Mehryar Mohri
Foundations of Machine Learning
Courant Institute of Mathematical Sciences
Homework assignment 2
Due: March 17, 2009

A. Support Vector Machines

1. Download and install the `libsvm` software library from:

   `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`

2. Download the `satimage` data set. Merge the training and validation set into one. We will refer to the resulting set as the training set from now on. Normalize all input vectors (training and test).

3. Consider the binary classification that consists of distinguishing class 6 from the rest. Use SVMs combined with polynomial kernels to solve this classification problem.

   To do that, randomly split the training data into ten equal-sized disjoint sets. For each value of the polynomial degree, $d = 1, 2, 3, 4$, plot the average cross-validation error plus or minus one standard deviation as a function of C (let the other parameters of polynomial kernels in `libsvm`, γ and c, be equal to their default values 1). Report the best value of the trade-off constant C measured on the validation set.

4. Let (C^*, d^*) be the best pair found previously. Fix C to be C^*. Plot the ten-fold cross-validation training and test errors for the hypotheses obtained as a function of d. Plot the average number of support vectors obtained as a function of d.

5. How many of the support vectors lie on the margin hyperplanes?

6. In the standard two-group classification, errors on positive or negative points are counted the same way. Suppose we wish to penalize an error on an negative point (false positive error) $k > 0$ times more than an error on a positive point. Give the dual optimization problem corresponding to SVMs modified in this way.

7. Assume that k is an integer. Show how you could use libsvm without writing any additional code to find the solution of the modified SVMs just described.

8. Apply the modified SVMs to the classification task previously examined and compare with your previous SVMs results for different values of k: $k = 2, 4, 8, 16$.

B. Kernels

1. Let $K: X \times X \to \mathbb{R}$ be a PDS kernel, show that for all $x, y \in X$,

$$|K(x,y)| \le \sqrt{K(x,x)K(y,y)}. \tag{1}$$

2. Show that the following kernels K are PDS:

    (a) $K(x,y) = \cos(x-y)$ over $\mathbb{R} \times \mathbb{R}$.

    (b) $K(x,y) = \cos(x^2-y^2)$ over $\mathbb{R} \times \mathbb{R}$.

    (c) $K(x,y) = (x+y)^{-1}$ over $]0,+\infty[\times]0,+\infty[$.

3. Show that the following kernels K are NDS:

    (a) $K(x,y) = [\sin(x-y)]^2$ over $\mathbb{R} \times \mathbb{R}$.

    (b) $K(x,y) = \log(x+y)$ over $]0,+\infty[\times]0,+\infty[$.

4. Is the kernel K defined over $\mathbb{R}^n \times \mathbb{R}^n$ by $K(x,y) = \|x-y\|^{3/2}$ PDS? Is it NDS? Prove your responses.