COM 599200 Statistical Learning
Homework #2

(Due April 13, 2018 noon to the TA at EECS 613)

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

1. (6%) Solve Problem 3(c) of HW #1 using QDA.

2. (8%) Solve Problem 7 of Chapter 4, but with the observed variance being $\hat{\sigma}^2 = 25$ for those that issued a dividend and $\hat{\sigma}^2 = 36$ for those that didn't.

3. (12%) Let

$$\mathrm{CV}_{(n)} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i; \mathcal{D} \setminus \{(x_i, y_i)\}\right)^2$$

be the leave-one-out cross-validation (LOOCV) error. Show that

$$\mathrm{CV}_{(n)} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{1 - h_i}\right)^2$$

where $\hat{y}_i = \hat{f}(x_i; \mathcal{D})$ is the i-th fitted value from the original least square fit (using the entire data set $\mathcal{D}$), and h_i is the leverage statistic.
(Hint: Fill in the details of the sketch proof shown in class.)

4. (10%+10%) 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 coefficient vector $\beta = (\beta_0, \ldots, \beta_p)^T$ is chosen such that $\|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda\|\beta\|^2$ is minimized for some $\lambda \geq 0$.

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

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

and that $\|\beta_\lambda\|\big|_{\lambda>0} \leq \|\beta_\lambda\|\big|_{\lambda=0}$

(b) Show that the training error is

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

and that it is an increasing function of λ.

5. (10%+8%) Suppose that the available data set is $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), (x_3, y_3)\} = \{(1, 2), (3, 7), (5, 8)\}$, and that B bootstrap data sets $\mathcal{D}^{*1}, \ldots, \mathcal{D}^{*B}$ (excluding those with only 1 distinct data point) are generated from $\mathcal{D}$, each with the same size as $\mathcal{D}$. Linear regression is performed on each data set to obtain coefficient estimates $\beta^{*1}, \ldots, \beta^{*B}$, where

(a) For $B \to \infty$, find the standard errors of the coefficient estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.

(b) Compare with the standard error estimates derived in Chapter 3 for original least squares (i.e., ridge regression with $\lambda = 0$), that is,

$$\widehat{\mathrm{SE}}(\hat{\beta}_0)^2 = \hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

and

$$\widehat{\mathrm{SE}}(\hat{\beta}_1)^2 = \frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

where $n = 3$ in this case.

6. (16%) Problem 5 of Chapter 6 with $x_{11} = -x_{12}$ and $x_{21} = -x_{22}$ and, in (b), $\hat{\beta}_1 = -\hat{\beta}_2$.