J. E. Coligan. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.*, 152(1):163–175, Jan 1994.

[1218] A. Passerini and P. Frasconi. Learning to discriminate between ligand-bound and disulfide-bound cysteines. *Protein Eng. Des. Sel.*, 17(4):367–373, 2004.

[1219] Andrea Passerini, Massimiliano Pontil, and Paolo Frasconi. New results on error correcting output codes of kernel machines. *IEEE Trans Neural Netw*, 15(1):45–54, Jan 2004.

[1220] R. Pastor-Satorras, E. D. Smith, and R. V. Solé. Evolving protein interaction networks through gene duplication. Technical report, Santa Fe Institute, 2002. Working paper 02-02-008.

[1221] D.J. Patterson, K. Yasuhara, and W.L. Ruzzo. Pre-mRNA secondary structure prediction aids splice site prediction. In Russ B. Altman, A. Keith Dunker, Lawrence Hunter, Kevin Lauerdale, and Teri E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing 2002*, pages 223–234. World Scientific, 2002.

[1222] S. Pavey, P. Johansson, L. Packer, J. Taylor, M. Stark, P.M. Pollock, G.J. Walker, G.M. Boyle, U. Harper, S.J. Cozzi, K. Hansen, L. Yudt, C. Schmidt, P. Hersey, K.A. Ellem, M.G. O'Rourke, P.G. Parsons, P. Meltzer, M. Ringner, and N.K. Hayward. Microarray expression profiling in melanoma reveals a BRAF mutation signature. *Oncogene*, 23(23):4060–4067, May 2004.

[1223] P. Pavlidis, T. S. Furey, M. Liberto, D. Haussler, and W. N. Grundy. Promoter Region-Based Classification of Genes. In *Pacific Symposium on Biocomputing*, pages 139–150, 2001.

[1224] P. Pavlidis, J. Weston, J. Cai, and W.N. Grundy. Gene functional classification from heterogeneous data. In *Proceedings of the Fifth Annual International Conference on Computational Biology*, pages 249–255, 2001.

[1225] P. Pavlidis, J. Weston, J. Cai, and W.S. Noble. Learning gene functional classifications from multiple data types. *J. Comput. Biol.*, 9(2):401–411, 2002.

[1226] Paul Pavlidis, Ilan Wapinski, and William Stafford Noble. Support vector machine classification on the web. *Bioinformatics*, 20(4):586–7, Mar 2004.

[1227] F. Pazos and A. Valencia. Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng.*, 9(14):609–614, 2001.

[1228] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Mateo, 1988. The classic original book on belief networks, which was certainly motivated by the idea that belief networks might have relevance to brains.

[1229] R. Pearlstein, R. Vaz, and D. Rampe. Understanding the structure-activity relationship of the human ether-a-go-go-related gene cardiac K+ channel. A model for bad behavior. *J. Med. Chem.*, 46(11):2017–2022, May 2003.

[1230] R. A. Pearlstein, R. J. Vaz, J. Kang, X.-L. Chen, M. Preobrazhenskaya, A. E. Shchekotikhin, A. M. Korolev, L. N. Lysenkova, O. V. Miroshnikova, J. Hendrix, and D. Rampe. Characterization of HERG potassium channel inhibition using CoMSiA 3D QSAR and homology modeling approaches. *Bioorg Med Chem Lett*, 13(10):1829–1835, May 2003.

[1231] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.

[1232] W. R. Pearson. Rapid and sensitive sequence comparisons with FASTP and FASTA. *Meth. Enzymol.*, 183:63–98, 1990.

[1233] M. Pellegrini, E. M. Marcotte, M. J. Thompson, D. Eisenberg, and T. O. Yeates. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc. Natl. Acad. Sci. USA*, 96:4285–4288, April 1999.

[1234] S. Peng, Q. Xu, X.B. Ling, X. Peng, W. Du, and L. Chen. Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines. *FEBS Lett.*, 555(2):358–362, 2003.

[1235] C. A. Pepperrell and P. Willett. Techniques for the calculation of three-dimensional structural similarity using inter-atomic distances. *J Comput Aided Mol Des*, 5(5):455–474, Oct 1991.

[1236] C. Perez-Iratxeta, P. Bork, and M. A. Andrade. Association of genes to genetically inherited diseases using data mining. *Nat. Genet.*, 31(3):316–319, Jul 2002.

[1237] D O Perkins, C Jeffries, and P Sullivan. Expanding the 'central dogma': the regulatory role of nonprotein coding genes and implications for the genetic liability to schizophrenia. *Molecular Psychiatry*, 10:69–78, 2005.

[1238] Z. E. Perlman, M. D. Slack, Y. Feng, T. J. Mitchison, L. F. Wu, and S. J. Altschuler. Multidimensional drug profiling by automated microscopy. *Science*, 306(5699):1194–1198, Nov 2004.

[1239] Emanuele Perola and Paul S Charifson. Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. *J Med Chem*, 47(10):2499–2510, May 2004.

[1240] C M. Perou, T. Sørlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lønning, A. L. Børresen-Dale, P. O. Brown, and D. Botstein. Molecular portraits of human breast tumours. *Nature*, 406(6797):747–752, Aug 2000.

[1241] N. Perry and V. J. van Geerestein. Database Searching on the basis of Three-Dimensional Similarity Using the sperm Program. *J Chem Inf Comput Sci*, 32:607–616, 1992.

[1242] Bjoern Peters, Huynh-Hoa Bui, Sune Frankild, Morten Nielson, Claus Lundegaard, Emrah Kostem, Derek Basch, Kasper Lamberth, Mikkel Harndahl, Ward Fleri, Stephen S Wilson, John Sidney, Ole Lund, Soren Buus, and Alessandro Sette. A community resource benchmarking predictions of peptide binding to MHC-I molecules. *PLoS Comput Biol*, 2(6):e65, Jun 2006.

[1243] Bjoern Peters and Alessandro Sette. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics*, 6:132, 2005.

[1244] Nico Pfeifer and Oliver Kohlbacher. Multiple instance learning allows mhc class ii epitope predictions across alleles. In *WABI '08: Proceedings of the 8th international workshop on Algorithms in Bioinformatics*, pages 210–221, Berlin, Heidelberg, 2008. Springer-Verlag.

[1245] Tho Hoan Pham, Kenji Satou, and Tu Bao Ho. Prediction and analysis of beta-turns in proteins by support vector machine. *Genome Inform Ser Workshop Genome Inform*, 14:196–205, 2003.

[1246] Tho Hoan Pham, Kenji Satou, and Tu Bao Ho. Support vector machines for prediction and analysis of beta and gamma-turns in proteins. *J. Bioinform. Comput. Biol.*, 3(2):343–58, Apr 2005.

[1247] D. L. Philips. A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.*, 9:84–97, 1962.

[1248] S. D. Pickett, J. S. Mason, and I. M. McLay. Diversity profiling and design using 3D pharmacophores : Pharmacophores-Derived Queries (PQD). *J. Chem. Inf. Comput. Sci.*, 36(6):1214–1223, 1996.

[1249] N. Piliouras, I. Kalatzis, N. Dimitropoulos, and D. Cavouras. Development of the cubic least squares mapping linear-kernel support vector machine classifier for improving the characterization of breast lesions on ultrasound. *Comput Med Imaging Graph*, 28(5):247–55, Jul 2004.

[1250] Y. Pilpel, P. Sudarsanam, and G. M. Church. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nature*, 29:153–159, 2001.

[1251] D. Pinkel and D. G. Albertson. Array comparative genomic hybridization and its applications in cancer. *Nat. Genet.*, 37 Suppl:S11–S17, Jun 2005.

[1252] D. Pinkel, R. Segraves, D. Sudar, S. Clark, I. Poole, D. Kowbel, C. Collins, W. L. Kuo, C. Chen, Y. Zhai, S. H. Dairkee, B. M. Ljung, J. W. Gray, and D. G. Albertson. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.*, 20(2):207–211, Oct 1998.

[1253] J. Platt. Fast training of support vector machines using sequential minimal optimization. In B. Schlkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, pages 185–208. MIT Press, Cambridge, MA, USA, 1999.

[1254] Dariusz Plewczynski, Adrian Tkacz, Adam Godzik, and Leszek Rychlewski. A support vector machine approach to the identification of phosphorylation sites. *Cell Mol Biol Lett*, 10(1):73–89, 2005.

[1255] Dariusz Plewczynski, Adrian Tkacz, Lucjan Stanislaw Wyrwicz, and Leszek Rychlewski. AutoMotif server: prediction of single residue post-translational modifications in proteins. *Bioinformatics*, 21(10):2525–7, May 2005.

[1256] N. Pochet, F. De Smet, J. A. K. Suykens, and B. L. R. De Moor. Systematic benchmarking of microarray data classification: assessing the role of non-linearity and dimensionality reduction. *Bioinformatics*, 20(17):3185–3195, Nov 2004.

[1257] J. Podani, Z.N. Oltvai, H. Jeong, B. Tombor, A.-L. Barabási, and E. Szathmáry. Comparable system-level organization of Archaea and Eukaryotes. *Nat. Genet.*, 29:54–56, 2001.

[1258] Poggio and Girosi. A Sparse Representation for Function Approximation. *Neural Comput*, 10(6):1445–54, Jul 1998.

[1259] Jonathan R. Pollack, Charles M. Perou, Ash A. Alizadeh, Michael B. Eisen, Alexander Pergamenschikov, Cheryl F. Williams, Stefanie S. Jeffrey, David Botstein, and Patrick O. Brown. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genetics*, 23:41–46, 1999.

[1260] W. Polonik. Measuring Mass Concentrations and Estimating Density Contour Clusters-An Excess Mass Approach. *Ann. Stat.*, 23(3):855–881, 1995.

[1261] W. Polonik. Minimum volume sets and generalized quantile processes. *Stochastic Processes and their Applications*, 69:1–24, 1997.

[1262] M. Pontil and A. Verri. Properties of support vector machines. *Neural Comput*, 10(4):955–74, May 1998.

[1263] J. Prados, A. Kalousis, J.C. Sanchez, L. Allard, O. Carrette, and M. Hilario. Mining mass spectra for diagnosis and biomarker discovery of cerebral accidents. *Proteomics*, 4(8):2320–2332, 2004.

[1264] K. N Bhanu Prakash, A. G. Ramakrishnan, S. Suresh, and Teresa W P Chow. Fetal lung maturity analysis using ultrasound image features. *IEEE Trans Inf Technol Biomed*, 6(1):38–45, Mar 2002.

[1265] Fernando Prez-Cruz, Carlos Bousoo-Calzn, and Antonio Arts-Rodrguez. Convergence of the IRWLS Procedure to the Support Vector Machine Solution. *Neural Comput*, 17(1):7–18, Jan 2005.

[1266] J. Qian, J. Lin, N. M. Luscombe, H. Yu, and M. Gerstein. Prediction of regulatory networks: genome-wide identification of transcription factor targets from gene expression data. *Bioinformatics*, 19(15):1917–1926, 2003.

[1267] J. Qian, N. M. Luscombe, and M. Gerstein. Protein Fold and Family Occurrence in Genomes: Power-Law Behaviour and Evolutionary Model. *J. Mol. Biol.*, 313:673–681, 2001.

[1268] Jian qiang Huang, Xiang xian Chen, and Le yu Wang. A novel method for tracking pedestrians from real-time video. *J Zhejiang Univ Sci*, 5(1):99–105, Jan 2004.

[1269] J. Qin, D. P. Lewis, and W. S. Noble. Kernel hierarchical gene clustering from microarray expression data. *Bioinformatics*, 19(16):2097–2104, 2003.

[1270] Zhong Qin, Qiang Yu, Jun Li, Zhi-Yi Wu, and Bing-Min Hu. Application of least squares vector machines in modelling water vapor and carbon dioxide fluxes over a cropland. *J Zhejiang Univ Sci*, 6(6):491–5, Jun 2005.

[1271] J. Qiu, J. Hue, A. Ben-Hur, J.-P. Vert, and W. S. Noble. A structural alignment kernel for protein structures. *Bioinformatics*, 23(9):1090–1098, May 2007.

[1272] S. Qiu, C. M. Adema, and T. Lane. A computational study of off-target effects of RNA interference. *Nucleic Acids Res.*, 33(6):1834–1847, 2005.

[1273] Yi Qu and Shizhong Xu. Supervised cluster analysis for microarray data based on multivariate Gaussian mixture. *Bioinformatics*, 20(12):1905–13, Aug 2004.

[1274] O Radulescu, S Lagarrigue, A Siegel, M Le Borgne, and P Veber. Topology and static response of interaction networks in molecular biology. *J.R.Soc.Interface*, Published online, 2005.

[1275] Gajendra P S Raghava and Joon H Han. Correlation and prediction of gene expression level from amino acid and dipeptide composition of its protein. *BMC Bioinformatics*, 6(1):59, Mar 2005.

[1276] S. Rahmann and C. Gräfe. Mean and variance of the Gibbs free energy of oligonucleotides in the nearest neighbor model under varying conditions. *Bioinformatics*, 20(17):2928–33, Nov 2004.

[1277] J Rahnenfuhrer, FS Domingues, J Maydt, and T. Lengauer. Calculating the statistical significance of changes in pathway activity from gene expression data. *Statistical Applications in Genetics and Molecular Biology*, 3(1):Article 16, 2004.

[1278] J.-C. Rain, L. Selig, H. De Reuse, V. Battaglia, C. Reverdy, S. Simon, G. Lenzen, F. Petel, J. Wojcik, V. Schächter, Y. Chemama, A. Labigne, and P. Legrain. The protein-protein interaction map of Helicobacter pylori. *Nature*, 409:211–215, 2001.

[1279] D. Rajwan and M. Feder. Universal finite memory machines for coding binary sequences. In *Proceedings of the Data Compression Conference (DCC 2000)*, pages 113–122, 2000.

[1280] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet. SimpleMKL. *J. Mach. Learn. Res.*, 9:2491–2521, 2008.

[1281] Alain Rakotomamonjy, Francis Bach, Stéphane Canu, and Yves Grandvalet. More efficiency in multiple kernel learning. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pages 775–782, New York, NY, USA, 2007. ACM.

[1282] L. Ralaivola, S. J. Swamidass, H. Saigo, and P. Baldi. Graph kernels for chemical informatics. *Neural Netw.*, 18(8):1093–1110, Sep 2005.

[1283] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C.H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J.P. Mesirov, T. Poggio, W. Gerald, M. Loda, E.S. Lander, and T.R. Golub. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl. Acad. Sci. USA*, 98(26):15149–15154, Dec 2001.

[1284] H. Rammensee, J. Bachmann, N. P. Emmerich, O. A. Bachor, and S. Stevanović. Syfpeithi: database for MHC ligands and peptide motifs. *Immunogenetics*, 50(3-4):213–219, Nov 1999.

[1285] H. G. Rammensee, T. Friede, and S. Stevanoviíc. MHC ligands and peptide motifs: first listing. *Immunogenetics*, 41(4):178–228, 1995.

[1286] J. Ramon and T. Gärtner. Expressivity versus efficiency of graph kernels. In T. Washio and L. De Raedt, editors, *Proceedings of the First International Workshop on Mining Graphs, Trees and Sequences*, pages 65–74, 2003.

[1287] H. Rangwala and G. Karypis. Profile-based direct kernels for remote homology detection and fold recognition. *Bioinformatics*, 21(23):4239–4247, Dec 2005.

[1288] F. Rapaport, E. Barillot, and J.-P. Vert. Classification of arrayCGH data using fused SVM. *Bioinformatics*, 24(13):i375–i382, Jul 2008.

[1289] F. Rapaport, A. Zynoviev, M. Dutreix, E. Barillot, and J.-P. Vert. Classification of microarray data using gene networks. *BMC Bioinformatics*, 8:35, 2007.

[1290] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*, 261(3):470–489, Aug 1996.

[1291] M. Rarey, S. Wefing, and T. Lengauer. Placement of medium-sized molecular fragments into active sites of proteins. *J Comput Aided Mol Des*, 10(1):41–54, Feb 1996.

[1292] Carl E. Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, December 2005.

[1293] G. Rätsch and S. Sonnenburg. Accurate splice site detection for Caenorhabditis elegans. In B. Schölkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 277–298. MIT Press, 2004.

[1294] G. Rätsch, S. Sonnenburg, and B. Schölkopf. RASE: recognition of alternatively spliced exons in C.elegans. *Bioinformatics*, 21(Suppl. 1):i369–i377, Jun 2005.

[1295] S. Raudys. How good are support vector machines? *Neural Netw*, 13(1):17–9, Jan 2000.

[1296] Pradeep Ravikumar, Han Liu, John Lafferty, and Larry Wasserman. Spam: Sparse additive models. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 1201–1208. MIT Press, Cambridge, MA, 2008.

[1297] Pradeep Ravikumar, Garvesh Raskutti, Martin Wainwright, and Bin Yu. Model selection in gaussian graphical models: High-dimensional consistency of $\boldsymbol{\ell_1}$-regularized mle. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 1329–1336. MIT Press, 2009.

[1298] P. A. Reche, J.-P. Glutting, and E. L. Reinherz. Prediction of MHC class I binding peptides using profile motifs. *Hum. Immunol.*, 63(9):701–709, Sep 2002.

[1299] Pedro A Reche, John-Paul Glutting, Hong Zhang, and Ellis L Reinherz. Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. *Immunogenetics*, 56(6):405–419, Sep 2004.

