Quiz 2, ECE7252, Spring 2008

1. Consider the following kernel function: $K(x_i, x_j) = (< x_i, x_j >)^2$, where $< x, y >= x^T y$ denotes the inner product of vectors x and y. Verify that for each of the following two feature mappings $\phi(x)$, it holds that $K(x_i, x_j) =< \phi(x_i), \phi(x_j) >$. Note in R^2, each vector $x_i = (x_{i1}, x_{i2})^T$. Show your calculation.

   (a)
$$\phi(x) : R^2 \rightarrow R^3, \phi(x) = \frac{1}{\sqrt{2}}(x_1^2 + x_2^2, 2x_1x_2, x_1^2 - x_2^2)^T; \tag{1}$$

   (b)
$$\phi(x) : R^2 \rightarrow R^4, \phi(x) = (x_1^2, x_1x_2, x_1x_2, x_2^2)^T. \tag{2}$$

2. We have a collection of data points X that we want to classify as data belonging to Class 1 (desired output +1), and another set of data points Y that we want to classify as data belonging to Class 2 (desired output -1). We can formulate this as a 2-class classification problem and use the single layer perceptron neural network to solve it. To make our objective uniform for every y point we define a corresponding point x, such that $y = -x$. Then the perceptron learning scenario becomes a problem where we can formulate learning such that $w^T x \geq 0$ for all x in X. Show that we can reformulate it as a minimization problem where the objective function is of the form: $J(w) = \frac{1}{2}(|w^T x| - w^T x)$. Show that this is equivalent to Rosenblatt's perceptron learning algorithm which was discussed in Slides 67-73, Lectures 10-11.

3. Consider a simple linear SVM classifier, $(w_1 x + w_0)$, a nonlinear SVM classifier, $(w_1\phi(x) + w_0)$, where $\phi(x) = (x, x^2)^T$. (a) Provide three inputs, x_1, x_2, x_3, and their associate labels (+1 or -1) such that they cannot be perfectly separated with the simple linear classifier, but are separable by the nonlinear classifier. (b) Mark the three points in the new feature space and the separating line.

4. In the SVM formulation we try to minimize $||w||^2$ such that $y_i(w^T x_i + b) \geq 1, i = 1, ..., n$ for the set of training examples denoted by $D = \{(x_i, y_i), i = 1, ..., n | y_i = -1 \text{ or } +1\}$. We can also consider from the viewpoint of robust optimization that every observed example,

x_i, is corrupted by noise, and the true example, x_i^*, is within a radius, ρ, of the observed sample, x_i, or $||x_i^* - x_i||^2 \leq \rho^2$. Since we are not sure which example within the above sphere is a true example, we require every example in the sphere to be classified correctly, or for all $||x_i^* - x_i||^2 \leq \rho^2$, $y_i(w^T x_i^* + b) \geq 0$. We now have an alternative optimization problem by searching for the decision boundary with the largest ρ, i.e. to maximize ρ, such that $||x_i^* - x_i||^2 \leq \rho^2$, $y_i(w^T x_i^* + b) \geq 0, i = 1, ..., n$.

5. The World Series is a seven-game series that terminates as soon as either team wins four games. Let X be the random variable that represents the outcome of a World Series between teams A and B; possible values of X are AAAA, BABABAB, and BBBAAAA, etc. Let Y be the number of games played in the series, which range from 4 to 7. Assuming that A and B are equally matched and that the games are independent, calculate $H(X)$, $H(Y)$, $H(Y|X)$, $H(X|Y)$ and $I(X,Y)$ [Hint: you need to compute $P(X)$ and $P(Y)$ based on all possible outcomes].