Problem1

1. The goal of the project

   This project is aimed at explore the quality rating for many well-known US hospitals. Since health care is an important issue for current society, an increasing number of US hospitals have emerged among different states, which support service in a diverse range of professional fields. This project hopes to help people decide which hospital is the best choice depending on many factors, such as charges, distance, patient reviews, doctor qualifications, etc. Through weighting these factors in an analysis model for different kinds of patients, quality for hospitals based on their needs is possible to be ranked.

2. Data collection

   The dataset is offered by Centers for Medicare & Medicaid Services (CMS). It contains information of 90 hospitals, including locations, hospital type, mortality records, safety of care national comparison, patient experience national comparison, etc. More information can be retrieved from its website.

3. Current/ Future tools

   PCA, LDA, linear/non-linear regression, decision tree

   (Note: This is only a preliminary proposal plan, it may be changed or modifies if I find something more interesting afterwards)

Problem 2

P2.

① MAP Decision. Rule.

Two class $c \in \{0, 1\}$. (y_i, x_i) are iid. drawn from.

$P_X(x|y=c) \sim N(\bar{x}^c, \underset{\substack{\uparrow \\ \text{same for both classes}}}{\Sigma})$ $\quad P_{XY}(x, y=c) = P_X(x|y=c)\, P(y=c)$

$$P(y=c|x) = \frac{P_{X,Y}(x, y=c)}{P_X(x)} = \frac{P_X(x|y=c)\, P(y=c)}{P_X(x)}.$$

MAP decision rule : $\hat{c}(x) = \begin{cases} 1 & P(y=1|x) \geq P(y=0|x) \\ 0 & \text{o/w}. \end{cases}$

$$P(y=1|x) \underset{<}{\geq} P(y=0|x) \quad \begin{matrix} (\hat{c}(x) = 1) \\ (\hat{c}(x) = 0) \end{matrix}$$

$$\frac{P(x|y=1)\, P(y=1)}{P_X\cancel{(x)}} \underset{\hat{c}_x=0}{\overset{\hat{c}_x=1}{\underset{<}{\geq}}} \frac{P(x|y=0)\, P(y=0)}{\cancel{P_X(x)}}$$

$P(y=1) = \pi_1 \quad P(y=0) = \pi_0.$

$$P(x|y=1) \sim N(\bar{x}^1, \Sigma) \Rightarrow \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-\tfrac{1}{2}(x-\bar{x}^1)^T \Sigma^{-1}(x-\bar{x}^1))$$

$$P(x|y=0) \sim N(\bar{x}^0, \Sigma) \Longrightarrow \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-\tfrac{1}{2}(x-\bar{x}^0)^T \Sigma^{-1}(x-\bar{x}^0)$$

$$\Rightarrow \cancel{\frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}}} \pi_1 \exp(-\tfrac{1}{2}(x-\bar{x}^1)^T \Sigma^{-1}(x-\bar{x}^1)) \underset{C(x)=0}{\overset{C(x)=1}{\underset{<}{\geq}}} \cancel{\frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}}} \pi_0 \exp(-\tfrac{1}{2}(x-\bar{x}^1)^T \Sigma^{-1}(x-\bar{x}^1))$$

$$\frac{\exp(-\tfrac{1}{2}(x-\bar{x}^1)^T \Sigma^{-1}(x-\bar{x}^1))\ \pi_1}{\exp(-\tfrac{1}{2}(x-\bar{x}^0)^T \Sigma^{-1}(x-\bar{x}^0))\ \pi_0} \underset{C(x)=0}{\overset{C(x)=1}{\underset{<}{\geq}}} 1$$

$$\log \Rightarrow \underset{①}{-}\tfrac{1}{2}(x-\bar{x}^1)^T \Sigma^{-1}(x-\bar{x}^1) + \tfrac{1}{2}(x-\bar{x}^0)^T \Sigma^{-1}(x-\bar{x}^0) + \ln\frac{\pi_1}{\pi_0} \underset{C(x)=0}{\overset{C(x)=1}{\underset{<}{\geq}}} 0$$