[1300] Kaushal Rege, Asif Ladiwala, Nihal Tugcu, Curt M Breneman, and Steven M Cramer. Parallel screening of selective and high-affinity displacers for proteins in ion-exchange systems. *J Chromatogr A*, 1033(1):19–28, Apr 2004.

[1301] Jon Reinders, Celine Delucinge Vivier, Gregory Theiler, Didier Chollet, Patrick Descombes, and Jerzy Paszkowski. Genome-wide, high-resolution DNA methylation profiling using bisulfite-mediated cytosine conversion. *Genome Research*, 18:469–76, 2008.

[1302] M. Remm, C.E. Storm, and E.L. Sonnhammer. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.*, 314(5):1041–1052, Dec 2001.

[1303] B. Ren, F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T. L. Volkert, C. J. Wilson, S. P. Bell, and R. A. Young. Genome-wide location and function of DNA binding proteins. *Science*, 290(5500):2306–2309, Dec 2000.

[1304] J. Ren and D.K. Stammers. HIV reverse transcriptase structures: designing new inhibitors and understanding mechanisms of drug resistance. *Trends Pharmacol Sci*, 26:4–7, 2005.

[1305] Y. Ren, W. Gong, Q. Xu, X. Zheng, D. Lin, Y. Wang, and T. Li. siRecords: an extensive database of mammalian siRNAs with efficacy ratings. *Bioinformatics*, Jan 2006.

[1306] J. D. M. Rennie and N. Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the 22nd international conference on Machine learning*, pages 713–719, New York, NY, USA, 2005. ACM Press.

[1307] Stefan A Rensing, Dana Fritzowsky, Daniel Lang, and Ralf Reski. Protein encoding genes in an ancient plant: analysis of codon usage, retained genes and splice sites in a moss, Physcomitrella patens. *BMC Genomics*, 6(1):43, Mar 2005.

[1308] I. Res, I. Mihalek, and O. Lichtarge. An evolution based classifier for prediction of protein interfaces without using protein structures. *Bioinformatics*, 21(10):2496–501, May 2005.

[1309] A. Reynolds, D. Leake, Q. Boese, S. Scaringe, W. S. Marshall, and A. Khvorova. Rational siRNA design for RNA interference. *Nat. Biotechnol.*, 22(3):326–330, Mar 2004.

[1310] P. Rice, I. Longden, and A. Bleasby. Emboss: the european molecular biology open software suite. *Trends Genet*, 16(6):276–277, Jun 2000.

[1311] Simon B Rice, Goran Nenadic, and Benjamin J Stapley. Mining protein function from text using term-based support vector machines. *BMC Bioinformatics*, 6 Suppl 1:S22, 2005.

[1312] Henning Riedesel, Bjrn Kolbeck, Oliver Schmetzer, and Ernst-Walter Knapp. Peptide binding at class I major histocompatibility complex scored with linear functions and support vector machines. *Genome Inform Ser Workshop Genome Inform*, 15(1):198–212, 2004.

[1313] Alessandro Rinaldo. Properties and refinements of the fused lasso. *Annals of Statistics*, 37(5B):2922–2952, 2009.

[1314] Risau-Gusman and Gordon. Generalization properties of finite-size polynomial support vector machines. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*, 62(5 Pt B):7092–9, Nov 2000.

[1315] S. Risau-Gusman and M. B. Gordon. Statistical mechanics of learning with soft margin classifiers. *Phys Rev E Stat Nonlin Soft Matter Phys*, 64(3 Pt 1):031907, Sep 2001.

[1316] J. Rissanen. Universal coding, information, prediction, and estimation. *IEEE Trans. Inform. Theory*, 30(4):629–636, Jul 1984.

[1317] J. Rissanen and G. Jr. Langdon. Universal modeling and coding. *IEEE Trans. Inform. Theory*, 27(1):12–23, Jan 1981.

[1318] J. Rissanen, T. P. Speed, and B. Yu. Density estimation by stochastic complexity. *IEEE Trans. Inform. Theory*, 38(2):315–323, Mar 1992.

[1319] J. J. Rissanen. Fisher information and stochastic complexity. *IEEE Trans. Inform. Theory*, 42(1):40–47, Jan 1996.

[1320] N. Robine, N. Uematsu, F. Amiot, X. Gidrol, E. Barillot, A. Nicolas, and V. Borde. Genome-wide redistribution of meiotic double-strand breaks in saccharomyces cerevisiae. *Mol. Cell. Biol.*, 27(5):1868–1880, Mar 2007.

[1321] J. Robinson, A. Malik, P. Parham, J. G. Bodmer, and S. G. Marsh. IMGT/HLA database–a sequence database for the human major histocompatibility complex. *Tissue Antigens*, 55(3):280–287, Mar 2000.

[1322] O. Roche, G. Trube, J. Zuegge, P. Pflimlin, A. Alanine, and G. Schneider. A virtual screening method for prediction of the HERG potassium channel liability of compound libraries. *Chembiochem*, 3(5):455–459, May 2002.

[1323] R. Rockafeller. *Convex Analysis.* Princeton Univ. Press, 1970.

[1324] D. Rogers and A. J. Hopfinger. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *J Chem Inf Comput Sci*, 34:854–866, 1994.

[1325] D. Rognan. Chemogenomic approaches to rational drug design. *Br. J. Pharmacol.*, 152:38–52, May 2007.

[1326] Thorsteinn Rögnvaldsson and Liwen You. Why neural networks should not be used for HIV-1 protease cleavage site prediction. *Bioinformatics*, 20(11):1702–9, Jul 2004.

[1327] F. James Rohlf. J. felsenstein, inferring phylogenies, sinauer assoc., 2004, pp. xx + 664. *J. Classif.*, 22(1):139–142, 2005.

[1328] C. Rolland, R. Gozalbes, A. Nicolaï, M.-F. Paugam, L. Coussy, F. Barbosa, D. Horvath, and F. Revah. G-protein-coupled receptor affinity prediction based on the use of a profiling dataset: Qsar design, synthesis, and experimental validation. *J. Med. Chem.*, 48(21):6563–6574, Oct 2005.

[1329] Derek G. Corneil Ronald C. Read. The graph isomorphism disease. *Journal of Graph Theory*, 1(4):339–363, 1977.

[1330] J. R. Rose, Jr. Turkett, W. H., I. C. Oroian, W. W. Laegreid, and J. Keele. Correlation of amino acid preference and mammalian viral genome type. *Bioinformatics*, 2005.

[1331] R. Rosenfeld, Q. Zheng, S. Vajda, and C. DeLisi. Flexible docking of peptides to class I major-histocompatibility-complex receptors. *Genet. Anal.*, 12(1):1–21, Mar 1995.

[1332] R. Rosipal and L. J. Trejo. Kernel partial least squares regression in reproducing kernel hilbert space. *J. Mach. Learn. Res.*, 2:97–123, 2001.

[1333] S. Rosset and J. Zhu. Piecewise linear regularized solution paths. *Annals of Statistics*, 35(3):1012–1030, 2007.

[1334] F. P. Roth, J. D. Hughes, P. W. Estep, and G. M. Church. Finding dna regulatory motifs within unaligned noncoding sequences clustered by whole-genome mrna quantitation. *Nat Biotechnol*, 16(10):939–945, October 1998.

[1335] Volker Roth. The generalized lasso: a wrapper approach to gene selection for microarray data. Technical report, Proceedings 14th International Conference on Automated Deduction (CADE-14), 252–255, 2002.

[1336] Volker Roth. The generalized lasso: a wrapper approach to gene selection for microarray data. In *Proc. CADE-14, 252–255*, 2002.

[1337] Volker Roth. The generalized LASSO. *IEEE Trans Neural Netw*, 15(1):16–28, Jan 2004.

[1338] Volker Roth and Bernd Fischer. The group-lasso for generalized linear models: uniqueness of solutions and efficient algorithms. In *ICML*, pages 848–855, 2008.

[1339] Volker Roth and Tilman Lange. Bayesian class discovery in microarray datasets. *IEEE Trans Biomed Eng*, 51(5):707–718, May 2004.

[1340] O. Rötzschke, K. Falk, S. Stevanović, G. Jung, and H. C. Rammensee. Peptide motifs of closely related HLA class I molecules encompass substantial differences. *Eur. J. Immunol.*, 22(9):2453–2456, Sep 1992.

[1341] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–6, Dec 2000.

[1342] G. Rücker and C. Rücker. Counts of All Walks as Atomic and Molecular Descriptors. *J Chem Inf Comput Sci*, 33:683–695, 1993.

[1343] S. Rudd and I. V. Tetko. Eclair–a web service for unravelling species origin of sequences sampled from mixed host interfaces. *Nucleic Acids Res*, 33(Web Server issue):W724–7, Jul 2005.

[1344] S. Ruepp, F. Boess, L. Suter, M. C. de Vera, G. Steiner, T. Steele, T. Weiser, and S. Albertini. Assessment of hepatotoxic liabilities by transcript profiling. *Toxicol Appl Pharmacol*, Jun 2005.

[1345] Nicole Rusk and Veronique Kiermer. Primer: Sequencing - the next generation. *Nature Methods*, 5:15, 2008.

[1346] R. B. Russell and G. J. Barton. Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins*, 14(2):309–323, Oct 1992.

[1347] Petra Rsch, Michaela Harz, Michael Schmitt, Klaus-Dieter Peschke, Olaf Ronneberger, Hans Burkhardt, Hans-Walter Motzkus, Markus Lankers, Stefan Hofer, Hans Thiele, and Jrgen Popp. Chemotaxonomic identification of single bacteria by micro-Raman spectroscopy: application to clean-room-relevant biological contaminations. *Appl Environ Microbiol*, 71(3):1626–37, Mar 2005.

[1348] Omowunmi Sadik, Walker H Land, Adam K Wanekaya, Michiko Uematsu, Mark J Embrechts, Lut Wong, Dale Leibensperger, and Alex Volykin. Detection and classification of organophosphate nerve agent simulants using support vector machines with multiarray sensors. *J Chem Inf Comput Sci*, 44(2):499–507, 2004.

[1349] J. Saeh, P. Lyne, B. Takasaki, and D. Cosgrove. Lead hopping using SVM and 3D pharmacophore fingerprints. *J Chem Inf Model*, 45(4):1122–1133, Jul 2005.

[1350] P. Saetrom. Predicting the efficacy of short oligonucleotides in antisense and RNAi experiments with boosted genetic programming. *Bioinformatics*, 20(17):3055–3063, 2004.

[1351] P. Saetrom and O. Snve. A comparison of siRNA efficacy predictors. *Biochem. Biophys. Res. Commun.*, 321(1):247–53, Aug 2004.

[1352] Y. Saeys, S. Degroeve, D. Aeyels, P. Rouz, and Y. Van de Peer. Feature selection for splice site prediction: A new method using EDA-based feature ranking. *BMC Bioinformatics*, 5(64), 2004.

[1353] Y. Saeys, S. Degroeve, D. Aeyels, Y. Van de Peer, and P. Rouze. Fast feature selection using a simple estimation of distribution algorithm: a case study on splice site prediction. *Bioinformatics*, 19(Suppl. 1):ii179–ii188, 2003.

[1354] H. Saigo, T. Kadowaki, and K. Tsuda. A linear programming approach for molecular QSAR analysis. In T. Gärtner, G. C. Garriga, and T. Meinl, editors, *International workshop on mining and learning with graphs (MLG)*, pages 85–96, 2006.

[1355] H. Saigo, S. Nowozin, T. Kadowaki, T. Kudo, and K. Tsuda. gBoost: a mathematical programming approach to graph classification and regression. *Mach. Learn.*, 75(1):69–89, 2009.

[1356] H. Saigo, J.-P. Vert, N. Ueda, and T. Akutsu. Protein homology detection using string alignment kernels. *Bioinformatics*, 20(11):1682–1689, 2004.

[1357] T. Saito, H. Yamada, and K. Yamamoto. On the data base etl9b of handprinted characters in jis chinese characters and its analysis. *IEICE Trans*, 68, 1985.

[1358] S. Saitoh. *Theory of reproducing Kernels and its applications.* Longman Scientific & Technical, Harlow, UK, 1988.

[1359] H. Salgado, S. Gama-Castro, M. Peralta-Gil, E. Díaz-Peredo, F. Sánchez-Solano, A. Santos-Zavaleta, I. Martínez-Flores, V. Jiménez-Jacinto, C. Bonavides-Martínez, J. Segura-Salazar, A. Martínez-Antonio, and J. Collado-Vides. RegulonDB (version 5.0): Escherichia coli K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res.*, 34(Database issue):D394–D397, Jan 2006.

[1360] N. Salim, J. Holliday, and P. Willett. Combination of fingerprint-based similarity coefficients using data fusion. *J Chem Inf Comput Sci*, 43(2):435–442, 2003.

[1361] J. Salomon and D. R. Flower. Predicting Class II MHC-Peptide binding: a kernel based approach using similarity scores. *BMC Bioinformatics*, 7:501, 2006.

[1362] Pamela A Sample, Michael H Goldbaum, Kwokleung Chan, Catherine Boden, Te-Won Lee, Christiana Vasile, Andreas G Boehm, Terrence Sejnowski, Chris A Johnson, and Robert N Weinreb. Using machine learning classifiers to identify glaucomatous change earlier in standard visual fields. *Invest Ophthalmol Vis Sci*, 43(8):2660–5, Aug 2002.

[1363] Marta Sanchez-Carbayo, Nicholas D Socci, Juan Jose Lozano, Wentian Li, Elizabeth Charytonowicz, Thomas J Belbin, Michael B Prystowsky, Angel R Ortiz, Geoffrey Childs, and Carlos Cordon-Cardo. Gene discovery in bladder cancer progression using cDNA microarrays. *Am. J. Pathol.*, 163(2):505–16, Aug 2003.

[1364] M. C. Sanguinetti and J. S. Mitcheson. Predicting drug-hERG channel interactions that cause acquired long QT syndrome. *Trends Pharmacol. Sci.*, 26(3):119–124, Mar 2005.

[1365] M.C. Sanguinetti and M. Tristani-Firouzi. hERG potassium channels and cardiac arrhythmia. *Nature*, 440(7083):463–469, Mar 2006.

[1366] K. Sanjiv and H. Martial. Discriminative Random Fields: A Discriminative Framework for Contextual Interaction in Classification. In *Proceedings of the Ninth IEEE International Conference on Computer Vision*, page 1150. IEEE Computer Society, 2003.

[1367] Deepak Sarda, Gek Huey Chua, Kuo-Bin Li, and Arun Krishnan. pSLIP: SVM based protein subcellular localization prediction using multiple physicochemical properties. *BMC Bioinformatics*, 6(1):152, Jun 2005.

[1368] Alexander P Sassi, Frank Andel, Hans-Marcus L Bitter, Michael P S Brown, Robert G Chapman, Jeraldine Espiritu, Alfred C Greenquist, Isabelle Guyon, Mariana Horchi-Alegre, Kathy L Stults, Ann Wainright, Jonathan C Heller, and John T Stults. An automated, sheathless capillary electrophoresis-mass spectrometry platform for discovery of biomarkers in human serum. *Electrophoresis*, 26(7-8):1500–12, Apr 2005.

[1369] Helga Satzinger. Theodor and Marcella Boveri: chromosomes and cytoplasm in heredity and development. *Nature Reviews Genetics*, 9:231–238, 2008.

[1370] L. K. Saul and S. T. Roweis. Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds. *J. Mach. Learn. Res.*, 4:119–155, 2003.

[1371] Dietmar Saupe and Dejan V. Vranic. 3d model retrieval with spherical harmonics and moments. In *Proceedings of the 23rd DAGM-Symposium on Pattern Recognition*, pages 392–397, London, UK, 2001. Springer-Verlag.

[1372] Charles L. Sawyers. The cancer biomarker problem. *Nature*, 452:548–552, 2008.

[1373] A.K. Saxena and P. Prathipati. Comparison of mlr, pls and ga-mlr in qsar analysis. *SAR. QSAR. Environ. Res.*, 14:433–445, 2003.

[1374] P. C. Scacheri, O. Rozenblatt-Rosen, N. J. Caplen, T. G. Wolfsberg, L. Umayam, J. C. Lee, C. M. Hughes, K. S. Shanmugam, A. Bhattacharjee, M. Meyerson, and F. S. Collins. Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proc. Natl. Acad. Sci. USA*, 101(7):1892–7, Feb 2004.

[1375] R. Schachtner, D. Lutter, P. Knollmller, A. M. Tom, F. J. Theis, G. Schmitz, M. Stetter, P. Gmez Vilda, and E. W. Lang. Knowledge-based gene expression classification via matrix factorization. *Bioinformatics*, 24(15):1688–1697, Aug 2008.

[1376] Claire Schalon, Jean-Sbastien Surgand, Esther Kellenberger, and Didier Rognan. A simple and fuzzy method to align and compare druggable ligand-binding sites. *Proteins*, 71(4):1755–1778, Jun 2008.

[1377] C. Schellewald and C. Schnorr. Probabilistic subgraph matching based on convex relaxation. In *EMMCVPR05*, pages 171–186, 2005.

[1378] Christian Schellewald, Stefan Roth, and Christoph Schnörr. Evaluation of convex optimization techniques for the weighted graph-matching problem in computer vision. In *Proceedings of the 23rd DAGM-Symposium on Pattern Recognition*, pages 361–368, London, UK, 2001. Springer-Verlag.

[1379] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown. Quantitative monitoring of gene expression patterns with a complementary dna microarray. *Science*, 270(5235):467–470, Oct 1995.

