1. $\pi_1 = \frac{31+52+68+101+144}{1000} = 0.396$, $\pi_0 = 0.604$

$$\hat{\mu}_1 = \frac{1}{n_1}\sum_{y_i=1} x_i = \frac{1}{396}(31\times5+52\times10+68\times15+101\times20+144\times30) = 20.2904$$

$$\hat{\mu}_0 = \frac{1}{n_0}\sum_{y_i=0} x_i = \frac{1}{604}(169\times5+148\times10+132\times15+99\times20+56\times30) = 13.1871$$

$$\hat{\sigma}_1^2 = \frac{1}{n_1-k}\sum_{y_i=1}(x_i-\hat{\mu}_1)^2$$

$$= \frac{1}{396-1}\left[31\times(5-20.2904)^2+52\times(10-20.2904)^2 + 68\times(15-20.2904)^2+101\times(20-20.2904)^2+144\times(30-20.2904)^2\right]$$

$$= 71.4977$$

$$\hat{\sigma}_2^2 = \frac{1}{n_2-k}\sum_{y_i=0}(x_i-\hat{\mu}_0)^2$$

$$= \frac{1}{604-1}\left[169\times(5-13.1871)^2+148\times(10-13.1871)^2 + 132\times(15-13.1871)^2+99\times(20-13.1871)^2+56\times(30-13.1871)^2\right]$$

$$= 55.8704$$

$$p_1(x) = \frac{\pi_1 \frac{1}{\sqrt{2\pi\hat{\sigma}_1^2}} \exp\left[-\frac{1}{2\hat{\sigma}_1^2}(x-\mu_1)^2\right]}{\sum_{l=0}^{1}\pi_l \frac{1}{\sqrt{2\pi\hat{\sigma}_l^2}} \exp\left[-\frac{1}{2\hat{\sigma}_l^2}(x-\mu_l)^2\right]}$$

$x=25 \Rightarrow p_1(25) = 0.6337$

2、 X : profit

$Y=1$ (Yes) $\quad P(Y=1)=0.8$

$Y=0$ (No) $\quad P(Y=0)=0.2$

$X|Y=1 \sim N(10,25)$

$X|Y=0 \sim N(0,36)$

$$P(Y=1|X=4) = \frac{f_{X|Y=1}(x)\,P(Y=1)}{f_X(x)}$$

$$= \frac{f_{X|Y=1}(x)\,P(Y=1)}{f_{X|Y=0}(x)\,P(Y=0) + f_{X|Y=1}(x)\,P(Y=1)}$$

$x=4$

$$= \frac{\frac{1}{\sqrt{2\pi\cdot 25}}\exp\left(\frac{-(4-10)^2}{2\cdot 25}\right)\times 0.8}{\frac{1}{\sqrt{2\pi\cdot 25}}\exp\left(\frac{-(4-10)^2}{2\cdot 25}\right)\cdot 0.8 + \frac{1}{\sqrt{2\pi\cdot 36}}\exp\left(\frac{-(4-0)^2}{2\cdot 36}\right)\times 0.2}$$

$$\doteq 0.7448$$

3. Recall that $\hat{y} = \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T y$

(i.e., $\hat{y}_i = \underline{x_i}^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T y$)

When (x_i, y_i) is left out,

$$\hat{y}_i^{(-i)} = \underline{x_i}^T(\underset{\sim}{X}^{(-i)T}\underset{\sim}{X}^{(-i)})^{-1}\underset{\sim}{X}^{(-i)T}y^{(-i)}$$

$$= \underline{x_i}^T(\underset{\sim}{X}^T\underset{\sim}{X} - \underline{x_i}\,\underline{x_i}^T)^{-1}(\underset{\sim}{X}^T y - \underline{x_i}^T y_i)$$

