Homework 6 Solution

Question 1

(a)

$$S = \begin{bmatrix} \frac{13}{3} & 8 \\ 8 & \frac{52}{3} \end{bmatrix}$$

The spectral decomposition of S is

$$S = 21.1410974 \begin{bmatrix} 0.4297717 \\ 0.9029376 \end{bmatrix} \begin{bmatrix} 0.4297717 \\ 0.9029376 \end{bmatrix}^\top + 0.5255693 \begin{bmatrix} -0.9029376 \\ 0.4297717 \end{bmatrix} \begin{bmatrix} -0.9029376 \\ 0.4297717 \end{bmatrix}^\top$$

Based on $\vec{u}_1 = \begin{bmatrix} 0.4297717 \\ 0.9029376 \end{bmatrix}$ and $\vec{u}_2 = \begin{bmatrix} -0.9029376 \\ 0.4297717 \end{bmatrix}$, the principal components are

$$\hat{y}_1 = 0.4297717x_1 + 0.9029376x_2$$
$$\hat{y}_2 = -0.9029376x_1 + 0.4297717x_2$$

(b) The proportion of total sample variance due to the first sample principal component is

$$\frac{\hat{\lambda}_1}{\hat{\lambda}_1 + \hat{\lambda}_2} = \frac{21.1410974}{21.1410974 + 0.5255693} = 97.57\%$$

(c) Based on loadings, as $|v_{11}| = 0.4297717 < |v_{12}| = 0.9029376$, the second variate contributes more to the determination of the first sample principal component.

(d) Based on sample correlations,

$$v_{11}\sqrt{\frac{\hat{\lambda}_1}{s_{11}}} = 0.4297717\sqrt{\frac{21.1410974}{\frac{13}{3}}} = 0.949$$
$$v_{12}\sqrt{\frac{\hat{\lambda}_1}{s_{22}}} = 0.9029376\sqrt{\frac{21.1410974}{\frac{52}{3}}} = 0.997$$

the second variate contributes a little bit more to the determination of the first sample principal component.

(e) The sample covariance matrix of the standardized dataset is equivalent to the sample correlation matrix of the original dataset.

Let

$$D = \begin{bmatrix} \frac{13}{3} & 0 \\ 0 & \frac{52}{3} \end{bmatrix}$$

Then the sample correlation matrix is

$$D^{-\frac{1}{2}} S D^{-\frac{1}{2}} = \begin{bmatrix} 10.92307690.92307691 \end{bmatrix}$$

The spectral decomposition is

$$S = 1.92307692 \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}^\top + 0.07692308 \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}^\top$$

The principal components are

$$\begin{aligned} \hat{y}_1 &= \frac{\sqrt{2}}{2} z_1 + \frac{\sqrt{2}}{2} z_2 \\ \hat{y}_2 &= -\frac{\sqrt{2}}{2} z_1 + \frac{\sqrt{2}}{2} z_2 \end{aligned}$$

The proportion of total sample variance due to the first sample principal component is

$$\begin{aligned} \frac{\hat{\lambda}_1}{\hat{\lambda}_1 + \hat{\lambda}_2} &= \frac{1.92307692}{1.92307692 + 0.07692308} \\ &= 96.15\% \end{aligned}$$

Based on loadings, these two variate contribute the same to the determination of the first sample principal component.

Based on the sample correlations

$$u_{11}\sqrt{\hat{\lambda}_1} = 0.9806 u_{12}\sqrt{\hat{\lambda}_1} = 0.9806$$

these two variate contribute the same to the determination of the first sample principal component.

Remark: For the standardized dataset, the comparisons of their contributions to the determination of $\hat{y}_i$ based on loadings and correlation coeffcients are the same.

Question 2 First let's calculate the second principal component.
By

$$\begin{aligned} \vec{v}_2^\top \vec{v}_1 &= 0 \\ \|\vec{v}_2\| &= 1 \end{aligned}$$

we have

$$\vec{v}_2 = \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$$

So the second principal component is

$$Y_2 = -\frac{\sqrt{2}}{2}X_1 + \frac{\sqrt{2}}{2}X_2$$

Then,

$$\begin{aligned} Var(Y_1) &= Var(\frac{\sqrt{2}}{2}X_1 + \frac{\sqrt{2}}{2}X_2) \\ &= (\frac{\sqrt{2}}{2})^2 Var(X_1) + (\frac{\sqrt{2}}{2})^2 Var(X_2) + 2\frac{\sqrt{2}}{2}\frac{\sqrt{2}}{2}Cov(X_1, X_2) \\ &= \frac{1}{2}Var(X_1) + \frac{1}{2}Var(X_2) + Cov(X_1, X_2) \end{aligned}$$