[1380] Douglas C. Schmidt and Larry E. Druffel. A fast backtracking algorithm to test directed graphs for isomorphism using distance matrices. *J. ACM*, 23(3):433–445, 1976.

[1381] Stefan Schmitt, Daniel Kuhn, and Gerhard Klebe. A new method to detect related function among proteins independent of sequence and fold homology. *J Mol Biol*, 323(2):387–406, Oct 2002.

[1382] G. Schneider and P. Wrede. Artificial neural networks for computer-based molecular design. *Prog Biophys Mol Biol*, 70(3):175–222, 1998.

[1383] Gisbert Schneider and Uli Fechner. Advances in the prediction of protein targeting signals. *Proteomics*, 4(6):1571–80, Jun 2004.

[1384] B . Schölkopf, C. Burges, and V. Vapnik. Extracting support data for a given task. In M. Fayyad and R. Uthurusamy, editors, *Proceedings of the First International Conference on Knowledge Discovery & Data Mining*. AAAI Press, 1995.

[1385] B. Schölkopf, C. Burges, and V. Vapnik. Incorporating invariances in support vector learning machines. In C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen, and B. Sendhoff, editors, *ICANN 96: Proceedings of the 1996 International Conference on Artificial Neural Networks*, pages 47–52, London, UK, 1996. Springer-Verlag.

[1386] B. Schölkopf, R. Herbrich, and A. J. Smola. A generalized representer theorem. In *Proceedings of the 14th Annual Conference on Computational Learning Theory*, volume 2011 of *Lecture Notes in Computer Science*, pages 416–426, Berlin / Heidelberg, 2001. Springer.

[1387] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the Support of a High-Dimensional Distribution. *Neural Comput.*, 13:1443–1471, 2001.

[1388] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* MIT Press, Cambridge, MA, 2002.

[1389] B. Schölkopf, A.J. Smola, and K.-R. Müller. Kernel principal component analysis. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, pages 327–352. MIT Press, 1999.

[1390] B. Schölkopf, K. Tsuda, and J.-P. Vert. *Kernel Methods in Computational Biology.* MIT Press, The MIT Press, Cambridge, Massachussetts, 2004.

[1391] B. Schölkopf, J. Weston, E. Eskin, C. Leslie, and W.S. Noble. A Kernel Approach for Learning from Almost Orthogonal Patterns. In *Proceedings of ECML 2002*, 2002.

[1392] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt. Support Vector Method for Novelty Detection. In S.A. Solla, T.K. Leen, and K.-R. Müller, editors, *Adv. Neural Inform. Process. Syst.*, volume 12, pages 582–588. MIT Press, 2000.

[1393] Dustin E. Schones and Keji Zhao. Genome-wide approaches to studying chromatin modifications. *Nature Reviews Genetics*, 9:179–191, 2008.

[1394] Erwin Schrdinger. *Qu'est-ce que la vie?* Christian Bourgois Editeur, 1986, 1944.

[1395] S. Schubert, A. Grnweller, V. A. Erdmann, and J. Kurreck. Local RNA target structure influences siRNA efficacy: systematic analysis of intentionally designed binding regions. *J. Mol. Biol.*, 348(4):883–893, May 2005.

[1396] O. Schueler-Furman, Y. Altuvia, A. Sette, and H. Margalit. Structure-based prediction of binding peptides to MHC class I molecules: application to a broad range of MHC alleles. *Protein Sci.*, 9(9):1838–1846, Sep 2000.

[1397] A. Schuffenhauer, P. Floersheim, P. Acklin, and E. Jacoby. Similarity metrics for ligands reflecting the similarity of the target proteins. *J. Chem. Inf. Comput. Sci.*, 43(2):391–405, 2003.

[1398] A. Schuffenhauer, J. Zimmermann, R. Stoop, J. J. van der Vyver, S. Lecchini, and E. Jacoby. An ontology for pharmaceutical ligands and its application for in silico screening and library design. *J. Chem. Inf. Comput. Sci.*, 42(4):947–955, 2002.

[1399] A. Schumacher, P. Kapranov, Z. Kaminsky, J. Flanagan, A. Assadzadeh, P. Yau, C. Virtanen, N. Winegarden, J. Cheng, T. Gingeras, and A. Petronis. Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Research*, 34:528–542, 2006.

[1400] Stephan C Schuster. Next-generation sequencing transforms today's biology. *Nature Methods*, 5:16–18, 2007.

[1401] D. S. Schwarz, G. Hutvagner, T. Du, Z. Xu, N. Aronin, and P. D. Zamore. Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, 115(2):199–208, Oct 2003.

[1402] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.

[1403] Holger Schwender, Manuela Zucknick, Katja Ickstadt, Hermann M Bolt, and G. E. N. I. C. A. network. A pilot study on the application of statistical classification procedures to molecular epidemiological data. *Toxicol Lett*, 151(1):291–9, Jun 2004.

[1404] Walter R. P. Scott, Ilario G. Tironi, Alan E. Mark, Salomon R. Billeter, Jens Fennen, Andrew E. Torda, Thomas Huber, and Peter Kruger. The gromos biomolecular simulation program package. *J. Phys. Chem. A*, 103:3596–3607, 1999.

[1405] C. Scovel and I. Steinwart. Fast Rates for Support Vector Machines. Technical report, Los Alamos National Laboratory, 2004.

[1406] M. Sebag and C. Rouveirol. Tractable Induction and Classification in First-Order Logic via Stochastic Matching. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence*, pages 888–893. Morgan Kaufmann, 1997.

[1407] J. Sebat, B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Mnr, H. Massa, M. Walker, M. Chi, N. Navin, R. Lucito, J. Healy, J. Hicks, K. Ye, A. Reiner, T. C. Gilliam, B. Trask, N. Patterson, A. Zetterberg, and M. Wigler. Large-scale copy number polymorphism in the human genome. *Science*, 305(5683):525–528, Jul 2004.

[1408] Jonathan Sebat. Major changes in our dna lead to major changes in our thinking. *Nat Genet*, 39(7 Suppl):S3–S5, Jul 2007.

[1409] M. Seeger. Covariance Kernels from Bayesian Generative Models. In *Adv. Neural Inform. Process. Syst.*, volume 14, pages 905–912, 2002.

[1410] Matthias Seeger. Gaussian processes for machine learning. *Int J Neural Syst*, 14(2):69–106, Apr 2004.

[1411] E. Segal, N. Friedman, N. Kaminski, A. Regev, and D. Koller. From signatures to models: understanding cancer using microarrays. *Nat Genet*, 37(6 Suppl):S38–45, 2005.

[1412] E. Segal, N. Friedman, D. Koller, and A. Regev. A module map showing conditional activity of expression modules in cancer. *Nat. Genet.*, 36(10):1090–1098, Oct 2004.

[1413] E. Segal, N. Friedman, D. Koller, and A. Regev. A module map showing conditional activity of expression modules in cancer. *Nat Genet*, 36(10):1090–8, 2004.

[1414] E. Segal, M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller, and N. Friedman. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.*, 34(2):166–176, Jun 2003.

[1415] M. R. Segal, K. D. Dahlquist, and B. R. Conklin. Regression approaches for microarray data analysis. *J. Comput. Biol.*, 10(6):961–980, 2003.

[1416] N. H. Segal, P. Pavlidis, C. R. Antonescu, R. G. Maki, W. S. Noble, D. DeSantis, J. M. Woodruff, J. J. Lewis, M. F. Brennan, A. N. Houghton, and C. Cordon-Cardo. Classification and Subtype Prediction of Adult Soft Tissue Sarcoma by Functional Genomics. *Am. J. Pathol.*, 163(2):691–700, Aug 2003.

[1417] N. H. Segal, P. Pavlidis, W. S. Noble, C. R. Antonescu, A. Viale, U. V. Wesley, K. Busam, H. Gallardo, D. DeSantis, M. F. Brennan, C. Cordon-Cardo, J. D. Wolchok, and A. N. Houghton. Classification of Clear-Cell Sarcoma as a Subtype of Melanoma by Genomic Profiling. *J. Clin. Oncol.*, 21(9):1775–1781, May 2003.

[1418] M. Seike, T. Kondo, K. Fujii, T. Okano, T. Yamada, Y. Matsuno, A. Gemma, S. Kudoh, and S. Hirohashi. Proteomic signatures for histological types of lung cancer. *Proteomics*, Jul 2005.

[1419] Douglas W. Selinger, Kevin J. Cheung, Rui Mei, Erik M. Johansson, Craig S. Richmond, Frederick R. Blattner, David J. Lockhart, and George M. Church. RNA expression analysis using a 30 base pair resolution Escherichia coli genome array. *Nat. Biotechnol.*, 18:1262–1268, 2000.

[1420] D. Semizarov, L. Frost, A. Sarthy, P. Kroeger, D. N. Halbert, and S. W. Fesik. Specificity of short interfering RNA determined through gene expression signatures. *Proc. Natl. Acad. Sci. USA*, 100(11):6347–52, May 2003.

[1421] T.Z. Sen, A. Kloczkowski, R.L. Jernigan, C. Yan, V. Honavar, K.M. Ho, C.Z. Wang, Y. Ihm, H. Cao, X. Gu, and D. Dobbs. Predicting binding sites of hydrolase-inhibitor complexes by combining several methods. *BMC Bioinformatics*, 5(205), 2004.

[1422] Pasak Senawongse, Andrew R Dalby, and Zheng Rong Yang. Predicting the phosphorylation sites using hidden markov models and machine learning methods. *J Chem Inf Model*, 45(4):1147–52, 2005.

[1423] J. H. Seol, A. Shevchenko, A. Shevchenko, and R. J. Deshaies. Skp1 forms multiple protein complexes, including RAVE, a regulator of V-ATPase assembly. *Nat Cell Biol*, 3(4):384–91, Apr 2001.

[1424] J.R. Serra, E.D. Thompson, and P.C. Jurs. Development of binary classification of structural chromosome aberrations for a diverse set of organic compounds from molecular structure. *Chem. Res. Toxicol.*, 16(2):153–163, 2003.

[1425] A. Sette, R. Chesnut, and J. Fikes. HLA expression in cancer: implications for T cell-based immunotherapy. *Immunogenetics*, 53(4):255–263, 2001.

[1426] A. Sette and J. Sidney. HLA supertypes and supermotifs: a functional perspective on HLA polymorphism. *Curr Opin Immunol*, 10(4):478–482, Aug 1998.

[1427] A. Sette and J. Sidney. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics*, 50(3-4):201–212, Nov 1999.

[1428] A. Sette, A. Vitiello, B. Reherman, P. Fowler, R. Nayersina, W. M. Kast, C. J. Melief, C. Oseroff, L. Yuan, J. Ruppert, J. Sidney, M. F. del Guercio, S. Southwood, R. T. Kubo, R. W. Chesnut, H. M. Grey, and F. V. Chisari. The relationship between class i binding affinity and immunogenicity of potential cytotoxic t cell epitopes. *J Immunol*, 153(12):5586–5592, Dec 1994.

[1429] S. Shabalina, A. Spiridonov, and A. Ogurtsov. Computational models with thermodynamic and composition features improve siRNA design. *BMC Bioinformatics*, 7(1):65, Feb 2006.

[1430] S. Shacham, Y. Marantz, S. Bar-Haim, O. Kalid, D. Warshaviak, N. Avisar, B. Inbal, A. Heifetz, M. Fichman, M. Topf, Z. Naor, S. Noiman, and O. M. Becker. PREDICT modeling and in-silico screening for G-protein coupled receptors. *Proteins*, 57(1):51–86, Oct 2004.

[1431] Shesha Shah and P. S. Sastry. Fingerprint classification using a feedback-based line detector. *IEEE Trans Syst Man Cybern B Cybern*, 34(1):85–94, Feb 2004.

[1432] Y.J. Shann, C. Cheng, C.H. Chiao, D.T. Chen, P.H. Li, and M.T. Hsu. Genome-wide mapping and characterization of hypomethylated sites in human tissues and breast cancer cell lines. *Genome Research*, 18:791–801, 2008.

[1433] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13(11):2498–2504, Nov 2003.

[1434] William Shannon, Robert Culverhouse, and Jill Duncan. Analyzing microarray data using cluster analysis. *Pharmacogenomics*, 4(1):41–52, Jan 2003.

[1435] Larry S. Shapiro and Michael Brady. Feature-based correspondence: an eigenvector approach. *Image Vision Comput.*, 10(5):283–288, 1992.

[1436] R. Sharan, S. Suthram, R.M. Kelley, T. Kuhn, S. McCuine, P. Uetz, T. Sittler, R.M. Karp, and T. Ideker. Conserved patterns of protein interaction in multiple species. *Proc. Natl. Acad. Sci. USA*, 102(6):1974–1979, Feb 2005.

[1437] Roded Sharan and Eugene W Myers. A motif-based framework for recognizing sequence families. *Bioinformatics*, 21 Suppl 1:i387–i393, Jun 2005.

[1438] J. Shawe-Taylor and N. Cristianini. On the Generalization of Soft Margin Algorithms. *IEEE Transactions on Information Theory*, 48(10):2721–2735, October 2002.

[1439] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis.* Cambridge University Press, New York, NY, USA, 2004.

[1440] R. She, F. Chen, K. Wang, M. Ester, J.L. Gardy, and F.S.L. Brinkman. Frequent-subsequence-based prediction of outer membrane proteins. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 436–445. ACM Press, 2003.

[1441] Felix B Sheinerman, Bissan Al-Lazikani, and Barry Honig. Sequence, structure and energetic determinants of phosphopeptide selectivity of SH2 domains. *J Mol Biol*, 334(4):823–841, Dec 2003.

[1442] Felix B Sheinerman, Elie Giraud, and Abdelazize Laoui. High affinity targets of protein kinase inhibitors have similar residues at the positions energetically important for binding. *J Mol Biol*, 352(5):1134–1156, Oct 2005.

[1443] Li Shen, Jie Yang, and Yue Zhou. Detection of PVCs with support vector machine. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*, 22(1):78–81, Feb 2005.

[1444] N. Sherashidze, S.V.N. Vishwanathan, T.H. Petri, K. Mehlhorn, and K.M. Borgwardt. Efficient graphlet kernels for large graph comparison. In *12th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 488–495, Clearwater Beach, Florida USA, 2009. Society for Artificial Intelligence and Statistics.

[1445] G. Sherlock, T. Hernandez-Boussard, A. Kasarskis, G. Binkley, J.C. Matese, S.S. Dwight, M. Kaloper, S. Weng, H. Jin, C.A. Ball, M.B. Eisen, and P.T. Spellman. The Stanford Microarray Database. *Nucleic Acids Res.*, 29(1):152–155, Jan 2001.

[1446] Charles J. Sherr. Principles of tumor suppression. *Cell*, 116:235–246, 2004.

[1447] D. Shi, D. S. Yeung, and J. Gao. Sensitivity analysis applied to the construction of radial basis function networks. *Neural Netw*, Jun 2005.

[1448] Lei Shi and Fabien Campagne. Building a protein name dictionary from full text: a machine learning term extraction approach. *BMC Bioinformatics*, 6(1):88, Apr 2005.

[1449] P.C. Shields. Universal redundancy rates do not exist. *IEEE Trans. Inform. Theory*, 39(2):520–524, Mar 1993.

[1450] Alistair Shilton, M. Palaniswami, Daniel Ralph, and Ah Chung Tsoi. Incremental training of support vector machines. *IEEE Trans Neural Netw*, 16(1):114–31, Jan 2005.

[1451] H. Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2):227–244, October 2000.

[1452] H. Shimodaira, K.-I. Noma, M. Nakai, and S. Sagayama. Dynamic time-alignment kernel in support vector machine. In *Adv. Neural. Inform. Process Syst.*, pages 921–928, 2001.

[1453] M. A. Shipp, K. N. Ross, P. Tamayo, A. P. Weng, J. L. Kutok, R. C. T. Aguiar, M. Gaasenbeek, M. Angelo, M. Reich, G. A. Pinkus, T. S. Ray, M. A. Koval, K. W. Last, A. Norton, T. A. Lister, J. Mesirov, D. S. Neuberg, E. S. Lander, J. C. Aster, and T. R. Golub. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat. Med.*, 8(1):68–74, 2002.

[1454] J. L. Shock, K. F. Fischer, and J. L. DeRisi. Whole-genome analysis of mrna decay in plasmodium falciparum reveals a global lengthening of mrna half-life during the intra-erythrocytic development cycle. *Genome Biol.*, 8(7):R134, 2007.

[1455] Ali Shoeb, Herman Edwards, Jack Connolly, Blaise Bourgeois, S. Ted Treves, and John Guttag. Patient-specific seizure onset detection. *Epilepsy Behav*, 5(4):483–98, Aug 2004.

[1456] N. Shulman and M. Feder. The Uniform Distribution as a Universal Prior. *IEEE Trans. Inform. Theory*, 50(6):1356 – 1362, Jun 2004.

[1457] Alexandra Shulman-Peleg, Maxim Shatsky, Ruth Nussinov, and Haim J. J. Wolfson. Multibind and mappis: webservers for multiple alignment of protein 3d-binding sites and their interactions. *Nucleic Acids Res*, 36:260–264, May 2008.

[1458] J. Sidney, M. F. del Guercio, S. Southwood, V. H. Engelhard, E. Appella, H. G. Rammensee, K. Falk, O. Rötzschke, M. Takiguchi, and R. T. Kubo. Several HLA alleles share overlapping peptide specificities. *J Immunol*, 154(1):247–259, Jan 1995.

