1.

(a)

Need to use numerical methods to solve β_0, β_1 (such as gradient descent)

$$\beta(k+1) = \beta(k) - \eta \frac{\nabla J(\beta(k))}{\|\nabla J(\beta(k))\|}, \quad \eta \text{ is step size.} \quad \nabla J(\beta(k)) = -\sum_{i=1}^{n} \frac{(2y_i-1)x_i}{1+e^{-(2y_i-1)x_i^T\beta(k)}}$$

① $\beta(k) = [10, 4]$

$$\nabla J(\beta(k)) = -\Bigg[33 \frac{\begin{bmatrix}1\\0.05\end{bmatrix}}{1+e^{-[1\ 0.05]\begin{bmatrix}10\\4\end{bmatrix}}} + 167 \frac{-1\times\begin{bmatrix}1\\0.05\end{bmatrix}}{1+e^{[1\ 0.05]\begin{bmatrix}10\\4\end{bmatrix}}}$$

$$+ 48 \frac{\begin{bmatrix}1\\0.1\end{bmatrix}}{1+e^{-[1\ 0.1]\begin{bmatrix}10\\4\end{bmatrix}}} + 152 \frac{\begin{bmatrix}1\\0.1\end{bmatrix}}{1+e^{[1\ 0.1]\begin{bmatrix}10\\4\end{bmatrix}}}$$

$$+ 62 \frac{\begin{bmatrix}1\\0.15\end{bmatrix}}{1+e^{-[1\ 0.15]\begin{bmatrix}10\\4\end{bmatrix}}} + 138 \cdot \frac{\begin{bmatrix}1\\0.15\end{bmatrix}}{1+e^{[1\ 0.15]\begin{bmatrix}10\\4\end{bmatrix}}}$$

$$+ 99 \frac{\begin{bmatrix}1\\0.2\end{bmatrix}}{1+e^{-[1\ 0.2]\begin{bmatrix}10\\4\end{bmatrix}}} + 101 \frac{\begin{bmatrix}1\\0.2\end{bmatrix}}{1+e^{[1\ 0.2]\begin{bmatrix}10\\4\end{bmatrix}}}$$

$$+ 132 \frac{\begin{bmatrix}1\\0.3\end{bmatrix}}{1+e^{-[1\ 0.3]\begin{bmatrix}10\\4\end{bmatrix}}} + 68 \cdot \frac{\begin{bmatrix}1\\0.3\end{bmatrix}}{1+e^{-[1\ 0.3]\begin{bmatrix}10\\4\end{bmatrix}}} \Bigg] = \begin{bmatrix}-373.9828\\-75.1481\end{bmatrix}$$

$$\|\nabla J(\beta(k))\| = \sqrt{(-373.9828)^2 + (-75.1481)^2} = 381.4582 \Rightarrow \frac{\nabla J(\beta(k))}{\|\nabla J(\beta(k))\|} = \begin{bmatrix}-0.98\\-0.197\end{bmatrix}$$

$$\Rightarrow \beta(k+1) = \begin{bmatrix}10\\4\end{bmatrix} - 0.1\begin{bmatrix}-0.98\\-0.197\end{bmatrix} = \begin{bmatrix}10.098\\4.0197\end{bmatrix}$$

② $\beta(k) = [1, 1]$

$$\nabla J(\beta(k)) = -\Bigg[33 \frac{\begin{bmatrix}1\\0.05\end{bmatrix}}{1+e^{-[1\ 0.05]\begin{bmatrix}1\\1\end{bmatrix}}} + 167 \frac{-1\times\begin{bmatrix}1\\0.05\end{bmatrix}}{1+e^{[1\ 0.05]\begin{bmatrix}1\\1\end{bmatrix}}}$$

$$+ 48 \frac{\begin{bmatrix}1\\0.1\end{bmatrix}}{1+e^{-[1\ 0.1]\begin{bmatrix}1\\1\end{bmatrix}}} + 152 \frac{\begin{bmatrix}1\\0.1\end{bmatrix}}{1+e^{[1\ 0.1]\begin{bmatrix}1\\1\end{bmatrix}}}$$

$$+ 62 \frac{\begin{bmatrix}1\\0.15\end{bmatrix}}{1+e^{-[1\ 0.15]\begin{bmatrix}1\\1\end{bmatrix}}} + 138 \cdot \frac{\begin{bmatrix}1\\0.15\end{bmatrix}}{1+e^{[1\ 0.15]\begin{bmatrix}1\\1\end{bmatrix}}}$$

