A weighted Kendall kernel

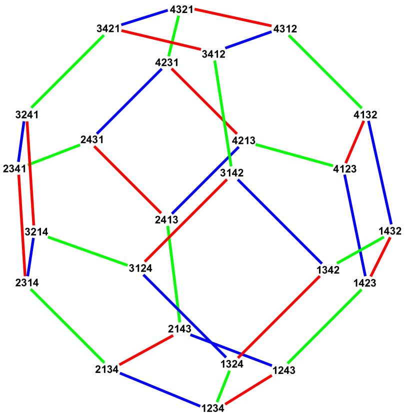
Yunlong Jiao & Jean-Philippe Vert

Learning on Distributions, Functions, Graphs and Groups
@ NIPS-2017, Dec. 8, 2017

Joint work with

Yunlong Jiao

Permutations



- Permutation:

$$\sigma : [1, n] \rightarrow [1, n]$$

- $\sigma(i)$ = rank of item i
- Composition

$$(\sigma_1 \sigma_2)(i) = \sigma_1(\sigma_2(i))$$

- $\mathbb{S}_n$ the symmetric group
- $|\mathbb{S}_n| = n!$

Learning over the symmetric group

- Assume your data are permutations and you want to learn

$$f : \mathbb{S}_n \to \mathbb{R}$$

- A solutions: embed $\mathbb{S}_n$ to a Euclidean or Hilbert space

$$\Phi : \mathbb{S}_n \to \mathcal{H}$$

  and learn a linear function

$$f(\sigma) = \beta^\top \Phi(\sigma)$$

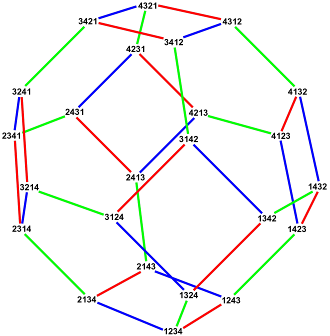
- The corresponding kernel is

$$K(\sigma_1, \sigma_2) = \Phi(\sigma_1)^\top \Phi(\sigma_2)$$

- A right-invariant kernel is invariant by renaming the items:

$$\forall \sigma_1, \sigma_2, \pi \in \mathbb{S}_n , \quad K(\sigma_1 \pi, \sigma_2 \pi) = K(\sigma_1, \sigma_2)$$

Related work



- Represent a permutation $x \in S_n$ by the vector of rank $\Phi(x) \in \mathbb{R}^n$
  - does not capture higher-order informations
- Diffusion kernel over the Cayley's graph (Kondor and Barbosa, 2010)
  - but complexity $O(n^{2n})$

Outline

Kendall and Mallows kernels

- Let $n_c(\sigma, \sigma')$ (resp. $n_d(\sigma, \sigma')$) the number of concordant (resp. discordant) pairs.
- The (rescaled) Kendall kernel (a.k.a. Kendall's τ correlation) is

$$K_\tau(\sigma, \sigma') = n_c(\sigma, \sigma')$$

- The Mallows kernel is

$$\forall \lambda \geq 0 \quad K_M^\lambda(\sigma, \sigma') = e^{-\lambda n_d(\sigma, \sigma')}$$

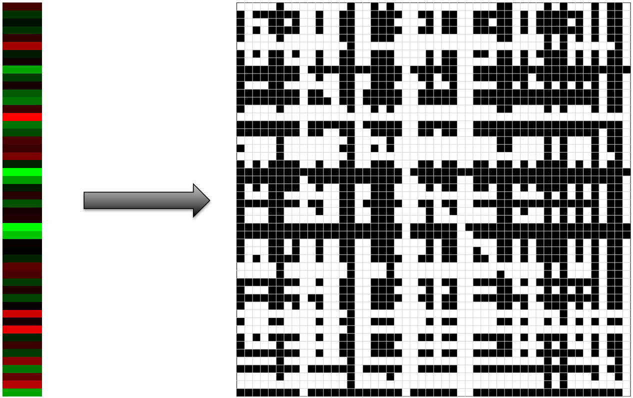
Theorem (Jiao and Vert, 2015, 2017)

The Kendall and Mallows kernels are positive definite.