Similarly,

$$\begin{aligned} Var(Y_2) &= Var(-\frac{\sqrt{2}}{2}X_1 + \frac{\sqrt{2}}{2}X_2) \\ &= (-\frac{\sqrt{2}}{2})^2 Var(X_1) + (\frac{\sqrt{2}}{2})^2 Var(X_2) + 2(-\frac{\sqrt{2}}{2})(\frac{\sqrt{2}}{2})Cov(X_1, X_2) \\ &= \frac{1}{2}Var(X_1) + \frac{1}{2}Var(X_2) - Cov(X_1, X_2) \end{aligned}$$

As $Corr(X_1, X_2) < 0$, we have $Cov(X_1, X_2) < 0$. Then

$$Var(Y_1) < Var(Y_2)$$

which violates the property of principal components. So it is not possible that $Corr(X_1, X_2) < 0$.

Question 3 The results based on S and R are quite different.
For the PCA based on S, the proportion of the total variance due to the first 3 principal components is 98.70%. And based on loadings, the first principal component is mainly determined by radiation. The second principal component is mainly determined by O3. And the third principal component is mainly determined by NO2. It is mainly because these three variates have much larger sample variances comparing with other variates.
For the PCA based on R, the proportion of the total variance due to the first 3 principal components is only 70.38%. Based on loadings, the first principal component mainly contains information about all the pollutants - CO,NO,NO2,O3 and HC. The second principal component is mainly determined by the contrast between the solar radiation, O3 and NO, NO2, HC. The third principal component is mainly determined by the contrast between wind, HC and NO.

Question 4 As $n_1 = n_2$, the pooled sample covariance matrix is

$$S_{pooled} = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$$

1. Suppose only x_{01} is observable. We will only use the first variate to construct the rule.

$$
\begin{aligned}
w &= s_{11}^{-1}(\bar{x}_{11} - \bar{x}_{21}) \\
&= \frac{1}{4}(6-0) \\
&= 1.5
\end{aligned}
$$

Then the Fisher's rule is: assign x_0 to class 1 if

$$
\begin{aligned}
wx_{01} &\geq \frac{1}{2}w(\bar{x}_{11} + \bar{x}_{21}) \\
1.5x_{01} &\geq \frac{1}{2}1.5(6+0) \\
x_{01} &\geq 3
\end{aligned}
$$

2. Similarly, suppose only x_{02} is observable. We will only use the second variate to construct the rule.

$$
\begin{aligned}
w &= s_{22}^{-1}(\bar{x}_{12} - \bar{x}_{22}) \\
&= \frac{1}{4}(0-6) \\
&= -1.5
\end{aligned}
$$

Then the Fisher's rule is: assign x_0 to class 1 if

$$
\begin{aligned}
wx_{02} &\geq \frac{1}{2}w(\bar{x}_{12} + \bar{x}_{22}) \\
-1.5x_{02} &\geq \frac{1}{2}(-1.5)(0+6) \\
x_{02} &\leq 3
\end{aligned}
$$

3.

$$
\begin{aligned}
\vec{w} &= S_{pooled}^{-1}(\vec{\bar{x}}_1 - \vec{\bar{x}}_2) \\
&= \begin{bmatrix} 3 \\ -3 \end{bmatrix}
\end{aligned}
$$

Then the Fisher's rule is: assign x_0 to class 1 if

$$
\begin{aligned}
\vec{w}^\top \vec{x}_0 &\geq \frac{1}{2}\vec{w}^\top(\vec{\bar{x}}_1 + \vec{\bar{x}}_2) \\
3x_{01} - 3x_{02} &\geq \frac{1}{2}\begin{bmatrix} 3 \\ -3 \end{bmatrix}^\top \begin{bmatrix} 6 \\ 6 \end{bmatrix} \\
x_{01} - x_{02} &\geq 0
\end{aligned}
$$

Then suppose both Classifier 1 and Classifier 2 classify x_0 to class 1, then we have

$$
\begin{aligned}
x_{01} &\geq 3 \\
x_{02} &\leq 3
\end{aligned}
$$

Then $x_{01} - x_{02} \geq 0$, which means Classifier 3 will also classify x_0 to class 1. Similarly, when both Classifier 1 and Classifier 2 classify x_0 to class 2, we can check that Classifier 3 will also classify x_0 to class 2.
Therefore, the x_0 does not exist.