[1459] J. Sidney, H. M. Grey, S. Southwood, E. Celis, P. A. Wentworth, M. F. del Guercio, R. T. Kubo, R. W. Chesnut, and A. Sette. Definition of an HLA-A3-like supermotif demonstrates the overlapping peptide-binding repertoires of common HLA molecules. *Hum Immunol*, 45(2):79–93, Feb 1996.

[1460] J. A. Siepen, S. E. Radford, and D. R. Westhead. Beta Edge strands in protein structure prediction and aggregation. *Protein Sci.*, 12(10):2348–2359, 2003.

[1461] B. W. Silverman. On the Estimation of a Probability Density Function by the Maximum Penalized Likelihood Method. *Ann. Stat.*, 10:795–810, 1982.

[1462] P. Simard, B. Victorri, Y. LeCun, and J. S. Denker. Tangent prop - a formalism for specifying selected invariances in an adaptive network. In J. E. Moody, S. J. Hanson, and R. Lippmann, editors, *Adv. Neural. Inform. Process Syst. 4*, pages 895–903. Morgan Kaufman, 1992.

[1463] R. E. Sinden. A proteomic analysis of malaria biology: integration of old literature and new technologies. *Int. J. Parasitol.*, 34(13-14):1441–1450, Dec 2004.

[1464] V. Sindhwani, P. Niyogi, and M. Belkin. Manifold Regularization: A Geometric Framework for Learning from Examples. Technical Report TR-2004-06, The University of Chicago, 2004.

[1465] Vikas Sindhwani, Subrata Rakshit, Dipti Deodhare, Deniz Erdogmus, Jose C Principe, and Partha Niyogi. Feature selection in MLPs and SVMs based on maximum output information. *IEEE Trans Neural Netw*, 15(4):937–48, Jul 2004.

[1466] A.C. Singer, S.S. Kozat, and M. Feder. Universal linear least squares prediction: upper and lower bounds. *IEEE Trans. Inform. Theory*, 48(8):2354 – 2362, Aug 2002.

[1467] R. Singh, J. Xu, and B. Berger. Pairwise global alignment of protein interaction networks by matching neighborhood topology. *The Proceedings of the 11th International Conference on Research in Computational Molecular Biology (RECOMB)*, 2007.

[1468] R. Singh, J. Xu, and B. Berger. Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proc. Natl. Acad. Sci. USA*, 105(35):12763–12768, Sep 2008.

[1469] K. Sjölander. Phylogenomic inference of protein molecular function: advances and challenges. *Bioinformatics*, 20(2):170–179, Jan 2004.

[1470] F. Slanina and M. Kotrla. Random networks created by biological evolution. *Phys. Rev. E*, 62(5):6170–6177, 2000.

[1471] S. Smale and D. Zhou. Estimating the approximation error in learning theory. *Analysis and Applications*, 1(1), 2003.

[1472] A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and TRECVid. In *MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, New-York, NY, USA, 2006. ACM Press.

[1473] P. A. Smith, M. J. Sorich, L. S C Low, R. A. McKinnon, and J. O. Miners. Towards integrated ADME prediction: past, present and future directions for modelling metabolism by UDP-glucuronosyltransferases. *J Mol Graph Model*, 22(6):507–17, Jul 2004.

[1474] T. Smith and M. Waterman. Identification of common molecular subsequences. *J. Mol. Biol.*, 147:195–197, 1981.

[1475] A. Smola and R. Kondor. Kernels and Regularization on Graphs. In B. Schölkopf and M.K. Warmuth, editors, *Proceedings of 16th Annual Conference on Computational Learning Theory*, pages 144–158. Springer-Verlag, 2003.

[1476] A.J. Smola, B. Schölkopf, and K.-R. Müller. The connection between regularization operators and support vector kernels. *Neural Networks*, 11(4):637–649, 1998.

[1477] G. K. Smyth. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, chapter Limma: linear model for microarray data, pages 397–420. Springer, New York, 2005.

[1478] G. K. Smyth and T. P. Speed. Normalization of cDNA microarray data. *Methods*, 31:265–273, 2003.

[1479] O. Snve, M. Nedland, S. H. Fjeldstad, H. Humberset, O. R. Birkeland, T. Grönfeld, and P. Saetrom. Designing effective siRNAs with off-target control. *Biochem. Biophys. Res. Commun.*, 325(3):769–73, Dec 2004.

[1480] R. V. Solé, R. Pastor-Satorras, E. D. Smith, and T. Kepler. A Model of Large-Scale Proteome Evolution. Technical report, Santa Fe Institute, 2001. Working paper 01-08-041.

[1481] S. Solinas-Toldo, S. Lampel, S. Stilgenbauer, J. Nickolenko, A. Benner, H. Dohner, T. Cremer, and P. Lichter. Matrix-based comparative genomic hybridization: Biochips to screen for genomic imbalances. *Genes Chromosomes Cancer*, 20:399–407, 1997.

[1482] M. Song and M. Clark. Development and evaluation of an in silico model for hERG binding. *J. Chem. Inf. Model.*, 46(1):392–400, 2006.

[1483] Minghu Song, Curt M Breneman, Jinbo Bi, N. Sukumar, Kristin P Bennett, Steven Cramer, and Nihal Tugcu. Prediction of protein retention times in anion-exchange chromatography systems using support vector regression. *J Chem Inf Comput Sci*, 42(6):1347–57, 2002.

[1484] Xiaowei Song, Arnold Mitnitski, Jafna Cox, and Kenneth Rockwood. Comparison of machine learning techniques with classical statistical models in predicting health outcomes. *Medinfo*, 11(Pt 1):736–40, 2004.

[1485] S. Sonnenburg, G. Rätsch, A. Jagota, and K.-R. Müller. New methods for splice-site recognition. In JR. Dorronsoro, editor, *Proc. International conference on artificial Neural Networks ? ICANN?02*, number 2415 in LNCS, pages 329–336. Springer Berlin, 2002.

[1486] Sören Sonnenburg, Gunnar Rätsch, Christin Schäfer, and Bernhard Schölkopf. Large scale multiple kernel learning. *J. Mach. Learn. Res.*, 7:1531–1565, 2006.

[1487] M. J. Sorich, J. O. Miners, R. A. McKinnon, D. A. Winkler, F. R. Burden, and P. A. Smith. Comparison of linear and nonlinear classification algorithms for the prediction of drug and chemical metabolism by human UDP-glucuronosyltransferase isoforms. *J Chem Inf Comput Sci*, 43(6):2019–24, 2003.

[1488] Michael J Sorich, Ross A McKinnon, John O Miners, David A Winkler, and Paul A Smith. Rapid prediction of chemical metabolism by human UDP-glucuronosyltransferase isoforms using quantum chemical descriptors derived with the electronegativity equalization method. *J Med Chem*, 47(21):5311–7, Oct 2004.

[1489] T. Sørlie, C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, T. Thorsen, H. Quist, J. C. Matese, P. O. Brown, D. Botstein, P. Eystein Lønning, and A. L. Børresen-Dale. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA*, 98(19):10869–10874, Sep 2001.

[1490] E. M. Southern. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.*, 98:503–517, 1975.

[1491] Edwin Southern, Kalim Mir, and Mikhail Shchepinov. Molecular interactions on microarrays. *Nature Genetics*, 21:5–9, 1999.

[1492] P.T. Spellman, G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein, and B. Futcher. Comprehensive Identification of Cell Cycle-regulated Genes of the Yeast Saccharomyces cerevisiae by Microarray Hybridization. *Mol. Biol. Cell*, 9:3273–3297, 1998.

[1493] L Van Speybroeck. From epigenesis to epigenetics: The case of C. H. Waddington. *Annals of the New York Academy of Sciences*, 981:61–81, 2002.

[1494] N. Srebro and T. Jaakkola. Weighted low-rank approximations. In T. Fawcett and N. Mishra, editors, *Proceedings of the Twentieth International Conference on Machine Learning*, pages 720–727. AAAI Press, 2003.

[1495] N. Srebro, J. D. M. Rennie, and T. S. Jaakkola. Maximum-margin matrix factorization. In L. K. Saul, Y. Weiss, and L. Bottou, editors, *Adv. Neural. Inform. Process Syst. 17*, pages 1329–1336, Cambridge, MA, 2005. MIT Press.

[1496] Nathan Srebro and Adi Shraibman. Rank, trace-norm and max-norm. In *COLT*, pages 545–560, 2005.

[1497] A. Srebrow and A. R. Kornblihtt. The connection between splicing and cancer. *Journal of Cell Science*, 119:2635–2641, 2006.

[1498] V. K. Srivastava and T. D. Dwivedi. Estimation of seemingly unrelated regression equations : A brief survey. *Journal of Econometrics*, 10(1):15–32, April 1979.

[1499] P. F. Stadler. Landscapes and Their Correlation Functions. *J. Math. Chem.*, 20:1–45, 1996.

[1500] P. F. Stadler. Spectral landscape theory. In J.P. Crutchfield and P. Schuster, editors, *Evolutionary Dynamics — Exploring the Interplay of Selection, Neutrality, Accident and Function.* Oxford University Press, New York, 1999.

[1501] Florence L Stahura and Jrgen Bajorath. Virtual screening methods that complement HTS. *Comb Chem High Throughput Screen*, 7(4):259–69, Jun 2004.

[1502] B.J. Stapley, L.A. Kelley, and M.J. Sternberg. Predicting the subcellular location of proteins from text using support vector machines. In Russ B. Altman, A. Keith Dunker, Lawrence Hunter, Kevin Lauerdale, and Teri E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing 2002*, pages 374–385. World Scientific, 2002.

[1503] V. Starkuviene and R. Pepperkok. The potential of high-content high-throughput microscopy in drug discovery. *Br. J. Pharmacol.*, 152(1):62–71, Sep 2007.

[1504] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, and S. Levy. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. *Bioinformatics*, 2005. To appear.

[1505] Alexander Statnikov, Constantin F Aliferis, and Ioannis Tsamardinos. Methods for multi-category cancer diagnosis from gene expression data: a comprehensive evaluation to inform decision support system development. *Medinfo*, 11(Pt 2):813–7, 2004.

[1506] Guido Steiner, Laura Suter, Franziska Boess, Rodolfo Gasser, Maria Cristina de Vera, Silvio Albertini, and Stefan Ruepp. Discriminating different classes of toxicants by transcript profiling. *Environ. Health Perspect.*, 112(12):1236–48, Aug 2004.

[1507] I. Steinwart. On the influence of the kernel on the consistency of support vector machines. *J. Mach. Learn. Res.*, 2:67–93, 2001.

[1508] I. Steinwart. Support Vector Machines are Universally Consistent. *J. Complexity*, 18:768–791, 2002.

[1509] I. Steinwart. Sparseness of Support Vector Machines. *J. Mach. Learn. Res.*, 4:1071–1105, 2003.

[1510] I. Steinwart. Consistency of support vector machines and other regularized kernel classifiers. *IEEE Trans. Inform. Theory*, 51(1):128–142, 2005.

[1511] I. Steinwart, D. Hush, and C. Scovel. An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels. Technical Report LA-UR 04-8274, Los Alamos National Laboratory, 2004.

[1512] I. Steinwart, D. Hush, and C. Scovel. A classification framework for anomaly detection. *J. Mach. Learn. Res.*, 6:211–232, 2005.

[1513] I. Steinwart, D. Hush, and C. Scovel. Density Level Detection is Classification. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems 17*. MIT Press, Cambridge, MA, 2005.

[1514] I. Steinwart and C. Scovel. Fast Rates for Support Vector Machines using Gaussian Kernels. Technical Report LA-UR 04-8796, Los Alamos National Laboratory, 2004.

[1515] C.J. Stone. Consistent nonparametric regression. *Ann. Stat.*, 8:1348–1360, 1977.

[1516] N. Stransky, C. Vallot, F. Reyal, I. Bernard-Pierrot, S. G. Diez de Medina, R. Segraves, Y. de Rycke, P. Elvin, A. Cassidy, C. Spraggon, A. Graham, J. Southgate, B. Asselain, Y. Allory, C. C. Abbou, D. G. Albertson, J.-P. Thiery, D. K. Chopin, D. Pinkel, and F. Radvanyi. Regional copy number-independent deregulation of transcription in cancer. *Nat. Genet.*, 38(12):1386–1396, Dec 2006.

[1517] M.R. Stratton and N. Rahman. The emerging landscape of breast cancer susceptibility. *Nature Genetics*, 40:17–22, 2008.

[1518] Daniel J Strauss, Wolfgang Delb, and Peter K Plinkert. Objective detection of the central auditory processing disorder: a new machine learning approach. *IEEE Trans Biomed Eng*, 51(7):1147–55, Jul 2004.

[1519] S. S. Strogatz. Exploring complex networks. *Nature*, 410:268–276, 2001.

[1520] Alexander Sturn, John Quackenbush, and Zlatko Trajanoski. Genesis: cluster analysis of microarray data. *Bioinformatics*, 18(1):207–8, Jan 2002.

[1521] A. I. Su, J. B. Welsh, L. M. Sapinoso, S. G. Kern, P. Dimitrov, H. Lapp, P. G. Schultz, S. M. Powell, C. A. Moskaluk, H. F.Jr. Frierson, and G. M. Hampton. Molecular Classification of Human Carcinomas by Use of Gene Expression Signatures. *Cancer Res.*, 61(20):7388–7393, 2001.

[1522] Yang Su, T.M. Murali, Vladimir Pavlovic, Michael Schaffer, and Simon Kasif. RankGene: identification of diagnostic genes based on expression data. *Bioinformatics*, 19(12):1578–1579, 2003.

[1523] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, 102(43):15545–15550, Oct 2005.

[1524] M. Sultan, D. A. Wigle, C. A. Cumbaa, M. Maziarz, J. Glasgow, M. S. Tsao, and I. Jurisica. Binary tree-structured vector quantization approach to clustering and visualizing microarray data. *Bioinformatics*, 18 Suppl 1:S111–9, 2002.

[1525] Y.F. Sun, X.D. Fan, and Y.D. Li. Identifying splicing sites in eukaryotic RNA: support vector machine approach. *Comput. Biol. Med.*, 33(1):17–29, 2003.

[1526] Zhenghong Sun, Xiaoli Fu, Lu Zhang, Xiaoli Yang, Feizhou Liu, and Gengxi Hu. A protein chip system for parallel analysis of multi-tumor markers and its application in cancer detection. *Anticancer Res*, 24(2C):1159–65, 2004.

[1527] R. M. Surabhi and R. B. Gaynor. RNA interference directed against viral and cellular targets inhibits human immunodeficiency Virus Type 1 replication. *J. Virol.*, 76(24):12963–12973, Dec 2002.

[1528] Jean-Sebastien Surgand, Jordi Rodrigo, Esther Kellenberger, and Didier Rognan. A chemogenomic analysis of the transmembrane binding cavity of human g-protein-coupled receptors. *Proteins*, 62(2):509–538, Feb 2006.

[1529] Edward H. Sussenguth. A graph-theoretic algorithm for matching chemical structures. *J. Chem. Doc.*, 5(1):36–43, 1963.

[1530] J. J. Sutherland, L. A. O'Brien, and D. F. Weaver. Spline-fitting with a genetic algorithm: a method for developing classification structure-activity relationships. *J. Chem. Inf. Comput. Sci.*, 43(6):1906–1915, 2003.

[1531] S. Suthram, T. Sittler, and T. Ideker. The plasmodium protein network diverges from those of other eukaryotes. *Nature*, 438(7064):108–112, Nov 2005.

[1532] J. A. Suykens, J. Vandewalle, and B. De Moor. Optimal control by least squares support vector machines. *Neural Netw*, 14(1):23–35, Jan 2001.

[1533] S. J. Swamidass, J. Chen, J. Bruand, P. Phung, L. Ralaivola, and P. Baldi. Kernels for small molecules and the prediction of mutagenicity, toxicity and anti-cancer activity. *Bioinformatics*, 21(Suppl. 1):i359–i368, Jun 2005.

[1534] M. Szafranski, Y. Grandvalet, and A. Rakotomamonjy. Composite kernel learning. In *International Conference on Machine Learning (ICML 2008)*, Helsinki Finlande, 07 2008.

[1535] Marie Szafranski, Yves Grandvalet, and Pierre Morizet-Mahoudeaux. Hierarchical penalization. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 1457–1464. MIT Press, Cambridge, MA, 2008.

[1536] Nobuhiro Takahashi, Mitsuaki Yanagida, Sally Fujiyama, Toshiya Hayano, and Toshiaki Isobe. Proteomic snapshot analyses of preribosomal ribonucleoprotein complexes formed at various stages of ribosome biogenesis in yeast and mammalian cells. *Mass Spectrom Rev*, 22(5):287–317, 2003.

[1537] Norikazu Takahashi and Tetsuo Nishi. Rigorous proof of termination of SMO algorithm for support vector machines. *IEEE Trans Neural Netw*, 16(3):774–6, May 2005.

[1538] Y. Takaoka, Y. Endo, S. Yamanobe, H. Kakinuma, T. Okubo, Y. Shimazaki, T. Ota, S. Sumiya, and K. Yoshikawa. Development of a method for evaluating drug-likeness and ease of synthesis using a data set in which compounds are assigned scores based on chemists' intuition. *J Chem Inf Comput Sci*, 43(4):1269–75, 2003.

[1539] Koichi Takeuchi and Nigel Collier. Bio-medical entity extraction using support vector machines. *Artif. Intell. Med.*, 33(2):125–37, Feb 2005.