[Apply matrix inversion lemma $(A - \underline{x}\underline{x}^T)^{-1} = A^{-1} + \dfrac{A^{-1}\underline{x}\underline{x}^TA^{-1}}{1-\underline{x}^TA^{-1}\underline{x}}$

$$\Rightarrow (\underset{\sim}{X}^T\underset{\sim}{X} - x_i x_i^T)^{-1} = (\underset{\sim}{X}^T\underset{\sim}{X})^{-1} + \frac{(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}x_i x_i^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}}{1 - \underline{x_i}^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underline{x_i}}$$

$$= (\underset{\sim}{X}^T\underset{\sim}{X})^{-1} + \frac{(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underline{x_i}\,\underline{x_i}^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}}{1-h_i}$$]

$$= \underline{x_i}^T\left[(\underset{\sim}{X}^T\underset{\sim}{X})^{-1} + \frac{(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underline{x_i}\,\underline{x_i}^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}}{1-h_i}\right](\underset{\sim}{X}^T y - \underline{x_i}^T y_i)$$

$$= \underline{x_i}^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T y - \underline{x_i}^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underline{x_i}^T y_i$$

$$+ \frac{(\underline{x_i}^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underline{x_i})(\underline{x_i}^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T y) - (\underline{x_i}^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underline{x_i})(\underline{x_i}^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underline{x_i}^T)y_i}{1-h_i}$$

$$= \hat{y}_i - h_i y_i + \frac{h_i\hat{y}_i - h_i^2 y_i + \hat{y}_i - h_i y_i - \hat{y}_i + h_i y_i}{1-h_i}$$

$$= \hat{y}_i - h_i y_i + \frac{\hat{y}_i - h_i y_i - (\hat{y}_i - h_i y_i)(1-h_i)}{1-h_i} \qquad \left[B + \frac{B - BA}{A} = \frac{B}{A}\right]$$

$$= \frac{\hat{y}_i - h_i y_i}{1-h_i} \quad \cdots (*)$$

Substitute $(*)$ into $CV_{(n)} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i^{(-i)})^2$.

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

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

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

4a
$$L = \| y - X\beta \|^2 + \lambda \|\beta\|^2$$
$$= (y - X\beta)^T (y - X\beta) + \lambda \beta^T \beta$$
$$\frac{\partial L}{\partial \beta} = 0 \Rightarrow -2X^T y + 2X^T X\beta + 2\lambda\beta = 0$$
$$\Rightarrow \hat{\beta} = (X^T X + \lambda I)^{-1} X^T y$$

5(a) - given data set $D = \{\underbrace{(3,7)}_{A}, \underbrace{(5,8)}_{B}, \underbrace{(10,15)}_{C}\}$

according to the given condition, the Bootstrap data sets could be

$- D^{*1} = \{A\ B\ C\} = \{(3,7), (5,8), (10,15)\}$

$\hat{\beta}_1^{*1} = 1.192$, $\hat{\beta}_0^{*1} = 2.85$ (Prob $(D^{*1}) = 6/27$)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$- D^{*2} = \{A\ B\ B\} = \{(3,7), (5,8), (5,8)\}$

Prob $(D^{*i}) = 3/24$; $i = 2 \cdots 7$

$\hat{\beta}_1^{*2} = 0.5$; $\hat{\beta}_0^{*2} = 5.50$

$D^{*3} = \{A\ A\ B\} = \{(3,7), (3,7), (5,8)\}$, $\hat{\beta}_1^{*3} = 0.5$, $\hat{\beta}_0^{*3} = 5.50$

$D^{*4} = \{A\ C\ C\} = \{(3,7), (10,15), (10,15)\}$; $\hat{\beta}_1^{*4} = 1.14$; $\hat{\beta}_0^{*4} = 3.58$

$D^{*5} = \{A\ A\ C\} = \{(3,7), (3,7), (10,15)\}$; $\hat{\beta}_1^{*5} = 1.14$; $\hat{\beta}_0^{*5} = 3.58$

$D^{*6} = \{B\ B\ C\} = \{(5,8), (5,8), (10,15)\}$; $\hat{\beta}_1^{*6} = 1.4$; $\hat{\beta}_0^{*6} = 1$