Theorem (Knight, 1966)

These two kernels for permutations can be evaluated in $O(n \log n)$ time.

Proof



Take $\Phi_\tau(\sigma) = \left(\mathbb{1}_{\sigma(i)<\sigma(j)}\right)_{1\leq i\neq j\leq n} \in \mathbb{R}^{n(n-1)}$ and

$$K_\tau(\sigma,\sigma') = \Phi_\tau(\sigma)^\top \Phi_\tau(\sigma') = \sum_{1\leq i\neq j\leq n} \mathbb{1}_{\sigma(i)<\sigma(j)} \mathbb{1}_{\sigma'(i)<\sigma'(j)} .$$

Weighted Kendall's τ correlation

- How to weight differently pairs based on their ranks?
- Given a weight function $w : [1, n]^2 \to \mathbb{R}$, weighted versions of the Kendall's τ have been proposed:

$$\sum_{1 \leq i \neq j \leq n} w(\sigma(i), \sigma(j)) \mathbb{1}_{\sigma(i) < \sigma(j)} \mathbb{1}_{\sigma'(i) < \sigma'(j)}$$

Shieh (1998)

$$\sum_{1 \leq i \neq j \leq n} w(\sigma(i), \sigma(j)) \frac{p_{\sigma(i)} - p_{\sigma'(i)}}{\sigma(i) - \sigma'(i)} \frac{p_{\sigma(j)} - p_{\sigma'(j)}}{\sigma(j) - \sigma'(j)} \mathbb{1}_{\sigma(i) < \sigma(j)} \mathbb{1}_{\sigma'(i) < \sigma'(j)}$$

Kumar and Vassilvitskii (2010)

$$\sum_{1 \leq i \neq j \leq n} w(i, j) \mathbb{1}_{\sigma(i) < \sigma(j)} \mathbb{1}_{\sigma'(i) < \sigma'(j)}$$

Vigna (2015)

- However, they are either not symmetric (1st and 2nd), or right-invariant (3rd)
- How to make a right-invariant, p.d. weighted Kendall correlation?

A right-invariant weighted Kendall kernel

Theorem

Let $W : \mathbb{N}^2 \times \mathbb{N}^2 \to \mathbb{R}$ be a p.d. kernel on $\mathbb{N}^2$, then the function $K_W : \mathbb{S}_n \times \mathbb{S}_n \to \mathbb{R}$ defined by

$$K_W(\sigma, \sigma') = \sum_{1 \leq i \neq j \leq n} W\left((\sigma(i), \sigma(j)), (\sigma'(i), \sigma'(j))\right) \mathbb{1}_{\sigma(i)<\sigma(j)} \mathbb{1}_{\sigma'(i)<\sigma'(j)}$$

is a right-invariant p.d. kernel on $\mathbb{S}_n$.

Corollary

If weights take the form $W((a,b),(c,d)) = U_{a,b}U_{c,d}$ for some matrix $U \in \mathbb{R}^{n \times n}$, then the function

$$K_U(\sigma, \sigma') = \sum_{1 \leq i \neq j \leq n} U_{\sigma(i),\sigma(j)} U_{\sigma'(i),\sigma'(j)} \mathbb{1}_{\sigma(i)<\sigma(j)} \mathbb{1}_{\sigma'(i)<\sigma'(j)} ,$$

is a right-invariant p.d. kernel on $\mathbb{S}_n$.

A right-invariant weighted Kendall kernel

Theorem

Let $W : \mathbb{N}^2 \times \mathbb{N}^2 \to \mathbb{R}$ *be a p.d. kernel on* $\mathbb{N}^2$*, then the function* $K_W : \mathbb{S}_n \times \mathbb{S}_n \to \mathbb{R}$ *defined by*

$$K_W(\sigma, \sigma') = \sum_{1 \leq i \neq j \leq n} W\left((\sigma(i), \sigma(j)), (\sigma'(i), \sigma'(j))\right) \mathbb{1}_{\sigma(i) < \sigma(j)} \mathbb{1}_{\sigma'(i) < \sigma'(j)}$$

is a right-invariant p.d. kernel on $\mathbb{S}_n$.