$$+ 99 \frac{\begin{bmatrix}1\\0.2\end{bmatrix}}{1+e^{-[1\ 0.2]\begin{bmatrix}1\\1\end{bmatrix}}} + 101 \frac{\begin{bmatrix}1\\0.2\end{bmatrix}}{1+e^{[1\ 0.2]\begin{bmatrix}1\\1\end{bmatrix}}}$$

$$+ 132 \frac{\begin{bmatrix}1\\0.3\end{bmatrix}}{1+e^{-[1\ 0.3]\begin{bmatrix}1\\1\end{bmatrix}}} + 68 \cdot \frac{\begin{bmatrix}1\\0.3\end{bmatrix}}{1+e^{-[1\ 0.3]\begin{bmatrix}1\\1\end{bmatrix}}} \Bigg] = \begin{bmatrix}-134.9811\\-38.2393\end{bmatrix} \Rightarrow \frac{\nabla J(\beta(k))}{\|\nabla J(\beta(k))\|} = \begin{bmatrix}-0.962\\-0.273\end{bmatrix}$$

$$\Rightarrow \beta(k+1) = \begin{bmatrix}1\\1\end{bmatrix} - 0.1\begin{bmatrix}-0.962\\-0.273\end{bmatrix} = \begin{bmatrix}1.0962\\1.0273\end{bmatrix}$$

(b)

LDA:

$$\hat{u}_1 = \frac{1}{n_1}\cdot\sum_{y_i=1} x_i = \frac{1}{374}\cdot\left(33\times5 + 48\times10 + 62\times15 + 99\times20 + 132\cdot30\right) = \underline{20.1}$$

$$\pi_1 = \frac{33+48+62+99+132}{1000} = \underline{0.374 \quad , \quad \pi_2 = 0.626}$$

$$\hat{u}_2 = \frac{1}{626}\cdot\left(167\times5 + 152\times10 + 138\times15 + 101\times20 + 68\times30\right) = \underline{13.6}$$

$$\hat{6}^2 = \frac{1}{998}\Big((5-20.1)^2\cdot33 + (5-13.6)^2\cdot167 + (10-20.1)^2\cdot48 + (10-13.6)^2\cdot152 + (15-20.1)^2\cdot62 + (15-13.6)^2\cdot138 + (20-20.1)^2\cdot99 +$$
$$(20-13.6)^2\cdot101 + (30-20.1)^2\cdot132 + (30-13.6)^2\cdot68\Big) = 64.12$$

$$P_1(25) = \frac{0.374\cdot\exp\left(-\frac{(25-20.1)^2}{2\times64.12}\right)}{0.374\cdot\exp\left(-\frac{(25-20.1)^2}{2\times64.12}\right) + 0.626\exp\left(-\frac{(25-13.6)^2}{2\times64.12}\right)} = 0.5772$$

QDA:

$$\hat{\sigma}_1^2 = \frac{1}{374-1}\times\sum_{y_i=1}(x_i-\hat{u}_1)^2 = \frac{1}{374-1}\cdot\Big\{33\times(5-20.1)^2 + 48\times(10-20.1)^2 + 62\times(15-20.1)^2$$
$$+ 99\times(20-20.1)^2 + 132\times(30-20.1)^2\Big\} = \underline{72.31} \ \#$$

$$\hat{\sigma}_2^2 = \frac{1}{626-1}\cdot\Big\{167\times(5-13.6)^2 + 152\times(10-13.6)^2 + 138\times(15-13.6)^2$$
$$+ 101\times(20-13.6)^2 + 68\times(30-13.6)^2\Big\} = \underline{59.23} \ \#$$

$$P_1(25) = 0.6025$$

2.

Decision rule of QDA

$$\hat{Y}(x) = \arg\max_{k\in\{0,1\}} P_k(x)$$

$$= \arg\max_{k\in\{0,1\}} \frac{\pi_k \frac{1}{\cancel{2\pi}\det(\Sigma_k)^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-u_k)^T\Sigma_k^{-1}(x-u_k)\right)}{\sum_{l=0}^{1}\pi_l\frac{1}{\cancel{2\pi}\det(\Sigma_l)^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-u_l)^T\Sigma_l^{-1}(x-u_l)\right)}$$

$$= \arg\max_{k\in\{0,1\}} \log\pi_k - \frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x-u_k)^T\Sigma_k^{-1}(x-u_k)$$