[1540] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. I.H.E.S.*, 81:73–203, 1995.

[1541] M. Talagrand. Majorizing measures: The generic chaining. *Ann. Probab.*, 24:1049–1103, 1996.

[1542] M. Talagrand. New concentration inequalities for product spaces. *Inventionnes Math.*, 126:505–563, 1996.

[1543] M. Talagrand. A New Look at Independence. *Ann. Probab.*, 24:1–34, 1996.

[1544] A. Talukder and D. Casasent. A closed-form neural network for discriminatory feature extraction from high-dimensional data. *Neural Netw*, 14(9):1201–18, Nov 2001.

[1545] G. Tang. siRNA and miRNA: an insight into RISCs. *Trends Biochem. Sci.*, 30(2):106–14, Feb 2005.

[1546] Thomas Tang, Jinbo Xu, and Ming Li. Discovering sequence-structure motifs from protein segments and two applications. *Pac Symp Biocomput*, pages 370–81, 2005.

[1547] Yuchun Tang, Bo Jin, and Yan-Qing Zhang. Granular support vector machines with association rules mining for protein homology prediction. *Artif. Intell. Med.*, Jul 2005.

[1548] Toshiyasu Tarumi, Gary W Small, Roger J Combs, and Robert T Kroutil. Remote detection of heated ethanol plumes by airborne passive Fourier transform infrared spectrometry. *Appl Spectrosc*, 57(11):1432–41, Nov 2003.

[1549] B. Taskar, C. Guestrin, and D. Koller. Max-Margin Markov Networks. In Sebastian Thrun, Lawrence Saul, and Bernhard Schölkopf, editors, *Advances in Neural Information Processing Systems 16*, Cambridge, MA, 2004. MIT Press.

[1550] S. Tavazoie, J. D. Hughes, M. J. Campbell, R. J. Cho, and G. M. Church. Systematic determination of genetic network architecture. *Nat. Genet.*, 1999.

[1551] D. M. J. Tax and R. P. W. Duin. Uniform Object Generation for Optimizing One-class Classifiers. *J. Mach. Learn. Res.*, 2:155–173, 2001.

[1552] W. R. Taylor. Protein structure comparison using bipartite graph matching and its application to protein structure classification. *Mol Cell Proteomics*, 1(4):334–339, April 2002.

[1553] J. Tegner, M. K. S. Yeung, J. Hasty, and J. J. Collins. Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling. *Proc. Natl. Acad. Sci. USA*, 100(10):5944–5949, May 2003.

[1554] R. D. Teixeira, A. P. Braga, R. H. Takahashi, and R. R. Saldanha. Recent advances in the MOBJ algorithm for training artificial neural networks. *Int J Neural Syst*, 11(3):265–70, Jun 2001.

[1555] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–23, Dec 2000.

[1556] Reiji Teramoto, Mikio Aoki, Toru Kimura, and Masaharu Kanaoka. Prediction of siRNA functionality using generalized string kernel and support vector machine. *FEBS Lett.*, 579(13):2878–82, May 2005.

[1557] Thorsten Thies and Frank Weber. Optimal reduced-set vectors for support vector machines with a quadratic kernel. *Neural Comput*, 16(9):1769–77, Sep 2004.

[1558] M. Thimm, A. Goede, S. Hougardy, and R. Preissner. Comparison of 2D similarity and 3D superposition. Application to searching a conformational drug database. *J Chem Inf Comput Sci*, 44(5):1816–1822, 2004.

[1559] Uwe Thissen, Blent Ustn, Willem J Melssen, and Lutgarde M C Buydens. Multivariate calibration with least-squares support vector machines. *Anal Chem*, 76(11):3099–105, Jun 2004.

[1560] R. Thomas and M. Kaufman. Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback circuits. *Chaos*, 11(1):180–195, 2001.

[1561] Sebastian Thrun and Lorien Pratt, editors. *Learning to learn.* Kluwer Academic Publishers, Norwell, MA, USA, 1998.

[1562] Sushil K Thukral, Paul J Nordone, Rong Hu, Leah Sullivan, Eric Galambos, Vincent D Fitzpatrick, Laura Healy, Michael B Bass, Mary E Cosenza, and Cynthia A Afshari. Prediction of nephrotoxicant action and identification of candidate toxicity-related biomarkers. *Toxicol Pathol*, 33(3):343–55, 2005.

[1563] Liang Tian and Afzel Noore. A novel approach for short-term load forecasting using support vector machines. *Int J Neural Syst*, 14(5):329–35, Oct 2004.

[1564] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc. B.*, 58(1):267–288, 1996.

[1565] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 67(1):91–108, 2005.

[1566] R. Tibshirani and P. Wang. Spatial smoothing and hot spot detection for cgh data using the fused lasso. *Biostatistics (Oxford, England)*, 9(1):18–29, January 2008.

[1567] N. Tiffin, J. F. Kelso, A. R. Powell, H. Pan, V. B. Bajic, and W. A. Hide. Integration of text- and data-mining using ontologies successfully selects disease gene candidates. *Nucleic Acids Res.*, 33(5):1544–1552, 2005.

[1568] A.N. Tikhonov. On the stability of inverse problems. *Doklady Akademii nauk SSSR*, 39(5):195–198, 1943.

[1569] A.N. Tikhonov. Solution of incorrectly problems and the regularization method. *Soviet Mathematics Doklady*, 4:1035–1038, 1963.

[1570] A.N. Tikhonov and V.Y. Arsenin. *Solutions of ill-posed problems.* W.H. Winston, Washington, D.C., 1977.

[1571] T.J. Tjalkens and F.M.J. Willems. A universal variable-to-fixed length source code based on Lawrence's algorithm. *IEEE Trans. Inform. Theory*, 38(2):247–253, Mar 1992.

[1572] Tj.J. Tjalkens and F. M. J. Willems. Implementing the Context-Tree Weighting Method: Arithmetic Coding,. In *Int. Conf. on Combinatorics, Information Theory and Statistics*, page 83, Portland, Maine, U.S.A, 18-20 1997.

[1573] M. Tobita, T. Nishikawa, and R. Nagashima. A discriminant model constructed by the support vector machine method for HERG potassium channel inhibitors. *Bioorg. Med. Chem. Lett.*, 15(11):2886–90, Jun 2005.

[1574] R. Todeschini and V. Consonni. *Handbook of Molecular Descriptors.* Wiley-VCH, New York, 2002.

[1575] Y. Tomari and P. D. Zamore. Perspective: machines for RNAi. *Genes Dev.*, 19(5):517–29, Mar 2005.

[1576] Scott A Tomlins, Rohit Mehra, Daniel R Rhodes, Xuhong Cao, Lei Wang, Saravana M Dhanasekaran, Shanker Kalyana-Sundaram, John T Wei, Mark A Rubin, Kenneth J Pienta, Rajal B Shah, and Arul M Chinnaiyan. Integrative molecular concept modeling of prostate cancer progression. *Nat Genet*, 39(1):41–51, Jan 2007.

[1577] Martin Tompa, Nan Li, Timothy L Bailey, George M Church, Bart De Moor, Eleazar Eskin, Alexander V Favorov, Martin C Frith, Yutao Fu, W James Kent, Vsevolod J Makeev, Andrei A Mironov, William Stafford Noble, Giulio Pavesi, Graziano Pesole, Mireille Rgnier, Nicolas Simonis, Saurabh Sinha, Gert Thijs, Jacques Van Helden, Mathias Vandenbogaert, Zhiping Weng, Christopher Workman, Chun Ye, and Zhou Zhu. Assessing computational tools for the discovery of transcription factor binding sites. *Nature Biotechnology*, 23:137–144, 2005.

[1578] J. C. Tong, G. L. Zhang, T. W. Tan, J. T. August, V. Brusic, and S. Ranganathan. Prediction of HLA-DQ3.2beta ligands: evidence of multiple registers in class II binding peptides. *Bioinformatics*, 22(10):1232–1238, May 2006.

[1579] Joo Chuan Tong, Tin Wee Tan, and Shoba Ranganathan. In silico grouping of peptide/hla class i complexes using structural interaction characteristics. *Bioinformatics*, 23(2):177–183, Jan 2007.

[1580] Richard W Tothill, Adam Kowalczyk, Danny Rischin, Alex Bousioutas, Izhak Haviv, Ryan K van Laar, Paul M Waring, John Zalcberg, Robyn Ward, Andrew V Biankin, Robert L Sutherland, Susan M Henshall, Kwun Fong, Jonathan R Pollack, David D L Bowtell, and Andrew J Holloway. An expression-based site of origin diagnostic method designed for clinical application to cancer of unknown origin. *Cancer Res.*, 65(10):4031–40, May 2005.

[1581] Joel A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50:2231–2242, 2004.

[1582] Joel A. Tropp, Anna C. Gilbert, and Martin J. Strauss. Algorithms for simultaneous sparse approximation: part i: Greedy pursuit. *Signal Process.*, 86(3):572–588, 2006.

[1583] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman. Missing value estimation methods for DNA microarrays. *Bioinformatics*, 17:520–525, 2001.

[1584] O. G. Troyanskaya, K. Dolinski, A. B. Owen, R. B. Altman, and D. Botstein. A bayesian framework for combining heterogeneous data sources for gene function prediction (in *saccharomyces cerevisiae*). *Proc. Natl. Acad. Sci. USA*, 100(14):8348–8353, 2003.

[1585] M. Truss, M. Swat, S. M. Kielbasa, R. Schäfer, H. Herzel, and C. Hagemeier. HuSiDa–the human siRNA database: an open-access database for published functional siRNA sequences and technical details of efficient transfer into recipient cells. *Nucleic Acids Res.*, 33(Database issue):D108–11, Jan 2005.

[1586] C.A. Tsai, C.H. Chen, T.C. Lee, I.C. Ho, U.C. Yang, and J.J. Chen. Gene selection for sample classifications in microarray experiments. *DNA Cell Biol.*, 23(10):607–614, 2004.

[1587] W.H. Tsai and K.S. Fu. Error-correcting isomorphisms of attributed relational graphs for pattern analysis. *SMC*, 9(12):757–768, December 1979.

[1588] I. W. Tsang and J. T. Kwok. Distance metric learning with kernels. In *Proceedings of the International Conference on Artificial Neural Networks*, pages 126–129, 2003.

[1589] A. Tsirigos and I. Rigoutsos. A sensitive, support-vector-machine method for the detection of horizontal gene transfers in viral, archaeal and bacterial genomes. *Nucleic Acids Res.*, 33(12):3699–707, 2005.

[1590] I. Tsochantaridis, T. Hofmann, T. Joachims, and Y. Altun. Support vector machine learning for interdependent and structured output spaces. In *Twenty-first international conference on Machine learning*. ACM Press, 2004.

[1591] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large margin methods for structured and interdependent output variables. *J. Mach. Learn. Res.*, 6:1453–1484, 2005.

[1592] K. Tsuda. Entire regularization path for graph data. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pages 919–926, New York, NY, USA, 2007. ACM.

[1593] K. Tsuda, S. Akaho, and K. Asai. The em Algorithm for Kernel Matrix Completion with Auxiliary Data. *J. Mach. Learn. Res.*, 4:67–81, 2003.

[1594] K. Tsuda, M. Kawanabe, G. Rätsch, S. Sonnenburg, and K.-R. Müller. A new discriminative kernel from probabilistic models. *Neural Computation*, 14(10):2397–2414, 2002.

[1595] K. Tsuda, T. Kin, and K. Asai. Marginalized Kernels for Biological Sequences. *Bioinformatics*, 18:S268–S275, 2002.

[1596] K. Tsuda and W.S. Noble. Learning kernels from biological networks by maximizing entropy. *Bioinformatics*, 20:i326–i333, 2004.

[1597] Daisuke Tsujinishi and Shigeo Abe. Fuzzy least squares support vector machines for multiclass problems. *Neural Netw*, 16(5-6):785–92, 2003.

[1598] A. B. Tsybakov. On Nonparametric Estimation of Density Level Sets. *Ann. Stat.*, 25:948–969, June 1997.

[1599] A. B. Tsybakov. *Introduction à l'estimation non-paramétrique*. Springer, 2004.

[1600] D. L. Tucker, N. Tucker, and T. Conway. Gene expression profiling of the ph response in escherichia coli. *J Bacteriol.*, 184(23):6551–6558, Dec 2002.

[1601] Nihal Tugcu, Asif Ladiwala, Curt M Breneman, and Steven M Cramer. Identification of chemically selective displacers using parallel batch screening experiments and quantitative structure efficacy relationship models. *Anal Chem*, 75(21):5806–16, Nov 2003.

[1602] Nihal Tugcu, Minghu Song, Curt M Breneman, N. Sukumar, Kristin P Bennett, and Steven M Cramer. Prediction of the effect of mobile-phase salt type on protein retention and selectivity in anion exchange systems. *Anal Chem*, 75(14):3563–72, Jul 2003.

[1603] Chun-Wei Tung and Shinn-Ying Ho. Popi: predicting immunogenicity of mhc class i binding peptides by mining informative physicochemical properties. *Bioinformatics*, 23(8):942–949, Apr 2007.

[1604] W. L. Tung and C. Quek. GenSo-FDSS: a neural-fuzzy decision support system for pediatric ALL cancer subtype identification using gene expression data. *Artif. Intell. Med.*, 33(1):61–88, Jan 2005.

[1605] B. A. Turlach, W. N. Venables, and S. J. Wright. Simultaneous variable selection. *Technometrics*, 47(3):349–363, 2005.

[1606] E. Turlin, M. Perrotte-piquemal, A. Danchin, and F. Biville. Regulation of the early steps of 3-phenylpropionate catabolism in Escherichia coli. *J. Mol. Microbiol. Biotechnol.*, 3(1):127–133, Jan 2001.

[1607] Bryan M. Turner. Cellular memory and the histone code. *Cell*, 111:285–291, 2002.

[1608] F. S. Turner, D. R. Clutterbuck, and C. A. M. Semple. Pocus: mining genomic sequence annotation to predict disease genes. *Genome Biol.*, 4(11):R75, 2003.

[1609] T. Tuschl, P.D. Zamore, R. Lehmann, D.P. Bartel, and P.A. Sharp. Targeted mRNA degradation by double-stranded RNA in vitro. *Genes Dev.*, 13(24):3191–7, Dec 1999.

[1610] Huey-Ming Tzeng, Jer-Guang Hsieh, and Yih-Lon Lin. Predicting nurses' intention to quit with a support vector machine: a new approach to set up an early warning mechanism in human resource management. *Comput Inform Nurs*, 22(4):232–42, 2004.

[1611] N. Ueda, K. F. Aoki-Kinoshita, A. Yamaguchi, T. Akutsu, and H. Mamitsuka. A Probabilistic Model for Mining Labeled Ordered Trees: Capturing Patterns in Carbohydrate Sugar Chains. *IEEE Transactions on Knowledge and Data Engineering*, 17(8):1051–1064, 2005.

[1612] P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleish, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J. M. Rothberg. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature*, 403:623–627, 2000.

[1613] K. Ui-Tei, Y. Naito, F. Takahashi, T. Haraguchi, H. Ohki-Hamazaki, A. Juni, R. Ueda, and K. Saigo. Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res.*, 32(3):936–948, Feb 2004.

[1614] J. R. Ullmann. An algorithm for subgraph isomorphism. *J. ACM*, 23(1):31–42, 1976.

[1615] S. Umeyama. An eigendecomposition approach to weighted graph matching problems. *IEEE Trans. Pattern Anal. Mach. Intell.*, 10(5):695–703, Sept. 1988.

[1616] Jayant S Vaidya. Breast cancer: an artistic view. *The Lancet Oncology*, 8:583–585, 2007.

[1617] G. Valentini. Gene expression data analysis of human lymphoma using support vector machines and output coding ensembles. *Artif. Intell. Med.*, 26(3):281–304, Nov 2002.

[1618] Anirudh Vallabhaneni and Bin He. Motor imagery task classification for brain computer interface applications using spatiotemporal principle component analysis. *Neurol Res*, 26(3):282–7, Apr 2004.

[1619] E. van Beers and P. Nederlof. Array-CGH and breast cancer. *Breast Cancer Research*, 8(3):210, 2006.

[1620] M. J. van de Vijver, Y. D. He, L. J. van't Veer, H. Dai, A. A. M. Hart, D. W. Voskuil, G. J. Schreiber, J. L. Peterse, C. Roberts, M. J. Marton, M. Parrish, D. Atsma, A. Witteveen, A. Glas, L. Delahaye, T. van der Velde, H. Bartelink, S. Rodenhuis, E. T. Rutgers, S. H. Friend, and R. Bernards. A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.*, 347(25):1999–2009, Dec 2002.

[1621] H. van der Holst. *Topological and spectral graph characterizations*. PhD thesis, Universiteit van Amsterdam, 1996.

[1622] J. van Helden, D. Gilbert, L. Wernisch, M. Schroeder, and S. J. Wodak. Application of regulatory sequence analysis and metabolic network analysis to the interpretation of gene expression data. In *JOBIM '00: Selected papers from the First International Conference on Computational Biology, Biology, Informatics, and Mathematics*, pages 147–164, London, UK, 2001. Springer-Verlag.

[1623] L. J. van 't Veer, H. Dai, M. J. van de Vijver, Y. D. He, A. A. M. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Witteveen, G. J. Schreiber, R. M. Kerkhoven, C. Roberts, P. S. Linsley, R. Bernards, and S. H. Friend. Gene expression profiling predicts clinical outcome of breast cancers. *Nature*, 415(6871):530–536, Jan 2002.