Corollary

If weights take the form $W((a, b), (c, d)) = U_{a,b} U_{c,d}$ *for some matrix* $U \in \mathbb{R}^{n \times n}$*, then the function*

$$K_U(\sigma, \sigma') = \sum_{1 \leq i \neq j \leq n} U_{\sigma(i),\sigma(j)} U_{\sigma'(i),\sigma'(j)} \mathbb{1}_{\sigma(i) < \sigma(j)} \mathbb{1}_{\sigma'(i) < \sigma'(j)} ,$$

is a right-invariant p.d. kernel on $\mathbb{S}_n$.

Examples

$U_{a,b}$ corresponds to the weight of (items ranked at) positions a and b in a permutation. Interesting choices include:

- *Top-k.* For some $k \in [1, n]$,

$$U_{a,b} = \begin{cases} 1 & \text{if } a \leq k \text{ and } b \leq k\,, \\ 0 & \text{otherwise.} \end{cases}$$

- *Additive.* For some $u \in \mathbb{R}^n$, take

$$U_{ij} = u_i + u_j$$

- *Multiplicative.* For some $u \in \mathbb{R}^n$, take

$$U_{ij} = u_i u_j$$

Theorem (Kernel trick)

The weighted Kendall kernel can be computed in $O(n \ln(n))$ for the top-k, additive or multiplicative weights.

Learning the weights U?

- K_U can be written as

$$K_U(\sigma, \sigma') = \Phi_U(\sigma)^\top \Phi_U(\sigma')$$

  with

$$\Phi_U(\sigma) = \left(U_{\sigma(i),\sigma(j)} \mathbb{1}_{\sigma(i)<\sigma(j)} \right)_{1 \leq i \neq j \leq n} \in \mathbb{R}^{n(n-1)}$$

- The function to be learned is

$$f(\sigma) = \beta^\top \Phi_U(\sigma)$$

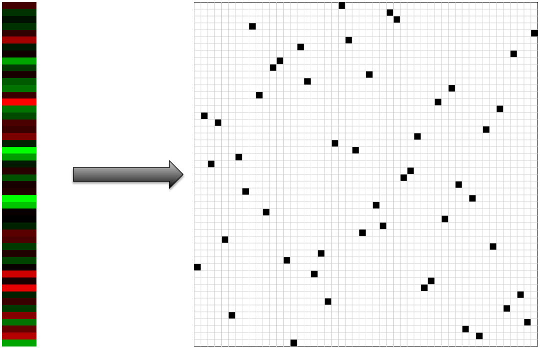
- We usually fit β by minimizing a (regularized) empirical risk

$$\min_{\beta} \sum_{i=1}^{n} \ell_i(\beta^\top \Phi_U(\sigma))$$

- Can we jointly optimize the weights:

$$\min_{\beta, U} \sum_{i=1}^{n} \ell_i(\beta^\top \Phi_U(\sigma))$$

Writing Φ_U as a function of U



Lemma

For any upper triangular matrix $U \in \mathbb{R}^{n \times n}$,

$$\Phi_U(\sigma) = \Pi_\sigma^\top U \Pi_\sigma \quad \text{with } (\Pi_\sigma)_{ij} = \mathbb{1}_{i=\sigma(j)}$$

Optimizing both β and U

- From this lemma we get that

$$
\begin{aligned}
f_{\beta,U}(\sigma) &= \left\langle \beta, \Phi^U(\sigma) \right\rangle_{\text{Frobenius}(n\times n)} \\
&= \left\langle \beta, \Pi_\sigma^\top U \Pi_\sigma \right\rangle_{\text{Frobenius}(n\times n)} \\
&= \left\langle \Pi_\sigma \otimes \Pi_\sigma, \text{vec}(U) \otimes (\text{vec}(\beta))^\top \right\rangle_{\text{Frobenius}(n^2\times n^2)}
\end{aligned}
$$