$D^{*7} = \{B\ C\ C\} = \{(5,8), (10,15), (10,15)\}$; $\hat{\beta}_1^{*7} = 1.4$; $\hat{\beta}_0^{*7} = 1$

we know that standard error of coefficient estimates are

$$\hat{SE}_B(\hat{\beta}_0)^2 = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\beta}_0^{*b} - \bar{\beta}_0^{*}\right)^2, \quad \hat{SE}_B(\hat{\beta}_1)^2 = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\beta}_1^{*b} - \bar{\beta}_1^{*}\right)^2$$

where $\bar{\beta}_{j=0,1}^{*} = \frac{1}{B} \sum_{b=1}^{B} \hat{\beta}_j^{*b}$ is the mean of bootstrap distribution

$\therefore$ as $B \to \infty$ $\bar{\beta}_0^{*} = \text{Mean} = \text{Prob}(D^{*1})\hat{\beta}_0^{*1} + \text{Prob}(D^{*2})\hat{\beta}_0^{*2} + \cdots + \text{Prob}(D^{*7})\hat{\beta}_0^{*7}$

$\Rightarrow \frac{6}{24} \times 2.85 + \frac{3}{24} \times [2 \times 5.50 + 2 \times 3.58 + 2 \times 1] = 3.229$

Similarly $\bar{\beta}_1^{*} = 1.059$

Thus as $B \to \infty$

$$\hat{SE}_B(\hat{\beta}_0) = \sqrt{\text{Variance}} = \Big[\text{Prob}(D^{*1})\big[\hat{\beta}_0^{*1} - \bar{\beta}_0^{*}\big]^2 + \text{Prob}(D^{*2})\big[\hat{\beta}_0^{*2} - \bar{\beta}_0^{*}\big]^2 + \dots + \text{Prob}(D^{*7})\big[\hat{\beta}_0^{*7} - \bar{\beta}_0^{*}\big]^2 \Big]^{1/2}$$

$$= \left[\frac{6}{24}\big[4 - 3.229\big]^2 + \frac{3}{24}\Big\{\big[3.333 - 3.229\big]^2 + \big[3.666 - 3.229\big]^2 + \dots + \big[4.333 - 3.229\big]^2\Big\}\right]$$

$= 1.6117$ —— (1)

Similarly

$$\hat{SE}_B(\hat{\beta}_1) = \Big[\text{prob}(D^{*1})\big[\hat{\beta}_1^{*1} - \bar{\beta}_1^{*}\big]^2 + \text{prob}(D^{*2})\big[\hat{\beta}_1^{*2} - \bar{\beta}_1^{*}\big]^2 + \dots + \text{Prob}(D^{*7})\big[\hat{\beta}_1^{*7} - \bar{\beta}_1^{*}\big]^2\Big]^{1/2}$$

$$= \frac{6}{24}\left(1 - 1.059\right)^2 + \frac{3.6}{24}\left(1 - 1.059\right)^2$$

$= 0.3367$ —— (2)

5 (b) for OLS,

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

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

where $\hat{\sigma}^2 = \frac{RSS}{n-2} = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$

$D = \{(3,7)(5,8)(10,15)\}$; $\hat{\beta}_1 = 1$ $\hat{\beta}_0 = 4$

$RSS = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 =$ $\Rightarrow \hat{\sigma}^2 = \frac{RSS}{3-2} = 1.0385$

$$\hat{SE}(\hat{\beta}_0) = \left(1 \left[\frac{1}{3} + \frac{\bar{x}^2}{\sum_{i=1}^{3} (x - \bar{x})^2} \right] \right)^{1/2} = 1.3357 \quad \text{—— (3)}$$

$$\hat{SE}(\hat{\beta}_1) = \left(\frac{1}{\sum_{i=1}^{3} (x_i - \bar{x})^2} \right)^{1/2} = 0.1999 \quad \text{—— (4)}$$

