

MVA "Kernel methods"

Homework 3

Jean-Philippe Vert

Due February 15, 2012

Exercise 1.

The 2-SVM with offset is a method for supervised binary classification. Given a training set $(x_i, y_i)_{i=1, \dots, n}$ of training patterns x_1, \dots, x_n in a space X endowed with a positive definite kernel K , and a set of corresponding labels $y_1, \dots, y_n \in \{-1, 1\}$, it solves the following problem:

$$\min_{f \in H_K, b \in \mathbb{R}} \left\{ \frac{1}{n} \sum_{i=1}^n L(f(x_i) + b, y_i) + \lambda \|f\|^2 \right\},$$

where $\|f\|$ is the norm of f in the RKHS H_K of the kernel K , and L is the *square hinge loss function*:

$$L(u, y) = \max(1 - uy, 0)^2.$$

Write the primal and dual problems associated to the 2-SVM, and interpret the values of the dual variables. Compare the result with the SVM studied in the course (1-SVM without offset).

Exercise 2.

Download the *Libras Movement Data Set*¹. Each instance is a video-recording of hand movements (for Brazilian sign language) encoded in 90 descriptors. The dataset contains 15 classes (different movements) with 24 instances for each class. Your objective is to design an automatic hand movement recognizer using a SVM and to evaluate its performance.

¹<http://archive.ics.uci.edu/ml/datasets/Libras+Movement>

Randomly split the full dataset into 80% of training examples (to train a model) and 20% of test examples (to assess its performance). Train a SVM on the training set with different kernels (e.g., linear, polynomial, Gaussian) and different regularization parameters. Note that you need to propose a strategy to use a binary SVM in order to solve a classification problem with 14 classes. For each kernel, plot the accuracy (percentage of good prediction) on the training set and on the test set, as a function of the regularization parameter, and comment the results.

Remarks: Use the software of your choice, there are many implementations of SVM that you will find on Google. I recommend libsvm which can be run from the command line, in python, matlab, R etc... Alternatively you can use environments for machine learning such as Spider on Matlab or PyML on Python. In R, the package svmPath computes the SVM solutions for all values of the regularization parameter in a single command, which can be useful here. Alternatively the kernlab package works well too².

²Tutorial with R and kernlab available at
<http://cbio.enscm.fr/~jvert/teaching/2011mines/index.html>