6.8.3

(a) As we increase s from 0, the training RSS will:

answer: Steadily decrease.

$\hat{\beta} = \arg\min \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2$, subject to the constraint that $\sum_{j=1}^{p} |\beta_j| \le s$. So, when $s=0$, the training RSS will be the maximum nub. when s increases, the train RSS will decrease.

(b) For test RSS, it will decrease at first and then strat increasing in a U shape. When $s=0$, the β all are 0. When s increases, the model becomes more & more flexible, the test RSS will decrease at first. But when move over a certain value, the model will overfit the traindata, and test RSS will increase again.

(c) For variance, it will steadily increase. When $s=0$, there is no variance in the model. When s increases, the model becomes more flexible. The estimator of β relies on the training data, so the variance increases.

(d) For (squared) bias, it will steadily decrease. The bias is opposite to the variance. When $s=0$, it has the largest bias. When the model becomes more flexible, the bias will decrease.

(e) For the irreducible error, it will remain constant. The irreducible error would be the error caused by the inherent noise. It will not change with the s.

6.8.5

(a) Ridge - optimization problem

$x_{11} = x_{12} = x \quad , x_{21} = x_{22} = -x \quad , y_1 = -y_2$

$$f(\beta_1, \beta_2) = \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

$$= (y_1 - \beta_0 - \beta_1 x_{11} - \beta_2 x_{12})^2 + (y_2 - \beta_0 - \beta_1 x_{21} - \beta_2 x_{22})^2 + \lambda(\beta_1^2 + \beta_2^2)$$

$$= (y_1 - \beta_1 x - \beta_2 x)^2 + (-y_1 - \beta_1(-x) - \beta_2(-x))^2 + \lambda(\beta_1^2 + \beta_2^2)$$

$$= 2(y_1 - \beta_1 x - \beta_2 x)^2 + \lambda(\beta_1^2 + \beta_2^2)$$

(b) Ridge coefficient estimates satisfy $\hat{\beta}_1 = \hat{\beta}_2$

$$\frac{\partial f}{\partial \beta_1} = 4(y-\beta_1 x-\beta_2 x)(-x)+2\lambda\beta_1 = 0$$

$$\frac{\partial f}{\partial \beta_2} = 4(y-\beta_1 x-\beta_2 x)(-x)+2\lambda\beta_2 = 0$$

$$2(y-\beta_1 x-\beta_2 x)-\lambda\beta_1 = 0$$

$$2(y-\beta_1 x-\beta_2 x)-\lambda\beta_2 = 0$$

$\therefore$ So the ridge coefficient estimates satisfiy

$$\lambda\beta_1 = \lambda\beta_2$$

$$\hat{\beta}_1 = \hat{\beta}_2$$

(C) Lasso optimization problem

$$g(\beta_1, \beta_2) = \sum_{i=1}^{n}(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij})^2+\lambda\sum_{j=1}^{p}|\beta_j|$$

$$= 2(y-\beta_1 x-\beta_2 x)^2+\lambda(|\beta_1|+|\beta_2|)$$

(d) $(\hat{\beta}_1, \hat{\beta}_2) = \underset{(\beta_1, \beta_2)}{\arg\min}\ 2(y-\beta_1 x-\beta_2 x)^2$, subject to constraint $|\beta_1|+|\beta_2| \le s$

when $\beta_1+\beta_2 = \frac{y}{x}$, there's no restriction

$\therefore$ $|\frac{y}{x}| \le s$, $\hat{\beta}_1+\hat{\beta}_2 = \frac{y}{x}$

$|\frac{y}{x}| > s$, $\hat{\beta}_1+\hat{\beta}_2 = s$, $\hat{\beta}_1 \ge 0$, $\hat{\beta}_2 \ge 0$, if $\frac{y}{x} > 0$; $\hat{\beta}_1+\hat{\beta}_2 = -s$,

$\hat{\beta}_1 \le 0$, $\hat{\beta}_2 \le 0$, if $\frac{y}{x} < 0$.

$\therefore$ lasso coefficient are not unique.

8.4.5

For majority vote, 6 value more than 0.5, 6 red & 4 green $\rightarrow$ Red

For average probability, $P(\text{Class is Red}|X) = \frac{0.1+0.15+0.2+0.2+0.55+0.6+0.6+0.65+0.7+0.75}{10}$

$= 0.45 \longrightarrow$ Green