Let $\delta_k = \log\pi_k - \frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x-u_k)^T\Sigma_k^{-1}(x-u_k)$

$\Rightarrow \hat{Y}(x) = \arg\max_{k\in\{0,1\}} \delta_k(x)$

Class 1:   class 0:

$(1.5, 2)$ $(-2, 0)$

$(1, 1)$ $(-1, 0)$

$(2, 0.5)$ $(-2, -1)$

$\hat{u}_0 = \left(\frac{4.5}{3}, \frac{3.5}{3}\right)$

$\hat{u}_1 = \left(\frac{-5}{3}, \frac{-1}{3}\right)$, $\pi_1 = \pi_2 = \frac{1}{2}$

$$\Sigma_0 = \frac{1}{2}\left(\left[0, \frac{2.5}{3}\right]^T\left[0, \frac{2.5}{3}\right] + \left[\frac{-1.5}{3}, \frac{-0.5}{3}\right]^T\left[\frac{-1.5}{3}, \frac{-0.5}{3}\right] + \left[\frac{1.5}{3}, \frac{-2}{3}\right]^T\left[\frac{1.5}{3}, \frac{-2}{3}\right]\right) = \begin{bmatrix} \frac{1}{4} & \frac{-1}{8} \\ \frac{-1}{8} & \frac{9}{12} \end{bmatrix}$$

$$\Sigma_1 = \frac{1}{2}\left(\left[\frac{-1}{3}, \frac{1}{3}\right]^T\left[\frac{-1}{3}, \frac{1}{3}\right] + \left[\frac{2}{3}, \frac{1}{3}\right]^T\left[\frac{2}{3}, \frac{1}{3}\right] + \left[\frac{-1}{3}, \frac{-2}{3}\right]^T\left[\frac{-1}{3}, \frac{-2}{3}\right]\right) = \begin{bmatrix} \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} \end{bmatrix}$$

0: $\delta_0 = -\frac{1}{2}(-u_0)^T \Sigma_0^{-1}(-u_0) + \log \pi_0 - \frac{1}{2}\log|\Sigma_0| = -4.94$

$\Rightarrow \delta_0 > \delta_1 \Rightarrow \hat{y} = 0$

1: $\delta_1 = -\frac{1}{2}(-u_1)^T \Sigma_1^{-1}(-u_1) + \log \pi_1 - \frac{1}{2}\log|\Sigma_1| = -8.5$

3.

X = profit

$Y = \begin{cases} 1 & (\text{Yes}) \\ 0 & (\text{No}) \end{cases}$ and $P(Y=1) = 0.8$, $P(Y=0) = 0.2$

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

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

$$P(Y=1 \mid X) = \frac{f_{X|Y=1}(x)\,P(Y=1)}{f_{X|Y}(x)} = \frac{0.8 \times N(10,25)}{0.8 \times N(10,25) + 0.2\,N(0,36)}$$

$$= \frac{\cancel{\frac{1}{\sqrt{2\pi}}}\frac{1}{5}\exp\left(\frac{-(x-10)^2}{50}\right) \times 0.8}{\cancel{\frac{1}{\sqrt{2\pi}}}\frac{1}{5}\exp\left(\frac{-(x-10)^2}{50}\right) \times 0.8 + \cancel{\frac{1}{\sqrt{2\pi}}}\frac{1}{6}\exp\left(\frac{-x^2}{72}\right) \times 0.2}$$

$$\Rightarrow P(Y=1 \mid X=4) = \frac{\frac{1}{5}\exp\left(\frac{-(4-10)^2}{50}\right) \times 0.8}{\frac{1}{5}\exp\left(\frac{-(4-10)^2}{50}\right) \times 0.8 + \frac{1}{6}\exp\left(\frac{-4^2}{72}\right) \times 0.2} \doteq 0.7448$$

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

When (x_i, y_i) is left out, the i^{th} fitted value is

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

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

Applying matrix inversion lemma : $(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}$

$$\Rightarrow (\underset{\sim}{X}^T\underset{\sim}{X} - x_i x_i^T)^{-1} = (\underset{\sim}{X}^T\underset{\sim}{X})^{-1} + (\underset{\sim}{X}^T\underset{\sim}{X})^{-1} x_i (1 - \underbrace{x_i^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}x_i}_{h_i})^{-1} x_i^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}$$

(Let $(\underset{\sim}{X}^T\underset{\sim}{X})^{-1} = A$)

$$\hat{y}_i^{(-i)} = x_i^T\left(A + A x_i(1-h_i)^{-1}x_i^T A\right)(\underset{\sim}{X}^T y - x_i y_i)$$

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

$$= x_i^T A \underset{\sim}{X}^T y - x_i^T A x_i y_i + \frac{x_i^T A x_i x_i^T A \underset{\sim}{X}^T y}{1-h_i} - \frac{x_i^T A x_i x_i^T A x_i y_i}{1-h_i}$$

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

The overall MSE is

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

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

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

5.