[1624] Laura J. Van't Veer and Ren Bernards. Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature*, 452:564–570, 2008.

[1625] V. Vapnik and O. Chapelle. Bounds on error expectation for support vector machines. *Neural Comput*, 12(9):2013–36, Sep 2000.

[1626] V. N. Vapnik. *Statistical Learning Theory.* Wiley, New-York, 1998.

[1627] Vladimir N. Vapnik. *The nature of statistical learning theory.* Springer-Verlag New York, Inc., New York, NY, USA, 1995.

[1628] V.N. Vapnik and A. Ya. Chervonenkis. Teoriya raspoznavaniya obrazov: Statisticheskie problemy obucheniya. (Russian) [Theory of Pattern Recognition: Statistical Problems of Learning]. Moscow: Nauka, 1974.

[1629] A. Varki, R. Cummings, J. Esko, H. Freeze, G. Hart, and J. Marth. *Essentials of glycobiology.* Cold Spring Harbor Laboratory Press, 1999.

[1630] A. Vazquez, A. Flammini, A. Maritan, and A. Vespignani. Modeling of protein interaction networks. E-print cond-mat/0108043, Aug 2001.

[1631] D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward, and K. D. Kopple. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.*, 45(12):2615–2623, Jun 2002.

[1632] Julian P. Venables. Aberrant and alternative splicing in cancer. *Cancer Research*, 64:7647–7654, 2004.

[1633] J. C. et al. Venter. The Sequence of the Human Genome. *Science*, 291(5507):1304–1351, 2001.

[1634] W. Vercoutere, S. Winters-Hilt, H. Olsen, D. Deamer, D. Haussler, and M. Akeson. Rapid discrimination among individual DNA hairpin molecules at single-nucleotide resolution using an ion channel. *Nat Biotechnol*, 19(3):248–52, Mar 2001.

[1635] L. Vermeulen, M.R. Sprick, K. Kemper, G. Stassi, and J.P. Medema. Cancer stem cells - old concepts, new insights. *Cell Death and Differentiation*, 15:947–58, 2008.

[1636] J.-P. Vert. Double mixture and universal inference. Technical Report DMA-00-15, Ecole Normale Supérieure, 2000.

[1637] J.-P. Vert. Adaptive context trees and text clustering. *IEEE Trans. Inform. Theory*, 47(5):1884–1901, Jul 2001.

[1638] J.-P. Vert. *Statistical Methods for Natural Language Modelling.* PhD thesis, Paris 6 University, 2001.

[1639] J.-P. Vert. Text categorization using adaptive context trees. In A. Gelbukh, editor, *Proceedings of the CICLing-2001 conference*, volume 2004 of *LNCS*, pages 423–436. Springer Verlag, 2001.

[1640] J.-P. Vert. Support vector machine prediction of signal peptide cleavage site using a new class of kernels for strings. In R. B. Altman, A. K. Dunker, L. Hunter, K. Lauerdale, and T. E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing 2002*, pages 649–660. World Scientific, 2002.

[1641] J.-P. Vert. A tree kernel to analyze phylogenetic profiles. *Bioinformatics*, 18:S276–S284, 2002.

[1642] J.-P. Vert. Kernel methods in computational biology. Technical Report ccsd-00012124, CNRS-HAL, Oct 2005.

[1643] J.-P. Vert. Kernel methods in bioinformatics : a survey. To appear, 2006.

[1644] J.-P. Vert. Kernel methods in genomics and computational biology. In G. Camps-Valls, J.-L. Rojo-Alvarez, and M. Martinez-Ramon, editors, *Kernel Methods in Bioengineering, Signal and Image Processing.* IDEA Group, 2007.

[1645] J.-P. Vert. The optimal assignment kernel is not positive definite. Technical Report 0801.4061, Arxiv, 2008.

[1646] J.-P. Vert, N. Foveau, C. Lajaunie, and Y. Vandenbrouck. An accurate and interpretable model for sirna efficacy prediction. *BMC Bioinformatics*, 7:520, 2006.

[1647] J.-P. Vert and L. Jacob. Machine learning for in silico virtual screening and chemical genomics: New strategies. *Combinatorial Chemistry & High Throughput Screening*, 11(8):677–685, September 2008.

[1648] J.-P. Vert and M. Kanehisa. Graph-driven features extraction from microarray data. Technical Report 0206055, Arxiv physics, 2002.

[1649] J.-P. Vert and M. Kanehisa. Extracting active pathways from gene expression data. *Bioinformatics*, 19:238ii–234ii, 2003.

[1650] J.-P. Vert and M. Kanehisa. Graph-driven features extraction from microarray data using diffusion kernels and kernel CCA. In S. Becker, S. Thrun, and K. Obermayer, editors, *Adv. Neural Inform. Process. Syst.*, pages 1449–1456. MIT Press, 2003.

[1651] J. P. Vert, T. Matsui, S. Satoh, and Y. Uchiyama. High-level feature extraction using SVM with walk-based graph kernel. In *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2009)*, 2009.

[1652] J.-P. Vert, J. Qiu, and W. S. Noble. A new pairwise kernel for biological network inference with support vector machines. *BMC Bioinformatics*, 8 Suppl 10:S8, 2007.

[1653] J.-P. Vert, H. Saigo, and T. Akutsu. Local alignment kernels for biological sequences. In B. Schölkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 131–154. MIT Press, The MIT Press, Cambridge, Massachussetts, 2004.

[1654] J.-P. Vert, R. Thurman, and W. S. Noble. Kernels for gene regulatory regions. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Adv. Neural. Inform. Process Syst.*, volume 18, pages 1401–1408, Cambridge, MA, 2006. MIT Press.

[1655] J.-P. Vert, K. Tsuda, and B. Schölkopf. A primer on kernel methods. In B. Schlkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 35–70. MIT Press, 2004.

[1656] J.-P. Vert and Y. Yamanishi. Supervised graph inference. In L. K. Saul, Y. Weiss, and L. Bottou, editors, *Adv. Neural Inform. Process. Syst.*, volume 17, pages 1433–1440. MIT Press, Cambridge, MA, 2005.

[1657] Jean-Philippe Vert, Francis Bach, and Theodoros Evgeniou. Low-rank matrix factorization with attributes, 2006.

[1658] R. Vert and J.-P. Vert. Consistency and convergence rates of one-class SVM and related algorithms. Technical Report 1414, LRI, Université Paris-Sud, 2005.

[1659] R. Vert and J.-P. Vert. Consistency and convergence rates of one-class SVMs and related algorithms. *J. Mach. Learn. Res.*, 7:817–854, 2006.

[1660] T. A. Vickers, S. Koo, C. F. Bennett, S. T. Crooke, N. M. Dean, and B. F. Baker. Efficient reduction of target RNAs by small interfering RNA and RNase H-dependent antisense agents. A comparative analysis. *J. Biol. Chem.*, 278(9):7108–18, Feb 2003.

[1661] A. Vinayagam, R. Knig, J. Moormann, F. Schubert, R. Eils, K.-H. Glatting, and S. Suhai. Applying Support Vector Machines for Gene Ontology based gene function prediction. *BMC Bioinformatics*, 5(1):116, Aug 2004.

[1662] A. Vincent-Salomon, C. Lucchesi, N. Gruel, V. Raynal, G. Pierron, R. Goudefroye, F. Reyal, F. Radvanyi, R. Salmon, J.-P. Thiery, X. Sastre-Garau, B. Sigal-Zafrani, A. Fourquet, and A. Delattre. Integrated genomic and transcriptomic analysis of ductal carcinoma in situ of the breast. *Clin. Cancer Res.*, 14(7):1956–1965, Apr 2008.

[1663] A. Vinokourov, J. Shawe-Taylor, and N. Cristianini. Finding Language-Independent Semantic Representation of Text using Kernel Canonical Correlation Analysis. Technical report, Neurocolt, 2002. NeuroCOLT Technical Report NC-TR-02-119.

[1664] A. Vinokourov, J. Shawe-Taylor, and N. Cristianini. Inferring a semantic representation of text via cross-language correlation analysis. In Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, editors, *Adv. Neural Inform. Process. Syst.* MIT Press, 2003.

[1665] S. V. N. Vishwanathan, N. N. Schraudolph, R. Kondor, and K. M. Borgwardt. Graph kernels. *J. Mach. Learn. Res.*, 10:1–41, 2009.

[1666] S. V. N. Vishwanathan and A. J. Smola. Fast kernels for string and tree matching. In B. Schölkopf, K. Tsuda, and J.-P. Vert, editors, *Kernel methods in computational biology*, pages 113–130. MIT Press, 2004.

[1667] S.V.N. Vishwanathan, K. Borgwardt, and N. Schraudolph. Fast Computation of Graph Kernels. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Adv. Neural Inform. Process. Syst.*, volume 19, pages 1–2, Cambridge, MA, 2007. MIT Press, Cambridge, MA.

[1668] A. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inform. Theory*, 13(2):260–269, 1973.

[1669] Kristian Vlahovicek, Lszl Kajn, Vilmos Agoston, and Sndor Pongor. The SBASE domain sequence resource, release 12: prediction of protein domain-architecture using support vector machines. *Nucleic Acids Res*, 33(Database issue):D223–5, Jan 2005.

[1670] Bert Vogelstein and Kenneth W Kinzler. Cancer genes and the pathways they control. *Nature Medicine*, 10:789–799, 2004.

[1671] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, December 2007.

[1672] C. von Mering, R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields, and P. Bork. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, 417(6887):399–403, May 2002.

[1673] M. Wagener, J. Sadowski, and J. Gasteiger. Autocorrelation of molecular surface properties for modeling. corticosteroid binding globulin and cytosolic. ah. receptor. activity by neural networks. *J. Am. Chem. Soc.*, 117:7769–7775, 1995.

[1674] A. Wagner. The Yeast Protein Interaction Network Evolves Rapidly and Contains Few Redundant Duplicate Genes. *Mol. Biol. Evol.*, 18:1283–1292, 2001.

[1675] M. Wagner, D. Naik, and A. Pothen. Protocols for disease classification from mass spectrometry data. *Proteomics*, 3(9):1692–1698, 2003.

[1676] M. Wagner, D.N. Naik, A. Pothen, S. Kasukurti, R.R. Devineni, B.L. Adam, O.J. Semmes, and G.L. Wright Jr. Computational protein biomarker prediction: a case study for prostate cancer. *BMC Bioinformatics*, 5(26), 2004.

[1677] G. Wahba. *Spline Models for Observational Data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1990.

[1678] Grace Wahba. Soft and hard classification by reproducing kernel Hilbert space methods. *Proc Natl Acad Sci U S A*, 99(26):16524–30, Dec 2002.

[1679] M. J. Wainwright. Sharp thresholds for high-dimensional and noisy recovery of sparsity. Technical Report 709, UC Berkeley, Department of Statistics, 2006.

[1680] T. Walter, J.-C. Klein, P. Massin, and A. Erignay. Detection of the median axis of vessels in retinal images. *European Journal of Ophthalmology*, 13(2), 2003.

[1681] W. Patrick Walters and Mark A. Murcko. Prediction of 'drug-likeliness'. *Adv. Drug Deliv. Rev.*, 54:255–271, 2002.

[1682] Haojun Wang, Chongxun Zheng, Ying Li, Huafeng Zhu, and Xiangguo Yan. Application of support vector machines to classification of blood cells. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*, 20(3):484–7, Sep 2003.

[1683] Hong Fang Wang and Edwin R. Hancock. Correspondence matching using kernel principal components analysis and label consistency constraints. *Pattern Recogn.*, 39(6):1012–1025, 2006.

[1684] J. Wang, W.-K. Sung, A. Krishnan, and K.-B. Li. Protein subcellular localization prediction for Gram-negative bacteria using amino acid subalphabets and a combination of multiple support vector machines. *BMC Bioinformatics*, 6(1):174, Jul 2005.

[1685] J. F. Wang, C. Z. Cai, C. Y. Kong, Z. W. Cao, and Y. Z. Chen. A computer method for validating traditional Chinese medicine herbal prescriptions. *Am J Chin Med*, 33(2):281–97, 2005.

[1686] Kai Wang, Ekachai Jenwitheesuk, Ram Samudrala, and John E Mittler. Simple linear model provides highly accurate genotypic predictions of HIV-1 drug resistance. *Antivir Ther*, 9(3):343–52, Jun 2004.

[1687] Long-Hui Wang, Juan Liu, Yan-Fu Li, and Huai-Bei Zhou. Predicting protein secondary structure by a support vector machine based on a new coding scheme. *Genome Inform Ser Workshop Genome Inform*, 15(2):181–90, 2004.

[1688] M. Wang, J. Yang, and K-C. Chou. Using string kernel to predict signal peptide cleavage site based on subsite coupling model. *Amino Acids*, 28(4):395–402, Jun 2005.

[1689] M. Wang, J. Yang, G.-P. Liu, Z.-J. Xu, and K.-C. Chou. Weighted-support vector machines for predicting membrane protein types based on pseudo-amino acid composition. *Protein Eng. Des. Sel.*, 17(6):509–516, 2004.

[1690] M-L. Wang, W-J. Li, M-L. Wang, and W-B. Xu. Support vector machines for prediction of peptidyl prolyl cis/trans isomerization. *J Pept Res*, 63(1):23–8, Jan 2004.

[1691] Ming-Lei Wang, Hui Yao, and Wen-Bo Xu. Prediction by support vector machines and analysis by Z-score of poly-L-proline type II conformation based on local sequence. *Comput. Biol. Chem.*, 29(2):95–100, Apr 2005.

[1692] R. F. Wang. Human tumor antigens: implications for cancer vaccine development. *J. Mol. Med.*, 77(9):640–655, Sep 1999.

[1693] S. Wang, H. Li, F. Qi, and Y. Zhao. Objective facial paralysis grading based on Pface and eigenflow. *Med Biol Eng Comput*, 42(5):598–603, Sep 2004.

[1694] Y. Wang, J.G.M. Klijn, Y. Zhang, A.M. Sieuwerts, M.P. Look, F. Yang, D. Talantov, M. Timmermans, M.E. Meijer-van Gelder, J. Yu, T. Jatkoe, E.M.J.J. Berns, D. Atkins, and J.A. Foekens. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*, 365(9460):671–679, 2005.

[1695] Yongmei Michelle Wang, Robert T Schultz, R. Todd Constable, and Lawrence H Staib. Nonlinear estimation and modeling of fMRI data using spatio-temporal support vector regression. *Inf Process Med Imaging*, 18:647–59, Jul 2003.

[1696] Yu Wang, Igor V Tetko, Mark A Hall, Eibe Frank, Axel Facius, Klaus F X Mayer, and Hans W Mewes. Gene selection from microarray data for cancer classification–a machine learning approach. *Comput. Biol. Chem.*, 29(1):37–46, Feb 2005.

[1697] Yuhang Wang, Fillia Makedon, and James Ford. A bipartite graph matching framework for finding correspondences between structural elements in two proteins. In *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2004.

[1698] Zhong Wang, Mark Gerstein, and Michael Snyder. Rna-seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10(1):57–63, Jan 2009.

[1699] J. J. Ward, L. J. McGuffin, B. F. Buxton, and D. T. Jones. Secondary structure prediction with support vector machines. *Bioinformatics*, 19(13):1650–1655, 2003.

[1700] Christopher A Waring and Xiuwen Liu. Face detection using spectral histograms and SVMs. *IEEE Trans Syst Man Cybern B Cybern*, 35(3):467–76, Jun 2005.

[1701] Jeffrey F Waring, Roger G Ulrich, Nick Flint, David Morfitt, Arno Kalkuhl, Frank Staedtler, Michael Lawton, Johanna M Beekman, and Laura Suter. Interlaboratory evaluation of rat hepatic gene expression changes induced by methapyrilene. *Environ Health Perspect*, 112(4):439–48, Mar 2004.

[1702] J. W. Warmke and B. Ganetzky. A family of potassium channel genes related to eag in Drosophila and mammals. *Proc Natl Acad Sci U S A*, 91(8):3438–3442, Apr 1994.

[1703] M. K. Warmuth, J. Liao, G. Rätsch, M. Mathieson, S. Putta, and C. Lemmen. Active learning with support vector machines in the drug discovery process. *J Chem Inf Comput Sci*, 43(2):667–673, 2003.

[1704] M. K. Warmuth, G. Rätsch, M. Mathieson, L. Liao, and C. Lemmen. Active learning in the drug discovery process. In T.G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Adv. Neural Inform. Process. Syst.*, volume 14, pages 1449–1456. MIT Press, 2002.

[1705] G.L. Warren, C.W. Andrews, A.M. Capelli, B. Clarke, J. LaLonde, M.H. Lambert, M. Lindvall, N. Nevins, S.F. Semus, S. Senger, G. Tedesco, I.D. Wall, J.M. Woolven, C.E. Peishoff, and M.S. Head. A critical assessment of docking programs and scoring functions. *J Med Chem*, 49(20):5912–5931, Oct 2006.

[1706] T. Watanabe, Y. Totoki, A. Toyoda, M. Kaneda, S. Kuramochi-Miyagawa, Y. Obata, H. Chiba, Y. Kohara, T. Kono, T. Nakano, M.A. Surani, Y. Sakaki, and H. Sasaki. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, 453:539–543, 2008.

[1707] Scott R Waterman and P. L C Small. Transcriptional expression of escherichia coli glutamate-dependent acid resistance genes gada and gadbc in an hns rpos mutant. *J Bacteriol*, 185(15):4644–4647, Aug 2003.