6 $n=2,\ p=2,\ x_{11}=-x_{12},\ x_{21}=-x_{22}$

$y_1+y_2=0,\ x_{11}+x_{21}=0,\ x_{12}+x_{22}=0$

$\hat{\beta}_0=0$

(a)
Ridge regression: minimize $(y_1-\hat{\beta}_1x_{11}-\hat{\beta}_2x_{12})^2+(y_2-\hat{\beta}_1x_{21}-\hat{\beta}_2x_{22})^2+\lambda(\hat{\beta}_1^2+\hat{\beta}_2^2)$ #

(b) $\hat{\beta}_1=-\hat{\beta}_2$

let $A=(y_1^2+\hat{\beta}_1^2x_{11}^2+\hat{\beta}_2^2x_{12}^2-2\hat{\beta}_1x_{11}y_1-2\hat{\beta}_2x_{12}y_1+2\hat{\beta}_1\hat{\beta}_2x_{11}x_{12})+(y_2^2+\hat{\beta}_1^2x_{21}^2+\hat{\beta}_2^2x_{22}^2-2\hat{\beta}_1x_{21}y_2-2\hat{\beta}_2x_{22}y_2$
$+2\hat{\beta}_1\hat{\beta}_2x_{21}x_{22})+\lambda\hat{\beta}_1^2+\lambda\hat{\beta}_2^2$

$$\frac{\partial A}{\partial\hat{\beta}_1}=(2\hat{\beta}_1x_{11}^2-2x_{11}y_1+2\hat{\beta}_2x_{11}x_{12})+(2\hat{\beta}_1x_{21}^2-2x_{21}y_2+2\hat{\beta}_2x_{21}x_{22})+2\lambda\hat{\beta}_1=0$$

set $x_{11}=-x_{12}=x_1,\ x_{21}=-x_{22}=x_2,\ x_1+x_2=0$

$\Rightarrow (\hat{\beta}_1x_1^2-x_1y_1-\hat{\beta}_2x_1^2)+(\hat{\beta}_1x_2^2-x_2y_2-\hat{\beta}_2x_2^2)+\lambda\hat{\beta}_1=0$

$\Rightarrow \hat{\beta}_1(x_1^2+x_2^2)-\hat{\beta}_2(x_1^2+x_2^2)+\lambda\hat{\beta}_1=x_1y_1+x_2y_2$

$[2\hat{\beta}_1x_1x_2-2\hat{\beta}_2x_1x_2]+\hat{\beta}_1(x_1^2+x_2^2)-\hat{\beta}_2(x_1^2+x_2^2)+\lambda\hat{\beta}_1=x_1y_1+x_2y_2+[2\hat{\beta}_1x_1x_2-2\hat{\beta}_2x_1x_2]$

$\Rightarrow \hat{\beta}_1\cancel{(x_1+x_2)^2}-\hat{\beta}_2\cancel{(x_1+x_2)^2}+\lambda\hat{\beta}_1=x_1y_1+x_2y_2+2\hat{\beta}_1x_1x_2-2\hat{\beta}_2x_1x_2$

$\Rightarrow \hat{\beta}_1=(x_1y_1+x_2y_2+2\hat{\beta}_1x_1x_2-2\hat{\beta}_2x_1x_2)/\lambda$ *

$$\frac{\partial A}{\partial\hat{\beta}_2}=2\hat{\beta}_2x_{12}^2-2x_{12}y_2+2\hat{\beta}_1x_{11}x_{12}+2\hat{\beta}_2x_{22}^2-2x_{22}y_2+2\hat{\beta}_1x_{21}x_{22}+2\lambda\hat{\beta}_2=0$$

set $x_{11}=-x_{12}=x_1,\ x_{21}=-x_{22}=x_2,\ x_1+x_2=0$

$\Rightarrow -\hat{\beta}_1(x_1^2+x_2^2)+\hat{\beta}_2(x_1^2+x_2^2)+\lambda\hat{\beta}_2=-x_1y_1-x_2y_2$