(a) $$\hat{\beta}_1 = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum_i x_i^2 - n\bar{x}^2} = \frac{21+40+150-3\times 6\times 10}{134 - 3\times 36} = \frac{31}{26} = 1.192$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 10 - 1.192\times 6 = 2.848$$

$$\text{training error} = \frac{1}{3}\sum_{i=1}^{3}(y_i - \hat{y}_i)^2 = \frac{(7-6.424)^2+(8-8.808)^2+(15-14.768)^2}{3}$$

$$= \frac{0.331776+0.652864+0.053824}{3} = 0.3462$$

(b) $$X = \begin{bmatrix} 1 & 3 \\ 1 & 5 \\ 1 & 10 \end{bmatrix},\quad X(X^TX)^{-1}X^T = \begin{bmatrix} 1 & 3 \\ 1 & 5 \\ 1 & 10 \end{bmatrix}\left(\begin{bmatrix} 1 & 1 & 1 \\ 3 & 5 & 10 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 1 & 5 \\ 1 & 10 \end{bmatrix}\right)\begin{bmatrix} 1 & 1 & 1 \\ 3 & 5 & 10 \end{bmatrix}$$

$$= \begin{bmatrix} 0.68 & 0.45 & -0.13 \\ 0.45 & 0.372 & 0.18 \\ -0.13 & 0.18 & 0.95 \end{bmatrix}$$

$$CV = \frac{1}{3}\sum_{i=1}^{3}\left(\frac{y_i - \hat{y}_i}{1-h_i}\right)^2 = \frac{1}{3}\left[\left(\frac{7-6.424}{1-0.68}\right)^2 + \left(\frac{8-8.808}{1-0.372}\right)^2 + \left(\frac{15-14.768}{1-0.95}\right)^2\right] = 8.808$$

(c) $\hat{\beta}_1^{*1} = 0.49925$ $\hat{\beta}_1^{*2} = 1.3983$ $\hat{\beta}_1^{*3} = 1.1428$

$\hat{\beta}_0^{*1} = 5.504$, $\hat{\beta}_0^{*2} = 1.015$, $\hat{\beta}_0^{*3} = 3.5714$

$$\widehat{Err}_{boot} = \frac{1}{3}\times\left[(y_1 - \hat{\beta}_0^{*2} - \hat{\beta}_1^{*2}x_1)^2 + (y_2 - \hat{\beta}_0^{*3} - \hat{\beta}_1^{*3}x_2)^2 + (y_3 - \hat{\beta}_0^{*1} - \hat{\beta}_1^{*1}x_3)^2\right]$$

$$= 8.3794$$

(d) $$\widehat{SE}_B(\hat{\beta}_0)^2 = \frac{1}{2}\sum_{b=1}^{3}\left(\hat{\beta}_0^{*b} - \overbrace{\frac{1}{3}\sum_{b'=1}^{3}\hat{\beta}_0^{*b'}}^{3.3635}\right)^2 = 5.0702$$

$$\widehat{SE}_B(\hat{\beta}_1)^2 = \frac{1}{2}\sum_{b=1}^{3}\Big(\hat{\beta}_1^{*b} - \underbrace{\frac{1}{3}\sum_{b'=1}^{3}\hat{\beta}_1^{*b'}}_{1.0135}\Big)^2 = 0.1486$$

$$\widehat{SE}(\hat{\beta}_0)^2 = \hat{\sigma}^2\left[\frac{1}{3} + \frac{\bar{x}^2}{\sum_{i=1}^{3}(x_i - \bar{x})^2}\right] = 1.0845\left[\frac{1}{3} + \frac{6^2}{26}\right] = 1.863$$

$$\widehat{SE}(\hat{\beta}_1)^2 = \frac{\hat{\sigma}^2}{\sum_{i=1}^{3}(x_i - \bar{x})^2} = \frac{1.0845}{26} = 0.0417$$