[1708] C. Watkins. Dynamic alignment kernels. In A.J. Smola, P.L. Bartlett, B. Schölkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*, pages 39–50. MIT Press, Cambridge, MA, 2000.

[1709] J. D. Watson and F. H. C. Crick. A Structure for Deoxyribose Nucleic Acid. *Nature*, 171:737, 1953.

[1710] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[1711] E. A. Weathers, M. E. Paulaitis, T. B. Woolf, and J. H. Hoh. Reduced amino acid alphabet is sufficient to accurately recognize intrinsically disordered protein. *FEBS Lett.*, 576(3):348–352, 2004.

[1712] B. L. Weber. Cancer genomics. *Cancer Cell*, 1(1):37–47, 2002.

[1713] Griffin Weber, Staal Vinterbo, and Lucila Ohno-Machado. Building an asynchronous web-based tool for machine learning classification. *Proc AMIA Symp*, pages 869–73, 2002.

[1714] Liyang Wei, Yongyi Yang, Robert M Nishikawa, and Yulei Jiang. A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications. *IEEE Trans Med Imaging*, 24(3):371–80, Mar 2005.

[1715] Nathanael Weill and Didier Rognan. Development and Validation of a Novel Protein– Ligand Fingerprint To Mine Chemogenomic Space: Application to G Protein-Coupled Receptors and Their Ligands. *Journal of Chemical Information and Modeling*, 49(4):1049–1062, 2009.

[1716] R A Weinberg. *The biology of cancer.* Garland Science, Taylor & Francis Group, LLC, 2007.

[1717] K. Weinberger, B. Packer, and L. Saul. Nonlinear Dimensionality Reduction by Semidefinite Programming and Kernel Matrix Factorization. In R. G. Cowell and Z. Ghahramani, editors, *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics, Jan 6-8, 2005, Savannah Hotel, Barbados*, pages 381–388. Society for Artificial Intelligence and Statistics, 2005.

[1718] K. Q. Weinberger, J. Blitzer, and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. In Y. Weiss, B. Schoelkopf, and J. Platt, editors, *Adv. Neural. Inform. Process Syst.*, number 18, Cambridge, MA, 2006. MIT Press.

[1719] K. Q. Weinberger and L. K. Saul. Unsupervised Learning of Image Manifolds by Semidefinite Programming. In *2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, volume 2, pages 988–995, 2004.

[1720] K. Q. Weinberger, F. Sha, and L. K. Saul. Learning a kernel matrix for nonlinear dimensionality reduction. In *ICML '04: Twenty-first international conference on Machine learning*, New York, NY, USA, 2004. ACM Press.

[1721] M. J. Weinberger, N. Merhav, and M. Feder. Optimal sequential probability assignment for individual sequences. *IEEE Trans. Inform. Theory*, 40(2):384–396, Mar 1994.

[1722] M. J. Weinberger, J. J. Rissanen, and M. Feder. A universal finite memory source. *IEEE Trans. Inform. Theory*, 41(3):643–652, May 1995.

[1723] S. R. Wellings and H. M. Jensen. On the origin and progression of ductal carcinoma in the human breast. *J. Natl. Cancer Inst.*, 50(5):1111–1118, May 1973.

[1724] Nils Weskamp, Eyke Hullermeier, Daniel Kuhn, and Gerhard Klebe. Multiple graph alignment for the structural analysis of protein active sites. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 4(2):310–320, 2007.

[1725] J. Weston, F. Pérez-Cruz, O. Bousquet, O. Chapelle, A. Elisseeff, and B. Schölkopf. Feature selection and transduction for prediction of molecular bioactivity for drug design. *Bioinformatics*, 19(6):764–771, 2003.

[1726] D. L. Wheeler, T Barrett, D.A Benson, and S.H. Bryant. Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, 31:28–33, 2006.

[1727] W. J. Wilbur. Boosting na ve Bayesian learning on a large subset of MEDLINE. *Proc AMIA Symp*, pages 918–22, 2000.

[1728] M. R. Wilkins, C. Pasquali, R. D. Appel, K. Ou, O. Golaz, J. C. Sanchez, J. X. Yan, A. A. Gooley, G. Hughes, I. Humphery-Smith, K. L. Williams, and D. F. Hochstrasser. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)*, 14(1):61–65, Jan 1996.

[1729] F. M. J. Willems. Universal data compression and repetition times. *IEEE Trans. Inform. Theory*, 35(1):54–58, Jan 1989.

[1730] F. M. J. Willems. Coding for a Binary Independent Piecewise-Identically Distributed Source. *IEEE Trans. Inform. Theory*, 42:2210–2217, nov 1996.

[1731] F. M. J. Willems, Y. M. Shtarkov, and T. J. Tjalkens. The Context Tree Weighting Method: Basic Properties. *IEEE Trans. Inform. Theory*, 41(3):653–664, May 1995.

[1732] F. M. J. Willems, Y. M. Shtarkov, and T. J. Tjalkens. Context Weighting for General Finite Context Sources. *IEEE Trans. Inform. Theory*, 42(5):1514–1520, 1996.

[1733] P. Willett. Chemical Similarity Searching. *J Chem Inf Comput Sci*, 38:983–996, 1998.

[1734] Peter Willett, Vivienne Winterman, and David Bawden. Implementation of nearest-neighbor searching in an online chemical structure search system. *Journal of Chemical Information and Computer Sciences*, 26(1):36–41, 1986.

[1735] C.K.I. Williams. Prediction with Gaussian Processes: From Linear Regression to Linear Prediction and Beyond. In M.I. Jordan, editor, *Learning and Inference in Graphical Models.* Kluwer Academic Press, 1998.

[1736] R.D. Williams, S.N. Hing, B.T. Greer, C.C. Whiteford, J.S. Wei, R. Natrajan, A. Kelsey, S. Rogers, C. Campbell, K. Pritchard-Jones, and J. Khan. Prognostic classification of relapsing favorable histology Wilms tumor using cDNA microarray expression profiling and support vector machines. *Genes Chromosomes Cancer*, 41(1):65–79, Sep 2004.

[1737] R.C. Williamson, A.J. Smola, and B. Schoelkopf. Entropy Numbers of Linear Function Classes. In *Proc. 13th Annu. Conference on Comput. Learning Theory*, pages 309–319. Morgan Kaufmann, San Francisco, 2000.

[1738] D. Wilton, P. Willett, K. Lawson, and G. Mullier. Comparison of ranking methods for virtual screening in lead-discovery programs. *J Chem Inf Comput Sci*, 43(2):469–74, 2003.

[1739] S. Winters-Hilt, W. Vercoutere, V.S. DeGuzman, D. Deamer, M. Akeson, and D. Haussler. Highly accurate classification of Watson-Crick basepairs on termini of single DNA molecules. *Biophys. J.*, 84(2):967–976, 2003.

[1740] Stephen Winters-Hilt and Mark Akeson. Nanopore cheminformatics. *DNA Cell Biol*, 23(10):675–83, Oct 2004.

[1741] E. A Winzeler. Applied systems biology and malaria. *Nat Rev Microbiol*, 4(2):145–151, Feb 2006.

[1742] D. M. Witten and R. Tibshirani. Covariance-regularized regression and classification for high dimensional problems. *J. R. Stat. Soc. Ser. B*, 71(3), 2009.

[1743] L. Wolf and A. Shashua. Learning over Sets using Kernel Principal Angles. *J. Mach. Learn. Res.*, 4:913–931, 2003.

[1744] B. Wu, T. Abbott, D. Fishman, W. McMurray, G. Mor, K. Stone, D. Ward, K. Williams, and H. Zhao. Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. *Bioinformatics*, 19(13):1636–1643, 2003.

[1745] Jiann-Ming Wu. Natural discriminant analysis using interactive Potts models. *Neural Comput*, 14(3):689–713, Mar 2002.

[1746] Z Wu, R A Irizarry, R Gentleman, F M Murillo, and F Spencer. A model based background adjustment for oligonucleotide expression arrays. Technical report, John Hopkins University, Department of Biostatistics Working Papers, Baltimore, MD, 2003.

[1747] A.D. Wyner and J. Ziv. Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. *IEEE Trans. Inform. Theory*, 35(6):1250–1258, Nov 1989.

[1748] H. Xia, Q. Mao, S. L. Eliason, S. Q. Harper, I. H. Martins, H. T. Orr, H. L. Paulson, L. Yang, R. M. Kotin, and B. L. Davidson. RNAi suppresses polyglutamine-induced neurodegeneration in a model of spinocerebellar ataxia. *Nat. Med.*, 10(8):816–820, Aug 2004.

[1749] T. Xia, J. SantaLucia, M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox, and D. H. Turner. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, 37(42):14719–35, Oct 1998.

[1750] Youshen Xia and Jun Wang. A one-layer recurrent neural network for support vector machine learning. *IEEE Trans Syst Man Cybern B Cybern*, 34(2):1261–9, Apr 2004.

[1751] Chao Xie and Martti T Tammi. Cnv-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics*, 10:80, 2009.

[1752] Dan Xie, Ao Li, Minghui Wang, Zhewen Fan, and Huanqing Feng. LOCSVMPSI: a web server for subcellular localization of eukaryotic proteins using SVM and profile of PSI-BLAST. *Nucleic Acids Res.*, 33(Web Server issue):W105–10, Jul 2005.

[1753] Lei Xie, Li Xie, and Philip E Bourne. A unified statistical model to support local sequence order independent similarity searching for ligand-binding sites and its application to genome-based drug discovery. *Bioinformatics*, 25(12):i305–i312, Jun 2009.

[1754] Q. Xie and A.R. Barron. Minimax redundancy for the class of memoryless sources. *IEEE Trans. Inform. Theory*, 43(2):646–657, Mar 1997.

[1755] Q. Xie and A.R. Barron. Asymptotic minimax regret for data compression, gambling, and prediction. *IEEE Trans. Inform. Theory*, 46(2):431 – 445, Mar 2000.

[1756] E. Xing and R. Karp. Motifprototyper: A bayesian profile model for motif families. *PNAS*, 101(29):10523–10528, 2004.

[1757] E. P. Xing, W. Wu, M. I. Jordan, and R. M. Karp. LOGOS: A modular Bayesian model for de novo motif detection. *J. Bioinform. Comput. Biol.*, 2:127–154, 2004.

[1758] E.P. Xing, A.Y. Ng, M.I. Jordan, and S. Russell. Distance metric learning with application to clustering with side-information. In S. Thrun S. Becker and K. Obermayer, editors, *Adv. Neural. Inform. Process Syst.*, volume 15, pages 505–512, Cambridge, MA, 2003. MIT Press.

[1759] M. Xiong, X. Fang, and J. Zhao. Biomarker Identification by Feature Wrappers. *Genome Res.*, 11(11):1878–1887, 2001.

[1760] Jian xiong Dong, Adam Krzyzak, and Ching Y Suen. Fast SVM training algorithm with decomposition on very large data sets. *IEEE Trans Pattern Anal Mach Intell*, 27(4):603–18, Apr 2005.

[1761] J. Xu and A. Hagler. Chemoinformatics and Drug Discovery. *Molecules*, 7:566–600, 2002.

[1762] Xiu-Qin Xu, Chon K Leow, Xin Lu, Xuegong Zhang, Jun S Liu, Wing-Hung Wong, Arndt Asperger, Sren Deininger, and Hon-Chiu Eastwood Leung. Molecular classification of liver cirrhosis in a rat model by proteomics and bioinformatics. *Proteomics*, 4(10):3235–45, Oct 2004.

[1763] C. X. Xue, R. S. Zhang, H. X. Liu, M. C. Liu, Z. D. Hu, and B. T. Fan. Support vector machines-based quantitative structure-property relationship for the prediction of heat capacity. *J Chem Inf Comput Sci*, 44(4):1267–74, 2004.

[1764] C. X. Xue, R. S. Zhang, H. X. Liu, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan. An accurate QSPR study of O-H bond dissociation energy in substituted phenols based on support vector machines. *J Chem Inf Comput Sci*, 44(2):669–77, 2004.

[1765] C. X. Xue, R. S. Zhang, H. X. Liu, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan. QSAR models for the prediction of binding affinities to human serum albumin using the heuristic method and a support vector machine. *J Chem Inf Comput Sci*, 44(5):1693–700, 2004.

[1766] C. X. Xue, R. S. Zhang, M. C. Liu, Z. D. Hu, and B. T. Fan. Study of the quantitative structure-mobility relationship of carboxylic acids in capillary electrophoresis based on support vector machines. *J Chem Inf Comput Sci*, 44(3):950–7, 2004.

[1767] L. Xue and J. Bajorath. Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. *Comb. Chem. High. Throughput Screen.*, 3(5):363–372, Oct 2000.

[1768] L. Xue, F. L. Stahura, J. W. Godden, and J. Bajorath. Fingerprint scaling increases the probability of identifying molecules with similar activity in virtual screening calculations. *J Chem Inf Comput Sci*, 41(3):746–753, 2001.

[1769] L. Xue, F. L. Stahura, J. W. Godden, and J. Bajorath. Mini-fingerprints detect similar activity of receptor ligands previously recognized only by three-dimensional pharmacophore-based methods. *J Chem Inf Comput Sci*, 41(2):394–401, 2001.

[1770] Y. Xue, Z. R. Li, C. W. Yap, L. Z. Sun, X. Chen, and Y. Z. Chen. Effect of molecular descriptor feature selection in support vector machine classification of pharmacokinetic and toxicological properties of chemical agents. *J Chem Inf Comput Sci*, 44(5):1630–8, 2004.

[1771] Y. Xue, C. W. Yap, L. Z. Sun, Z. W. Cao, J. F. Wang, and Y. Z. Chen. Prediction of P-glycoprotein substrates by a support vector machine approach. *J Chem Inf Comput Sci*, 44(4):1497–505, 2004.

[1772] Ya Xue, David Dunson, and Lawrence Carin. The matrix stick-breaking process for flexible multi-task learning. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pages 1063–1070, New York, NY, USA, June 2007. ACM.

[1773] Ya Xue, Xuejun Liao, Lawrence Carin, and Balaji Krishnapuram. Multi-task learning for classification with dirichlet process priors. *Journal of Machine Learning Research*, 8:2007, January 2007.

[1774] Y. Yabuki, T. Muramatsu, T. Hirokawa, H. Mukai, and M. Suwa. GRIFFIN: a system for predicting GPCR-G-protein coupling selectivity using a support vector machine and a hidden Markov model. *Nucleic Acids Res.*, 33(Web Server issue):W148–53, Jul 2005.

[1775] T. Yamada and S. Morishita. Accelerated off-target search algorithm for siRNA. *Bioinformatics*, 21(8):1316–24, Apr 2005.

[1776] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, and M. Kanehisa. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, 24(13):i232–i240, Jul 2008.

[1777] Y. Yamanishi, F. Bach, and J.-P. Vert. Glycan classification with tree kernels. *Bioinformatics*, 23(10):1211–1216, May 2007.

[1778] Y. Yamanishi, J.-P. Vert, and M. Kanehisa. Heterogeneous data comparison and gene selection with kernel canonical correlation analysis. In B. Schlkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 209–230. MIT Press, 2004.

[1779] Y. Yamanishi, J.-P. Vert, and M. Kanehisa. Protein network inference from multiple genomic data: a supervised approach. *Bioinformatics*, 20:i363–i370, 2004.

[1780] Y. Yamanishi, J.-P. Vert, and M. Kanehisa. Supervised enzyme network inference from the integration of genomic data and chemical information. *Bioinformatics*, 21:i468–i477, 2005.

[1781] Y. Yamanishi, J.-P. Vert, A. Nakaya, and M. Kanehisa. Extraction of correlated gene clusters from multiple genomic data by generalized kernel canonical correlation analysis. *Bioinformatics*, 19(Suppl. 1):i323–i330, 2003.

[1782] C. Yan, D. Dobbs, and V. Honavar. A two-stage classifier for identification of protein-protein interface residues. *Bioinformatics*, 20(Suppl. 1):i371–i378, 2004.

[1783] C. Yan, V. Honavar, and D. Dobbs. Identification of interface residues in protease-inhibitor and antigen-antibody complexes: a support vector machine. *Neural Comput. & Applic.*, 13:123–129, 2004.

[1784] Y. Yang. Minimax nonparametric classification — Part I: rates of convergence. *IEEE Trans. Inform. Theory*, 45(7):2271–2284, 1999.

[1785] Y. H. Yang, S. Dudoit, P. Luu, D. M. Lin, V. Peng, J. Ngai, and T. P. Speed. Normalization for cdna microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res*, 30(4), February 2002.

[1786] Z. R. Yang and K.-C. Chou. Bio-support vector machines for computational proteomics. *Bioinformatics*, 20(5):735–741, 2004.

[1787] Zheng Rong Yang. Biological applications of support vector machines. *Brief Bioinform*, 5(4):328–38, Dec 2004.

[1788] Zheng Rong Yang. Prediction of caspase cleavage sites using Bayesian bio-basis function neural networks. *Bioinformatics*, 21(9):1831–7, May 2005.

[1789] Chen Yanover and Tomer Hertz. Predicting protein-peptide binding affinity by learning peptide-peptide distance functions. In *RECOMB*, pages 456–471, 2005.

[1790] X. Yao, C. Parnot, X. Deupi, V. R. P. Ratnala, G. Swaminath, D. Farrens, and B. Kobilka. Coupling ligand structure to specific conformational switches in the beta2-adrenoceptor. *Nat. Chem. Biol.*, 2(8):417–422, Aug 2006.

