COM 599200 Statistical Learning
Homework #4

(Due June 11, 2018 at the beginning of class.)

Note: Detailed derivations are required to obtain a full score for each problem. (Total 80%)

1. (6%) Problem 5 of Chapter 8.

2. (10%) Suppose you wish to carry out a linear discriminant analysis (two classes) using a vector of transformations of the input variables $\phi(x)$. Since $\phi(x)$ is high-dimensional, you will use a regularized within-class covariance matrix $\mathbf{W}_\phi + \gamma\mathbf{I}$. Show that the model can be estimated using just the inner products $K(x, x') = \langle \phi(x), \phi(x') \rangle$. Hence, the kernel property of support vector machines is also shared by regularized linear discriminant analysis.

3. (4%+8%+6%+4%+8%) Suppose you are given 6 one-dimensional points: 3 with negative labels $x_1 = -1$, $x_2 = 0$, $x_3 = 1$ and 3 with positive labels $x_4 = -3$, $x_5 = -2$, $x_6 = 3$. Note that the data cannot be perfectly separated in $\mathbb{R}$, but, by applying the following feature map $\phi(x) = (x, x^2)$ (which transforms points in $\mathbb{R}$ to points in $\mathbb{R}^2$), a linear separator can perfectly separate the points in the new $\mathbb{R}^2$ feature space induced by ϕ.

(a) Find the kernel $K(x, x')$ that corresponds to the feature map ϕ.

(b) Construct a maximal margin separating hyperplane in the new feature space. This hyperplane will be a line in $\mathbb{R}^2$, which can be parameterized by its normal equation, i.e. $w_1 Y_1 + w_2 Y_2 + c = 0$ for appropriate choices of w_1, w_2, and c. Here, $(Y_1, Y_2) = \phi(X)$ is the result of applying the feature map ϕ to the original feature X. Give the values for w_1, w_2, and c, and compute the margin for your hyperplane.
(Hint: You do not need to solve any optimization problem to find the maximum margin hyperplane. Note that the line must pass somewhere between $(-2, 4)$ and $(-1, 1)$. There will only be two support vectors in this case.)

(c) Apply ϕ to the data and plot the points in the new $\mathbb{R}^2$ feature space. On the plot of the transformed points, plot the separating hyperplane and the margin, and circle the support vectors.

(d) Draw the decision boundary of the separating hyperplane in the original $\mathbb{R}$ feature space.

(e) Suppose that the point $x_7 = 4$ is added to the negative class. Find the support vector classifier for $C = 4$.

4. (12%) Problem 2 (a)-(d) of Chapter 10 in the textbook with the value 0.8 replaced with 0.6.

5. (12%) Consider Problem 3 of Chapter 10 in the textbook with Observation 2 changed to $(X_1, X_2) = (1, 2)$ and Observation 4 changed to $(5, 3)$. Initialize by taking the first 3 observations as the first cluster and the other 3 observations as the second cluster. Plot the observations and their class labels at the initialization and also after each iteration of the K-means clustering algorithm.

6. (10%) Consider a special case of the Gaussian mixture model in which the covariance matrices $\mathbf{\Sigma}_k$ of the components are all constrained to have a common value $\mathbf{\Sigma}$. Derive the EM equations for maximizing the likelihood function under such a model.