①

$$\underline{-\frac{1}{2}\cancel{x^T}\Sigma^{-1}x} + \underline{\frac{1}{2}\bar{x}^{1T}\Sigma^{-1}x} + \underline{\frac{1}{2}x^T\Sigma^{-1}\bar{x}^1} - \frac{1}{2}\bar{x}^{1T}\Sigma^{-1}\bar{x}^1 +$$

$$\underline{\frac{1}{2}x^T\cancel{\Sigma^{-1}}x} - \underline{\frac{1}{2}\bar{x}^{0T}\Sigma^{-1}x} - \underline{\frac{1}{2}x^T\Sigma^{-1}\bar{x}^0} + \frac{1}{2}\bar{x}^{0T}\Sigma^{-1}\bar{x}^0 + \ln\frac{\lambda_1}{\lambda_0} \underset{C(x)=0}{\overset{C(x)=1}{\gtrless}} 0$$

$$\frac{1}{2}\underline{x^T\Sigma^{-1}(\bar{x}^1-\bar{x}^0) + \frac{1}{2}(\bar{x}^{1T}-\bar{x}^{0T})}\Sigma^{-1}x + \frac{1}{2}\bar{x}^{0T}\Sigma^{-1}\bar{x}^0$$

$$-\frac{1}{2}\bar{x}^{1T}\Sigma^{-1}\bar{x}^1 + \ln\frac{\lambda_1}{\lambda_0} \underset{C(x)=0}{\overset{C(x)=1}{\gtrless}} 0$$

$$\circledast \quad x^T\underbrace{\Sigma^{-1}(\bar{x}^1-\bar{x}^0)}_{\theta} + \underbrace{\frac{1}{2}\bar{x}^{0T}\Sigma^{-1}\bar{x}_0 - \frac{1}{2}\bar{x}^{1T}\Sigma^{-1}\bar{x}^1 + \log\frac{\lambda_1}{\lambda_0}}_{\theta_0} \underset{C(x)=0}{\overset{C(x)=1}{\gtrless}} 0$$

$$\Rightarrow \quad \theta^T x + \theta_0 \underset{C(x)=0}{\overset{C(x)=1}{\gtrless}} 0 \qquad \theta = \Sigma^{-1}(\bar{x}^1-\bar{x}^0)$$

$$\theta_0 = \frac{1}{2}\bar{x}^{0T}\Sigma^{-1}\bar{x}^0 - \frac{1}{2}\bar{x}^{1T}\Sigma^{-1}\bar{x}^1 + \log\frac{\lambda_1}{\lambda_0}$$

---

② Minimum-distance decision rule.
(Euclidean)

Assign x to class C_i : $i = \operatorname{argmin}(x-\bar{x}^j)^T(x-\bar{x}^j)$ $\quad j = 1, 2$

Normal $\quad C_0 \sim N(\bar{x}_0, \Sigma) \qquad C_1 \sim (\bar{x}_1, \Sigma) \qquad \Sigma = 6^2 I$

if $(x-\bar{x}^0)^T(x-x^0) > (x-\bar{x}^1)^T(x-\bar{x}^1)$

assign x to class 1

O/W assign x to class 0.

Compare it with Bayesian rule,

as show in ① $g = -\frac{1}{2}(x-\bar{x}^1)^T\Sigma^{-1}(x-\bar{x}^1) + \frac{1}{2}(x-\bar{x}^0)^T\Sigma^{-1}(x-\bar{x}^0) + \cancel{\ln\frac{\lambda_1}{\lambda_0}}$

λ_1, λ_0 are the same in this case.

$$(X-\bar{X}^1)^T \Sigma^{-1} (X-\bar{X}^1) \underset{\text{class } 1}{\overset{\text{class } 0}{\gtrless}} (X-\bar{X}^0)^T \Sigma^{-1} (X-\bar{X}^0) .$$

this is called Minimum Mahalanobis distance classifier.

Unlike Euclidean decision rule. $(\Sigma = \sigma^2 I)$, Mahalonobis distance decision rule is equvalent to bayesian decision rule, which is more general as a weighted form of Euclidean decision rule.

Problem 3

1,2,3. See codes

4. For the first case that two class has the same covariance, the classification error using Bayesian classification rule is 24, while the error for logistic regression is 32. They look similar as the data can be easily separated by a linear line. Here, Bayesian rule can be simplified as a linear classification and logistic regression is a generalized linear model, which have similar performance for this scenario.

5. After changing the covariance matrix, the points from two classes are overlapped in the centre part. In this case, Bayesian rule works no longer as a linear model, instead, considering the quadratic term, it is equivalent to Mahalanobis distance rule, which can classify the overlapping points based on the distance between them and mean for each class. On the other hand, logistic regression still works as a linear model and cannot distinguish the overlapping parts very well. Therefore, as the test errors shows, error for Bayesian rule is 275, while error for logistic regression is 489. This result validates our analysis.