

Homework 5

Jean-Philippe Vert

Due March 9, 2010

1 Playing with SVM

The goal is to play a bit with support vector machines. Download the spam data from the course homepage (`data.txt` and `labels.txt`). This is a classical pattern recognition problem. Randomly select a training set of 100 points, and use the rest as a test set. Train a SVM on the training set with different kernels (e.g., linear, polynomial, Gaussian) and different regularization parameters. 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.

2 Sobolev norm

Let $H = C_2([0, 1])$ be the set of twice continuously differentiable functions $f : [0, 1] \rightarrow \mathbb{R}$, and $H_1 \subset H$ be the set of functions $f \in H$ that satisfy:

$$f(0) = f'(0) = 0.$$

Show that H_1 endowed with the norm:

$$\|f\|_{H_1}^2 = \int_0^1 f''(t)^2 dt$$

is a reproducing kernel Hilbert space (RKHS), and compute the reproducing kernel K_1 .