E1 213 Pattern Recognition and Neural networks

Problem Sheet 8

1. Consider the following pattern recognition problem in $\Re^2$:

   Class $+1$: (1,0), (0,1)
   Class -1: (0,0)

   a. **Guess** the optimal separating hyperplane. Give some intuitive explanation for why it is the optimal hyperplane. Then solve the optimization problem and verify your guess. (Here, you are using a linear SVM and hence it would be easier for you to solve the primal).

   b. Suppose we add one more pattern of class '+1' which is given by (1,1). Will the optimal hyperplane change? Explain your answer.

   c. In the above, suppose we change the separability constraints to $y_i[w^T x_i + b] \geq K$ where K is some fixed positive number. Would the equation of the optimal separating hyperplane change? Explain.

2. If $W^T X + b = 0$ is the equation of an optimal hyperplane, then $(kW)^T X + kb = 0$ is also the equation of an optimal hyperplane. Does this mean that kW, kb is another optimal solution to the SVM problem?

3. Consider a pattern recognition problem in $\Re^2$, for an SVM, with the following training samples:

   Class $+1$: $(0.5, 0.5)$, $(0, 2)$, $(0, 1)$, $(2, 2)$
   Class -1: $(0, 0)$

   (a) Consider the hyperplane given by $W = [2\ 2]^T$ and $b = -1$. Is this a separating hyperplane? Do you think this is a separating hyperplane with maximum margin?

   (b) Suppose we want to solve this problem using the slack variables, ξ_i. That is, we want to maximize $0.5 W^T W + C \sum \xi_i$ subject to constraints $1 - y_i[W^T x_i + b] - \xi_i \leq 0$ and $\xi_i \geq 0$. Take $C = 1$. For the hyperplane given by $W = [2\ 2]^T$ and $b = -1$, find the smallest values for ξ_i, $i = 1, 2, 3, 4, 5$ so that all constraints are satisfied.

At these W, b, ξ_i what is the value of the objective function being minimized. Now consider another hyperplane given by $W = [0\ 1]^T$ and $b = -1$. (That is, the hyperplane is a line parallel to x-axis and passing through (0,1)). For these W, b find the smallest possible values for ξ_i so that all constraints are satisfied. What is the value of the objective function at these W, b, ξ_i. Based on all this, can you say whether the separating hyperplane given in part (a) could be the optimal solution to the optimization problem (which includes the slack variables ξ_i).

(c). Suppose in a 2-class problem the training data is linearly separable. We obtain the optimal solution of the C-SVM (that is, the SVM with slack variables). Would this necessarily be a separating hyperplane?

4. For a linear SVM, let W^*, b^* be the optimal hyperplane and let μ_i^* be the optimal lagrange multipliers. Show that

$$(W^*)^T W^* = \sum_i \mu_i^*$$