COM 525000 Statistical Learning
Homework #3

(Due December 24, 2020 at the beginning of class.)

Note: Detailed derivations are required to obtain a full score for each problem. (Total 100%)

1. (10%+6%+8%) Suppose that the input and output variables of the n training data points can be expressed as $\mathbf{X} = (x_1, \ldots, x_n)^T$ and $\mathbf{y} = (y_1, \ldots, y_n)^T$, respectively. In ridge regression, the intercept β_0 and the coefficient vector $\beta = (\beta_1, \ldots, \beta_p)^T$ are chosen such that $\|\mathbf{y} - \mathbf{X}\beta - \mathbf{1}\beta_0\|^2 + \lambda\|\beta\|^2$ is minimized for some $\lambda \geq 0$, where $\mathbf{1}$ is an n-dimensional all-one vector.

(a) Show that the resulting coefficient estimate is given by

$$\hat{\beta}_\lambda = (\mathbf{X}_c^T\mathbf{X}_c + \lambda\mathbf{I})^{-1}\mathbf{X}_c^T\mathbf{y}_c,$$

where $(\mathbf{X}_c, \mathbf{y}_c)$ is the centered data set.

(b) Show that $\|\beta_\lambda\|\big|_{\lambda>0} \leq \|\beta_\lambda\|\big|_{\lambda=0}$

(c) Show that the training error is

$$\overline{\text{err}} = \frac{1}{n}\mathbf{y}_c\big[\mathbf{I} - \mathbf{X}_c(\mathbf{X}_c^T\mathbf{X}_c + \lambda\mathbf{I})^{-1}\mathbf{X}_c^T\big]^2\mathbf{y}_c$$

and that it is an increasing function of λ.

2. (8%+8%) Let us consider the data set

$$\mathcal{D} = \{(x_{i1}, x_{i2}, y_i)\}_{i=1}^n = \{(0, -2, -1), (3, 1, 2), (1, 2, 4)\}.$$

(a) Find the solution for lasso with regularizer weight $\lambda = 10$.

(b) Find the solution for local regression with $k = 2$ using only the first predictor (i.e., x_{i1}, for $i = 1, \ldots, n$). Plot your solution.

3. (8%+6%+8%) Consider $n = 4$ data points $\{(x_{i1}, x_{i2}, x_{i3})\}_{i=1}^n$ given by $(-1, -1, -1)$, $(0, 0, 0)$, $(2, 1, -1)$, $(1, 1, 0)$.

(a) What are the first and second principal components (write down the actual vector)?

(b) Find the projection of the data points onto the 1D subspace associated with the first principal component direction, and compute the variance of the projected data?

(c) Suppose that $y_1 = -3$, $y_2 = 1$, $y_3 = 5$, and $y_4 = 2$. Find the first partial least squares (PLS) direction and compare with the first principal component direction.

4. (4%+8%) Let us consider the basis representation for cubic splines with K interior knots, where

$$f(X) = \sum_{j=0}^{3} \beta_j X^j + \sum_{k=1}^{K} \beta_{k+3}(X - \xi_k)_+^3.$$

(a) Prove that the linear boundary conditions for natural cubic splines imply that $\beta_2 = \beta_3 = 0$ and $\sum_{k=1}^{K} \beta_{k+3} = 0$, $\sum_{k=1}^{K} \xi_k \beta_{k+3} = 0$.

(b) Show that the natural cubic spline can be represented by the K basis functions $b_0(X) = 1$, $b_1(X) = X$, and $b_{k+1}(X) = d_k(X) - d_{K-1}(X)$, for $k = 1, \ldots, K-1$, where

$$d_k(X) = \frac{(X - \xi_k)_+^3 - (X - \xi_K)_+^3}{\xi_K - \xi_k}.$$

5. (8%+8%) Consider the labeled data set $\mathcal{D} = \{(x_{i1}, x_{i2}, y_i)\}_{i=1}^n = \{(-1,-1,1), (-3,0,0), (2,-1,4), (1,0,2), (-1,2,0)\}$.

(a) 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 piecewise linear regression with a single knot at 0 for f_1 and an ordinary least square for f_2, respectively. Find the fits for f_1 and f_2 in the first iteration of the backfitting algorithm.

(b) Using the recursive binary splitting algorithm, find a regression tree that splits the predictor space intwo 3 regions, and provide the estimate in each region.

6. (2%+2%+2%+2%+2%) Problem 1 of Chapter 7.