[1791] X. J. Yao, A. Panaye, J. P. Doucet, R. S. Zhang, H. F. Chen, M. C. Liu, Z. D. Hu, and B. T. Fan. Comparative study of QSAR/QSPR correlations using support vector machines, radial basis function neural networks, and multiple linear regression. *J Chem Inf Comput Sci*, 44(4):1257–66, 2004.

[1792] C. W. Yap, C. Z. Cai, Y. Xue, and Y. Z. Chen. Prediction of torsade-causing potential of drugs by support vector machine approach. *Toxicol Sci*, 79(1):170–7, May 2004.

[1793] C. W. Yap and Y. Z. Chen. Prediction of Cytochrome P450 3A4, 2D6, and 2C9 Inhibitors and Substrates by Using Support Vector Machines. *J Chem Inf Model*, 45(4):982–92, 2005.

[1794] C.H. Yeang, S. Ramaswamy, P. Tamayo, S. Mukherjee, R.M. Rifkin, M. Angelo, M. Reich, E. Lander, J. Mesirov, and T. Golub. Molecular classification of multiple tumor types. *Bioinformatics*, 17(Suppl. 1):S316–S322, 2001.

[1795] Wen-Chun Yeh, Yung-Ming Jeng, Cheng-Han Li, Po-Huang Lee, and Pai-Chi Li. Liver steatosis classification using high-frequency ultrasound. *Ultrasound Med Biol*, 31(5):599–605, May 2005.

[1796] J. W. Yewdell and J. R. Bennink. Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annu. Rev. Immunol.*, 17:51–88, 1999.

[1797] Y. Yi, C. Li, C. Miller, and A. L. George. Strategy for encoding and comparison of gene expression signatures. *Genome Biol.*, 8(7):R133, 2007.

[1798] S. M. Yiu, Prudence W. H. Wong, T.W. Lam, Y.C. Mui, H. F. Kung, Marie Lin, and Y. T. Cheung. Filtering of Ineffective siRNAs and Improved siRNA Design Tool. *Bioinformatics*, 21(2):144–151, Jan 2005. To appear.

[1799] Bauke Ylstra, Paul Van den Ijssel, Beatriz Carvalho, Ruud H Brakenhoff, and Gerrit A Meijer. BAC to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array CGH). *Nucleic Acids Research*, 34:445–450, 2006.

[1800] S. H. Yook, H. Jeong, Y. Tu, and A.-L. Barabási. Weighted evolution networks. *Phys. Rev. Lett.*, 86(25):5835–5838, 2001.

[1801] Seungtai Yoon, Zhenyu Xuan, Vladimir Makarov, Kenny Ye, and Jonathan Sebat. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res*, 19(9):1586–1592, Sep 2009.

[1802] Y. Yoon, J. Song, S.H. Hong, and J.Q. Kim. Analysis of multiple single nucleotide polymorphisms of candidate genes related to coronary heart disease susceptibility by using support vector machines. *Clin. Chem. Lab. Med.*, 41(4):529–534, 2003.

[1803] N. Yosef, R. Sharan, and W.S. Noble. Improved network-based identification of protein orthologs. *Bioinformatics*, 24(16):i200–i206, Aug 2008.

[1804] J. A. Young and E. A. Winzeler. Using expression information to discover new drug and vaccine targets in the malaria parasite Plasmodium falciparum. *Pharmacogenomics*, 6(1):17–26, Jan 2005.

[1805] Jason A Young, Quinton L Fivelman, Peter L Blair, Patricia de la Vega, Karine G Le Roch, Yingyao Zhou, Daniel J Carucci, David A Baker, and Elizabeth A Winzeler. The Plasmodium falciparum sexual development transcriptome: a microarray analysis using ontology-based pattern identification. *Mol Biochem Parasitol*, 143(1):67–79, Sep 2005.

[1806] B. Yu. Lower bounds on expected redundancy for nonparametric classes. *IEEE Trans. Inform. Theory*, 42(1):272–275, Jan 1996.

[1807] C. Yu, N. Zavaljevski, F. J. Stevens, K. Yackovich, and J. Reifman. Classifying noisy protein sequence data: a case study of immunoglobulin light chains. *Bioinformatics*, 21(Supp 1):i495–i501, Jun 2005.

[1808] C.-S. Yu, C.-J. Lin, and J.-K. Hwang. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.*, 13(5):1402–1406, 2004.

[1809] C.S. Yu, J.Y. Wang, J.M. Yang, P.C. Lyu, C.J. Lin, and J.K. Hwang. Fine-grained protein fold assignment by support vector machines using generalized npeptide coding schemes and jury voting from multiple-parameter sets. *Proteins*, 50(4):531, 6 2003.

[1810] H. Yu, P. Braun, M. A. Yildirim, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J.-F. Rual, A. Dricot, A. Vazquez, R. R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrzikapa, C. Fan, A.-S. de Smet, A. Motyl, M. E. Hudson, J. Park, X. Xin, M. E. Cusick, T. Moore, C. Boone, M. Snyder, F. R. Roth, A.-L. Barabási, J. Tavernier, D. E. Hill, and M. Vidal. High-quality binary protein interaction map of the yeast interactome network. *Science*, 322(5898):104–110, Oct 2008.

[1811] J.-k. Yu, S. Zheng, Y. Tang, and L. Li. An integrated approach utilizing proteomics and bioinformatics to detect ovarian cancer. *J Zhejiang Univ Sci B*, 6(4):227–31, Apr 2005.

[1812] J. S. Yu, S. Ongarello, R. Fiedler, X. W. Chen, G. Toffolo, C. Cobelli, and Z. Trajanoski. Ovarian cancer identification based on dimensionality reduction for high-throughput mass spectrometry data. *Bioinformatics*, 21(10):2200–9, May 2005.

[1813] J.K. Yu, Y.D. Chen, and S. Zheng. An integrated approach to the detection of colorectal cancer utilizing proteomics and bioinformatics. *World J. Gastroenterol.*, 10(21):3127–3131, 2004.

[1814] Kai Yu, Volker Tresp, and Anton Schwaighofer. Learning gaussian processes from multiple tasks. In *ICML '05: Proceedings of the 22nd international conference on Machine learning*, pages 1012–1019, New York, NY, USA, 2005. ACM.

[1815] Kun Yu, Nikolai Petrovsky, Christian Schnbach, Judice Y L Koh, and Vladimir Brusic. Methods for prediction of peptide binding to MHC molecules: a comparative study. *Mol Med*, 8(3):137–148, Mar 2002.

[1816] Shipeng Yu, Volker Tresp, and Kai Yu. Robust multi-task learning with t-processes. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pages 1103–1110, New York, NY, USA, 2007. ACM.

[1817] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B*, 68(1):49–67, 2006.

[1818] Ming Yuan and Yi Lin. Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.

[1819] Ming Yuan and Yi Lin. On the non-negative garrotte estimator. *Journal Of The Royal Statistical Society Series B*, 69(2):143–161, 2007.

[1820] Y. Yuan, L. Guo, L. Shen, and J. S. Liu. Predicting gene expression from sequence: A reexamination. *PLoS Comput. Biol.*, 3(11):e243, 2007.

[1821] Z. Yuan, K. Burrage, and J.S. Mattick. Prediction of protein solvent accessibility using support vector machines. *Proteins*, 48(3):566–570, 2002.

[1822] Z. Yuan, J.S. Mattick, and R.D. Teasdale. SVMtm: support vector machines to predict transmembrane segments. *J. Comput. Chem.*, 25(5):632, 6 2004.

[1823] N. M. Zaki, S. Deris, and R. Illias. Application of string kernels in protein sequence classification. *Appl. Bioinformatics*, 4(1):45–52, 2005.

[1824] P.D. Zamore, T. Tuschl, P.A. Sharp, and D.P. Bartel. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, 101(1):25–33, Mar 2000.

[1825] Linda M Zangwill, Kwokleung Chan, Christopher Bowd, Jicuang Hao, Te-Won Lee, Robert N Weinreb, Terrence J Sejnowski, and Michael H Goldbaum. Heidelberg retina tomograph measurements of the optic disc and parapapillary retina for detecting glaucoma analyzed by machine learning classifiers. *Invest Ophthalmol Vis Sci*, 45(9):3144–51, Sep 2004.

[1826] M. Zaslavskiy, F. Bach, and J. P. Vert. GRAPHM: Graph matching package, 2008. Available at `http://cbio.ensmp.fr/graphm`.

[1827] M. Zaslavskiy, F. Bach, and J. P. Vert. A path following algorithm for graph matching. In A. Elmoataz, O. Lezoray, F. Nouboud, and D. Mammass, editors, *Image and Signal Processing, Proceedings of the 3rd International Conference, ICISP 2008*, volume 5099 of *LNCS*, pages 329–337. Springer Berlin / Heidelberg, 2008.

[1828] M. Zaslavskiy, F. Bach, and J.-P. Vert. A path following algorithm for the graph matching problem. Technical Report 00232851, HAL, 2008. To appear in IEEE Trans. Pattern Anal. Mach. Intell.

[1829] N. Zavaljevski, F.J. Stevens, and J. Reifman. Support vector machines with selective kernel scaling for protein classification and identification of key amino acid positions. *Bioinformatics*, 18(5):689–696, 2002.

[1830] Arnold Zellner. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57(298):348–368, 1962.

[1831] V. V. Zernov, K. V. Balakin, A. A. Ivaschenko, N. P. Savchuk, and I. V. Pletnev. Drug discovery using support vector machines. The case studies of drug-likeness, agrochemical-likeness, and enzyme inhibition predictions. *J Chem Inf Comput Sci*, 43(6):2048–56, 2003.

[1832] G. L. Zhang, A. M. Khan, K. N. Srinivasan, J. T. August, and V. Brusic. MULTIPRED: a computational system for prediction of promiscuous HLA binding peptides. *Nucleic Acids Res/*, 33(Web Server issue):W172–W179, Jul 2005.

[1833] H. H. Zhang, Y. Liu, Y. Wu, and J. Zhu. Variable selection for multicategory SVM via adaptive sup-norm regularization. *Electronic Journal of Statistics*, 2:149–167, 2008.

[1834] Li Zhang, Weida Zhou, and Licheng Jiao. Hidden space support vector machines. *IEEE Trans Neural Netw*, 15(6):1424–34, Nov 2004.

[1835] Li Zhang, Weida Zhou, and Licheng Jiao. Wavelet support vector machine. *IEEE Trans Syst Man Cybern B Cybern*, 34(1):34–9, Feb 2004.

[1836] Lu-Da Zhang, Shi-Guang Su, Lai-Sheng Wang, Jun-Hui Li, and Li-Ming Yang. Study on application of Fourier transformation near-infrared spectroscopy analysis with support vector machine (SVM). *Guang Pu Xue Yu Guang Pu Fen Xi*, 25(1):33–5, Jan 2005.

[1837] Qidong Zhang, Sukjoon Yoon, and William J Welsh. Improved method for predicting beta-turn using support vector machine. *Bioinformatics*, 21(10):2370–4, May 2005.

[1838] S.-W. Zhang, Q. Pan, H.-C. Zhang, Y-L. Zhang, and H.-Y. Wang. Classification of protein quaternary structure with support vector machine. *Bioinformatics*, 19(18):2390–2396, 2003.

[1839] T. Zhang. Statistical behavior and consistency of classification methods based on convex risk minimization. *Ann. Stat.*, 32:56–134, 2004.

[1840] X. H-F. Zhang, K. A. Heller, I. Hefter, C. S. Leslie, and L. A. Chasin. Sequence Information for the Splicing of Human Pre-mRNA Identified by Support Vector Machine Classification. *Genome Res.*, 13(12):2637–2650, 2003.

[1841] Yang Zhang. Progress and challenges in protein structure prediction. *Curr. Opin. Struct. Biol.*, 18(3):342–348, June 2008.

[1842] Z. Zhang and M.G. Grigorov. Similarity networks of protein binding sites. *Proteins*, 62(2):470–478, Feb 2006.

[1843] Z. Zhang, S. Kochhar, and M. G. Grigorov. Descriptor-based protein remote homology identification. *Protein Sci.*, 42(2):431–444, 2005.

[1844] Zhengyou ZHANG. Iterative point matching for registration of free-form curves. Technical report, Institut National de Recherche en Informatique et en Automatique (INRIA), 1992.

[1845] C. Y. Zhao, R. S. Zhang, H. X. Liu, C. X. Xue, S. G. Zhao, X. F. Zhou, M. C. Liu, and B. T. Fan. Diagnosing anorexia based on partial least squares, back propagation neural network, and support vector machines. *J Chem Inf Comput Sci*, 44(6):2040–6, 2004.

[1846] P. Zhao, G. Rocha, and B. Yu. Grouped and hierarchical model selection through composite absolute penalties. *Ann. Stat.*, To appear.(To appear.), 2009. To appear.

[1847] P. Zhao and B. Yu. On model selection consistency of lasso. *J. Mach. Learn. Res.*, 7:2541, 2006.

[1848] Y. Zhao, C. Pinilla, D. Valmori, R. Martin, and R. Simon. Application of support vector machines for T-cell epitopes prediction. *Bioinformatics*, 19(15):1978–1984, 2003.

[1849] Y. Zhao, C. Pinilla, D. Valmori, R. Martin, and R. Simon. Application of support vector machines for T-cell epitopes prediction. *Bioinformatics*, 19(15):1978–1984, Oct 2003.

[1850] Wenming Zheng, Li Zhao, and Cairong Zou. Foley-Sammon optimal discriminant vectors using kernel approach. *IEEE Trans Neural Netw*, 16(1):1–9, Jan 2005.

[1851] D. Zhou. The covering number in learning theory. *J. Complexity*, 18:739–767, 2002.

[1852] GuoDong Zhou, Dan Shen, Jie Zhang, Jian Su, and SoonHeng Tan. Recognition of protein/gene names from text using an ensemble of classifiers. *BMC Bioinformatics*, 6 Suppl 1:S7, 2005.

[1853] GuoDong Zhou, Jie Zhang, Jian Su, Dan Shen, and ChewLim Tan. Recognizing names in biomedical texts: a machine learning approach. *Bioinformatics*, 20(7):1178–90, May 2004.

[1854] Xin Zhou and K. Z. Mao. LS Bound based gene selection for DNA microarray data. *Bioinformatics*, 21(8):1559–64, Apr 2005.

[1855] G. Zhu, P. T. Spellman, T. Volpe, P. O. Brown, D. Botstein, T. N. Davis, and B. Futcher. Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. *Nature*, 406:90–94, 2000.

[1856] H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein, and M. Snyder. Global analysis of protein activities using proteome chips. *Science*, 293(5537):2101–5, Sep 2001.

[1857] J. Zhu and T. Hastie. Classification of gene microarrays by penalized logistic regression. *Biostatistics*, 5(3):427–43, Jul 2004.

[1858] J. Zhu and T. Hastie. Kernel Logistic Regression and the Import Vector Machine. *Journal of Computational & Graphical Statistics*, 14(1):185–205, Mar 2005.

[1859] J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani. 1-norm support vector machines. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Adv. Neural. Inform. Process Syst.*, volume 16, Cambridge, MA, 2004. MIT Press.

[1860] Lingyun Zhu, Baoming Wu, and Changxiu Cao. Introduction to medical data mining. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*, 20(3):559–62, Sep 2003.

[1861] S. C. Zhu, Y. Wu, and D. Mumford. FRAME: Filters, Random field And Maximum Entropy: — Towards a Unified Theory for Texture Modeling. *Int'l Journal of Computer Vision*, 27(2):1–20, 1998.

[1862] S. C. Zhu, Z. N. Wu, and D. Mumford. Minimax Entropy Principle and Its Application to Texture Modeling. *Neural Comput.*, 9(8):1627–1660, 1997.

[1863] Shanfeng Zhu, Keiko Udaka, John Sidney, Alessandro Sette, Kiyoko F. Aoki-Kinoshita, and Hiroshi Mamitsuka. Improving MHC binding peptide prediction by incorporating binding data of auxiliary MHC molecules. *Bioinformatics*, 22(13):1648–1655, 2006.

[1864] A. Zien and C. Ong. Multiclass multiple kernel learning. In Zoubin Ghahramani, editor, *Proceedings of the 24th Annual International Conference on Machine Learning (ICML 2007)*, pages 1191–1198. Omnipress, 2007.

[1865] A. Zien, G. Rätsch, S. Mika, B. Schölkopf, T. Lengauer, and K.-R. Müller. Engineering support vector machine kernels that recognize translation initiation sites. *Bioinformatics*, 16(9):799–807, 2000.

[1866] J. Ziv and A. Lempel. Compression of individual sequences via variable-rate coding. *IEEE Trans. Inform. Theory*, 24(5):530–536, Sep 1978.

[1867] Simeone Zomer, Christelle Guillo, Richard G Brereton, and Melissa Hanna-Brown. Toxicological classification of urine samples using pattern recognition techniques and capillary electrophoresis. *Anal Bioanal Chem*, 378(8):2008–20, Apr 2004.

[1868] Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101:1418–1429, December 2006.

[1869] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society B*, 67:301–320, 2005.

[1870] M. Zuker. On finding all suboptimal foldings of an RNA molecule. *Science*, 244(4900):48–52, Apr 1989.

[1871] J. Zupan and J. Gasteiger. *Neural Networks in Chemistry and Drug Design.* Wiley-VCH, 1999.