Mehryar Mohri
Foundations of Machine Learning
Courant Institute of Mathematical Sciences
Homework assignment 3
Due: March 28th, 2008
Credit: Ashish Rastogi, Afshin Rostamizadeh
Ameet Talwalkar, and Eugene Weinstein.

1. **SVMs**:

   (a) Download and install `libsvm` from

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

   (b) Download the `pendigits` data set. The task is to predict the digit label $(0-9)$ based on the features computed over the digit image. The data is comma-delimited, with the last item being the label. Normalize the input data so that all feature values are between -1 and 1.

   (c) Train and test a SVM using polynomial kernels and 10-fold cross validation. For each setting of the polynomial degree $d = 1, 2, 3, 4$, plot the average error as the data set size is changed from 50 to 1000 data points (keep the first n points of the data set).

   (d) Repeat the learning experiment with radial basis function (RBF) kernels. Use the script `grid.py` packaged with `libsvm` to do a sweep over the space of parameters (C, γ), where C is the SVM learning parameter and γ is the coefficient in the RBF kernel. Report the values of C and γ that yield the highest accuracy under 10-fold cross validation. Also report the accuracy achieved.

   (e) Let (C^*, γ^*) be the best parameters found in the previous exercise. With C fixed at C^*, plot the 10-fold cross-validation accuracy as the γ parameter is varied.

   (f) Suppose you wish to use support vector machines to solve a learning problem where some training data points are more important than others. More formally, assume that each training point consists of a triplet (x_i, y_i, p_i), where $0 \leq p_i \leq 1$ is the importance of the ith point. Rewrite the primal SVM constrained optimization problem so that the penalty for mis-labeling a point x_i is scaled by the priority p_i. Then carry this modification through the derivation of the dual solution.

2. **Kernels:**

   (a) Given a data set $x_1, \ldots, x_m$ and a kernel $k(x_i, x_j)$ with a Gram matrix K such that $K_{ij} = k(x_i, x_j)$. Give a map $\Phi(\cdot)$ such that if K is positive semidefinite then $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$.

   (b) Show the converse of the previous statement: that if there exists a mapping $\Phi(x)$, then the matrix K is positive semidefinite.

   (c) Let us define a *difference kernel* as $k(x, x') = ||x - x'||$ for $x, x' \in \mathbb{R}^n$. Show that this kernel is not positive definite symmetric (PDS).

   (d) The *cosine kernel* is defined as $k(x, x') = \cos \angle(x, x')$. Show that the cosine kernel is PDS.