- This is symmetric in U and β
- Note that $\Pi_\sigma^\top = (\Pi_\sigma)^{-1} = \Phi_{\sigma^{-1}}$, hence

$$f_{\beta,U}(\sigma) = f_{U,\beta}(\sigma^{-1})$$

- We propose to alternate optimization in U and β
  - For U fixed, optimize β with $K_U(\sigma_1, \sigma_2)$
  - For β fixed, optimize U with $K_\beta(\sigma_1^{-1}, \sigma_2^{-1})$

The representation point of view

$$f_{\beta,U}(\sigma) = \left\langle \Pi_\sigma \otimes \Pi_\sigma, \mathrm{vec}(U) \otimes (\mathrm{vec}(\beta))^\top \right\rangle_{\text{Frobenius}(n^2 \times n^2)}$$

- A particular rank-1 linear model for the embedding

$$\Sigma_\sigma = \Pi_\sigma \otimes \Pi_\sigma \in (\{0,1\})^{n^2 \times n^2}$$

- Σ is the direct sum of the second-order and first-order permutation representations:

$$\Sigma \cong \tau_{(n-2,1,1)} \oplus \tau_{(n-1,1)}$$

- This generalizes SUQUAN (Le Morvan and Vert, 2017) which considers the first-order representation Π_σ only:

$$h_{\beta,w}(\sigma) = \left\langle \Pi_\sigma, w \otimes \beta^\top \right\rangle_{\text{Frobenius}(n \times n)}$$

Conclusion

- A right-invariant, positive definite weighted version of the Kendall kernel
- Weights can be learned
- This is equivalent to learning a rank-1 tensor on a second-order representation of $\mathbb{S}_n$
- An intriguing connection between quantile normalization and Kendall's τ
- Ongoing work:
  - Experimental validation
  - Computatioinally tractable generalizations to higher orders

THANKS

References

Y. Jiao and J.-P. Vert. The Kendall and Mallows kernels for permutations. In *Proceedings of The 32nd International Conference on Machine Learning*, volume 37 of *JMLR:W&CP*, pages 1935–1944, 2015. URL `http://jmlr.org/proceedings/papers/v37/jiao15.html`.

Y. Jiao and J.-P. Vert. The Kendall and Mallows kernels for permutations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. doi: 10.1109/TPAMI.2017.2719680. URL `http://dx.doi.org/10.1109/TPAMI.2017.2719680`.

W. R. Knight. A computer method for calculating Kendall's tau with ungrouped data. *J. Am. Stat. Assoc.*, 61(314):436–439, 1966. URL `http://www.jstor.org/stable/2282833`.

R. I. Kondor and M. S. Barbosa. Ranking with kernels in fourier space. In A. T. Kalai and M. Mohri, editors, *COLT 2010 - The 23rd Conference on Learning Theory, Haifa, Israel, June 27-29, 2010*, pages 451–463. Omnipress, 2010.

R. Kumar and S. Vassilvitskii. Generalized distances between rankings. In *Proceedings of the 19th International Conference on World Wide Web (WWW-10)*, pages 571–580. ACM, 2010. doi: 10.1145/1772690.1772749.

M. Le Morvan and J.-P. Vert. Supervised quantile normalisation. Technical Report 1706.00244, arXiv, 2017.

G. S. Shieh. A weighted Kendall's tau statistic. *Statistics & Probability Letters*, 39(1):17–24, 1998. doi: 10.1016/s0167-7152(98)00006-6. URL `http://dx.doi.org/10.1016/S0167-7152(98)00006-6`.

S. Vigna. A weighted correlation index for rankings with ties. In *Proceedings of the 24th International Conference on World Wide Web (WWW-15)*, pages 1166–1176. ACM, 2015. doi: 10.1145/2736277.2741088.