$[-2\hat{\beta}_1x_1x_2+2\hat{\beta}_2x_1x_2]-\hat{\beta}_1(x_1^2+x_2^2)+\hat{\beta}_2(x_1^2+x_2^2)+\lambda\hat{\beta}_2=-x_1y_1-x_2y_2+[-2\hat{\beta}_1x_1x_2+2\hat{\beta}_2x_1x_2]$

$\Rightarrow -\hat{\beta}_1\cancel{(x_1+x_2)^2}+\hat{\beta}_2\cancel{(x_1+x_2)^2}+\lambda\hat{\beta}_2=-(x_1y_1+x_2y_2+2\hat{\beta}_1x_1x_2-2\hat{\beta}_2x_1x_2)$

$\Rightarrow \hat{\beta}_2=-(x_1y_1+x_2y_2+2\hat{\beta}_1x_1x_2-2\hat{\beta}_2x_1x_2)/\lambda$ * $\Rightarrow \hat{\beta}_1=-\hat{\beta}_2$ QED *

6. (c)

Lasso : minimize $(y_1-\hat{\beta}_1 x_{11}-\hat{\beta}_2 x_{12})^2+(y_2-\hat{\beta}_1 x_{21}-\hat{\beta}_2 x_{22}^2)+\lambda(|\hat{\beta}_1|+|\hat{\beta}_2|)$ #

(d) Let $C=(y_1^2+\hat{\beta}_1^2 x_{11}^2+\hat{\beta}_2^2 x_{12}^2-2\hat{\beta}_1 x_{11}y_1-2\hat{\beta}_2 x_{12}y_1+2\hat{\beta}_1\hat{\beta}_2 x_{11}x_{12})$

$+(y_2^2+\hat{\beta}_1^2 x_{21}^2+\hat{\beta}_2^2 x_{22}^2-2\hat{\beta}_1 x_{21}y_2-2\hat{\beta}_2 x_{22}y_2+2\hat{\beta}_1\hat{\beta}_2 x_{21}x_{22})+\lambda|\hat{\beta}_1|+\lambda|\hat{\beta}_2|$

$$\frac{\partial C}{\partial \hat{\beta}_1}=0 \Rightarrow \hat{\beta}_1(x_1+x_2)^2-\hat{\beta}_2(x_1+x_2)^2+\lambda\frac{|\hat{\beta}_1|}{\hat{\beta}_1}=x_1y_1+x_2y_2+2\hat{\beta}_1 x_1x_2-2\hat{\beta}_2 x_1x_2$$

where $\begin{cases} x_{11}=-x_{12}=x_1 \\ x_{21}=-x_{22}=x_2 \\ x_1+x_2=0 \end{cases}$

$$\Rightarrow \frac{|\hat{\beta}_1|}{\hat{\beta}_1}=(x_1y_1+x_2y_2+2\hat{\beta}_1 x_1x_2-2\hat{\beta}_2 x_1x_2)/\lambda \quad \#$$

$$\frac{\partial C}{\partial \hat{\beta}_2}=0 \Rightarrow -\hat{\beta}_1(x_1+x_2)^2+\hat{\beta}_2(x_1+x_2)^2+\lambda\frac{|\hat{\beta}_2|}{\hat{\beta}_2}=-(x_1y_1+x_2y_2+2\hat{\beta}_1 x_1x_2-2\hat{\beta}_2 x_1x_2)$$

where $\begin{cases} x_{11}=-x_{12}=x_1 \\ x_{21}=-x_{22}=x_2 \\ x_1+x_2=0 \end{cases}$

$$\Rightarrow \frac{|\hat{\beta}_2|}{\hat{\beta}_2}=-(x_1y_1+x_2y_2+2\hat{\beta}_1 x_1x_2-2\hat{\beta}_2 x_1x_2)/\lambda \quad \#$$

$$\Rightarrow \frac{|\hat{\beta}_1|}{\hat{\beta}_1}=-\frac{|\hat{\beta}_2|}{\hat{\beta}_2}$$

$\Rightarrow$ Lasso just requires that $\hat{\beta}_1$ and $\hat{\beta}_2$ are opposite sign. #