Question 5

$$S_{pooled} = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$$

1. Suppose only x_{01} is observable. We will only use the first variate to construct the rule.

$$\begin{aligned} w &= s_{11}^{-1}(\bar{x}_{11} - \bar{x}_{21}) \\ &= \frac{1}{4}(6 - 0) \\ &= 1.5 \end{aligned}$$

Then the Fisher's rule is: assign x_0 to class 1 if

$$\begin{aligned} wx_{01} &\geq \frac{1}{2}w(\bar{x}_{11} + \bar{x}_{21}) \\ 1.5x_{01} &\geq \frac{1}{2}1.5(6 + 0) \\ x_{01} &\geq 3 \end{aligned}$$

2. Similarly, suppose only x_{02} is observable. We will only use the second variate to construct the rule.

$$\begin{aligned} w &= s_{22}^{-1}(\bar{x}_{12} - \bar{x}_{22}) \\ &= \frac{1}{4}(6 - 0) \\ &= 1.5 \end{aligned}$$

Then the Fisher's rule is: assign x_0 to class 1 if

$$\begin{aligned} wx_{02} &\geq \frac{1}{2}w(\bar{x}_{12} + \bar{x}_{22}) \\ 1.5x_{02} &\geq \frac{1}{2}1.5(6 + 0) \\ x_{02} &\geq 3 \end{aligned}$$

3.

$$\begin{aligned} \vec{w} &= S_{pooled}^{-1}(\vec{\bar{x}}_1 - \vec{\bar{x}}_2) \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{aligned}$$

Then the Fisher's rule is: assign x_0 to class 1 if

$$\begin{aligned} \vec{w}^\top \vec{x}_0 &\geq \frac{1}{2}\vec{w}^\top(\vec{\bar{x}}_1 + \vec{\bar{x}}_2) \\ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \vec{x}_0 &\geq \frac{1}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}^\top \begin{bmatrix} 6 \\ 6 \end{bmatrix} \\ x_{01} + x_{02} &\geq 6 \end{aligned}$$

Then suppose both Classifier 1 and Classifier 2 classify x_0 to class 1, then we have

$$x_{01} \geq 3$$
$$x_{02} \geq 3$$

Then $x_{01} + x_{02} \geq 6$, which means Classifier 3 will also classify x_0 to class 1. Similarly, when both Classifier 1 and Classifier 2 classify x_0 to class 2, we can check that Classifier 3 will also classify x_0 to class 2.
Therefore, the x_0 does not exist.

Question 6 When we compare π_1 and π_2, as the Fisher's rule allocates $\vec{x}_0$ to π_2, we have

$$(\vec{x}_0 - \bar{\vec{x}}_1)^\top S_{pooled}^{-1}(\vec{x}_0 - \bar{\vec{x}}_1) > (\vec{x}_0 - \bar{\vec{x}}_2)^\top S_{pooled}^{-1}(\vec{x}_0 - \bar{\vec{x}}_2)$$

which means for $\vec{x}_0$, the Mahalanobis distance to $\bar{\vec{x}}_1$ is greater than that to $\bar{\vec{x}}_2$.
When we compare π_2 and π_3, similarly, we have

$$(\vec{x}_0 - \bar{\vec{x}}_2)^\top S_{pooled}^{-1}(\vec{x}_0 - \bar{\vec{x}}_2) > (\vec{x}_0 - \bar{\vec{x}}_3)^\top S_{pooled}^{-1}(\vec{x}_0 - \bar{\vec{x}}_3)$$

which means for $\vec{x}_0$, the Mahalanobis distance to $\bar{\vec{x}}_2$ is greater than that to $\bar{\vec{x}}_3$.
Therefore, when we compare π_1 and π_3, as

$$(\vec{x}_0 - \bar{\vec{x}}_1)^\top S_{pooled}^{-1}(\vec{x}_0 - \bar{\vec{x}}_1) > (\vec{x}_0 - \bar{\vec{x}}_2)^\top S_{pooled}^{-1}(\vec{x}_0 - \bar{\vec{x}}_2) > (\vec{x}_0 - \bar{\vec{x}}_3)^\top S_{pooled}^{-1}(\vec{x}_0 - \bar{\vec{x}}_3)$$

we know that $\vec{x}_0$ is allocated to π_3

Question 7 The Fisher's rule is: assign $\vec{x}_0$ to Non-Multiple-Sclerosis group if

$$\begin{bmatrix} 0.02340633 \\ -0.03446657 \\ 0.21027081 \\ -0.08393327 \\ -0.25345069 \end{bmatrix}^\top \vec{x}_0 \geq -23.23392$$

The confusion matrix is

Actual / Predict	Non-Multiple-Sclerosis	Multiple-Sclerosis
Non-Multiple-Sclerosis	66	3
Multiple-Sclerosis	7	22

The apparent error rate is the proportion of observed cases incorrectly predicted, which is

$$\frac{3+7}{n_1+n_2} = \frac{10}{98} = 0.102$$

The expected actual error rate by Lachenbruch's holdout is an estimate based on leave-one-out method. At each step, all but one observation is used to complete the classification rule, and this rule is then used to classify the omitted observation. We repeat this procedure for each observation, and the expected actual error rate is given by

$$\frac{\text{number of misclassified observations}}{n_1+n_2} = \frac{13}{98} = 0.133$$