Mehryar Mohri
Introduction to Machine Learning
Courant Institute of Mathematical Sciences
Homework assignment 3
November 05, 2011
Due: November 19, 2011

**Support vector machines**

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

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

2. Download the `Spambase` data set:

   `http://archive.ics.uci.edu/ml/datasets/Spambase`

   Use the `libsvm` scaling tool to scale the features of all the data. Use the first 3067 instances for training, the rest (1534 instances) for testing.

3. Use SVMs combined with polynomial kernels to solve the classification problem.

   To do that, randomly split the training data into ten equal-sized disjoint sets. For each value of the polynomial degree, $d = 1, 2, 3$, 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`, $\gamma$ and $c$, be equal to their default values). 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. Suppose we wish to penalize false positive errors (non-spam messages labeled as spam) twice as much as false negatives. How would you modify the data without writing new code to determine the correct solution?

**Sequence kernels**

1. Let $\Sigma = \{a, b\}$ and let $|x|$ denote the length of a sequence $x$ in $\Sigma^*$.

   For any $x \in \Sigma^*$, let $c_k(x)$ denote the number of times the prefix of length $k$ of $x$ appears in $x$ where $1 \leq k \leq |x|$. We also define $c_k(x) = 0$ for $k > |x|$. For example, for $x = ababa$, $c_1(x) = 3$, $c_2(x) = 2$ and $c_3(x) = 2$, $c_4(x) = c_5(x) = 1$ and $c_6(x) = 0$. Let $K$ be the kernel defined over $\Sigma^* \times \Sigma^*$ by

   $$K(x, x') = \sum_{k \geq 1} c_k(x) c_k(x'). \tag{1}$$

   for all $(x, x') \in \times\Sigma^* \times \Sigma^*$.

   (a) Show that $K$ is a rational kernel of the form $T \circ T^{-1}$ (show the corresponding transducer $T$).

   (b) Suppose we wish to modify the similarity measure to assign larger weights to longer prefixes in the following way by defining $K_\lambda$ for all $(x, x')$ by

   $$K_\lambda(x, x') = \sum_{k \geq 1} \frac{1}{\lambda^{2k}} c_k(x) c_k(x'). \tag{2}$$

   where $\lambda \in (0, 1)$ is fixed. Show how a transducer $T_\lambda$ can be derived from the transducer $T$ of the previous question to obtain $K_\lambda = T_\lambda \circ T_\lambda^{-1}$.