COM 525000 Statistical Learning
Homework #4

(Due January 7, 2020 before noon to the TA at EECS 613.)

Note: Detailed derivations are required to obtain a full score for each problem. (Total 70%)

1. (10%) Let us consider the generalized additive model

$$y_i = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \epsilon_i,$$

for $i = 1, \ldots, n$. Here, we use the backfitting algorithm to fit a cubic polynomial for f_1 and a ridge regression model with tuning parameter λ for f_2. Find the fits for f_1, f_2, and β_0 in the first iteration of the backfitting algorithm.

2. (8%+8%) Suppose we fit a curve with basis functions $b_1(X) = X$, $b_2(X) = (X - 1)^2 I(X \geq 1)$. (Note that $I(X \geq 1)$ equals 1 for $X \geq 1$ and 0 otherwise.) We fit the linear regression model

$$Y = \beta_0 + \beta_1 b_1(X) + \beta_2 b_2(X) + \epsilon.$$

(a) Suppose that we obtain coefficient estimates $\hat{\beta}_0 = 1$, $\hat{\beta}_1 = -1$, $\hat{\beta}_2 = 2$. Sketch the estimated curve between $X = -2$ and $X = 2$. Note the intercepts, slopes, and other relevant information.

(b) By fitting the model to the three data points $(X, Y) = (-2, -1), (1, 2), (2, 7)$, find the resulting coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$.

3. (4%+8%+6%+4%) 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.

4. (6%+6%+10%) (a) Problem 4 of Chapter 8 in the textbook.

(b) Suppose that the data set associated with Problem 4(a) is $\mathcal{D} = \{(x_{i1}, x_{i2}, y_i)\}_{i=1}^{8} = \{(\frac{1}{8}, \frac{1}{4}, 3), (\frac{3}{8}, \frac{1}{8}, 10), (\frac{3}{8}, \frac{3}{8}, 0), (\frac{1}{8}, \frac{7}{8}, 12), (\frac{5}{16}, \frac{9}{16}, 18), (\frac{5}{8}, \frac{1}{4}, 4), (\frac{7}{8}, \frac{1}{4}, 2), (\frac{6}{8}, \frac{3}{4}, 9)\}$ and that the tree obtained in Problem 4(a) is the tree obtained in the first iteration of the boosting algorithm in Algorithm 8.2 of the text book, where $\lambda = 0.5$. Find the tree (with 3 splits) in the second iteration of the boosting algorithm.