MH4501
Assignment 4
April 29, 2023

Name: Singh Jasraj
Matric: U1940558D
Collaborators: N.A.

Problem 1

a) The plot of the densities can be seen in Figure 1.

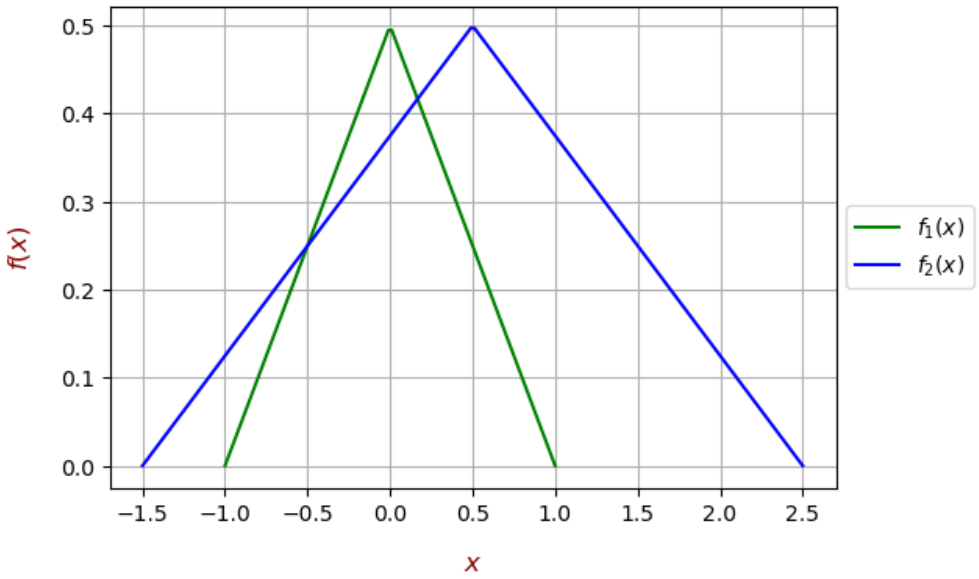

Figure 1: Plots of the densities $f_1(x)$ and $f_2(x)$.

b)

$$\begin{aligned}
\mathcal{R}_1 &= \{x \in [-3/2, 5/2] : p_1 f_1(x)\, c(2|1) \geq p_2 f_2(x)\, c(1|2)\} \\
&= \{x \in [-3/2, 5/2] : f_1(x) \geq f_2(x)\} \\
&= [-1/2, 1/6] \\
\mathcal{R}_2 &= \{[-3/2, 5/2] : p_1 f_1(x)\, c(2|1) < p_2 f_2(x)\, c(1|2)\} \\
&= [-3/2, -1/2) \cup (1/6, 5/2]
\end{aligned}$$

Problem 2

a)

$$\begin{aligned} S_{\text{pool}} &= \frac{(n_1 - 1) S_1 + (n_2 - 1) S_2}{n_1 + n_2 - 2} \\ &= \frac{16S_1 + 20S_2}{36} \\ &= \begin{pmatrix} 50/9 & 0 & 2 \\ 0 & 41/9 & 0 \\ 2 & 0 & 40/9 \end{pmatrix} \\ d_{12}(x_0) &= (\bar{x}_1 - \bar{x}_2)^T S_{\text{pool}}^{-1} x_0 - \frac{1}{2} (\bar{x}_1 - \bar{x}_2)^T S_{\text{pool}}^{-1} (\bar{x}_1 + \bar{x}_2) \\ &= \begin{pmatrix} 0.527326 & -0.175609 & -0.214797 \end{pmatrix} x_0 - 2.421522 \\ k &= \ln\left(\frac{p_2}{p_1} \frac{c(1|2)}{c(2|1)}\right) \\ &= \ln(4) \approx 1.386294 \end{aligned}$$

Assign x_0 to π_1 if $d_{12}(x_0) > k$, i.e., $\begin{pmatrix} 0.527326 & -0.175609 & -0.214797 \end{pmatrix} x_0 - 2.421522 > \ln(4)$. Else, assign to π_2.

b) For $x_0 = \begin{pmatrix} 11 & 8 & 10 \end{pmatrix}^T$, $d_{12}(x_0) = -0.173774 < \ln(4) = k$. Hence, we asign to π_2.

Problem 3

Let $\bar{x}_{[k]}^{(i)}$ denote the mean of the k^{th} cluster after the i^{th} iteration.

$$\bar{x}_{[1]}^{(0)} = \begin{pmatrix} 1.8 \\ 2.3 \end{pmatrix},$$
$$\bar{x}_{[2]}^{(0)} = \begin{pmatrix} 4.125 \\ 5.375 \end{pmatrix}$$

a) Cluster 1 consists of points #1 and #2, and Cluster 2 consists of points #3, #4, #5, #6 and #7.

$$\bar{x}_{[1]}^{(1)} = \begin{pmatrix} 1.2 \\ 1.45 \end{pmatrix},$$
$$\bar{x}_{[2]}^{(1)} = \begin{pmatrix} 3.9 \\ 5.1 \end{pmatrix}$$

b) Cluster 1 consists of points #1 and #2, and Cluster 2 consists of points #3, #4, #5, #6 and #7.

$$\bar{x}_{[1]}^{(1)} = \begin{pmatrix} 1.2 \\ 1.45 \end{pmatrix},$$
$$\bar{x}_{[2]}^{(1)} = \begin{pmatrix} 3.9 \\ 5.1 \end{pmatrix}$$

c) Since the algorithm converged after the first 2 steps (cluster centroids and cluster assignment stopped changing), we don't need to perform any more iterations.

Problem 4

a) The dendogram using complete linkage is shown in Figure 2.

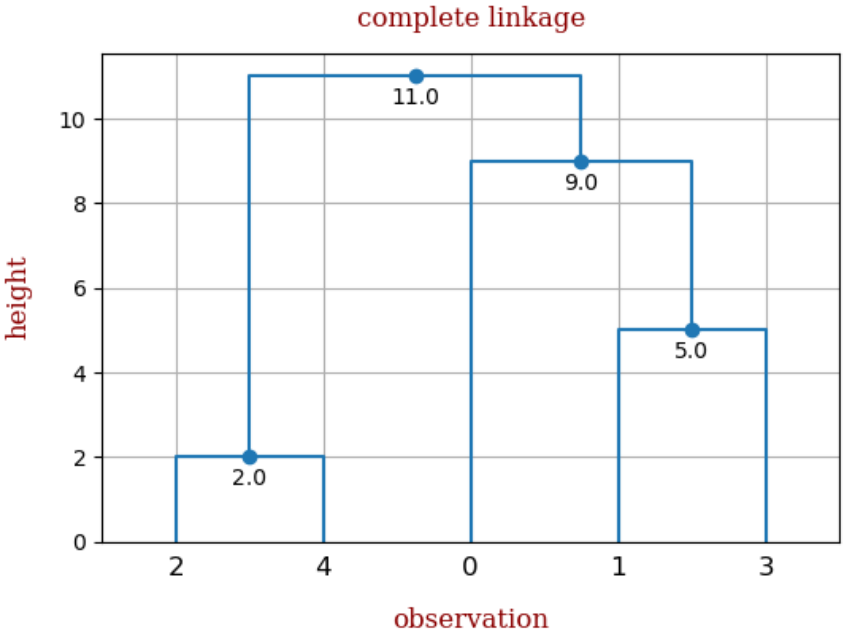

Figure 2: Dendogram using complete linkage hierarchical clustering.

b) The dendogram using single linkage is shown in Figure 3.

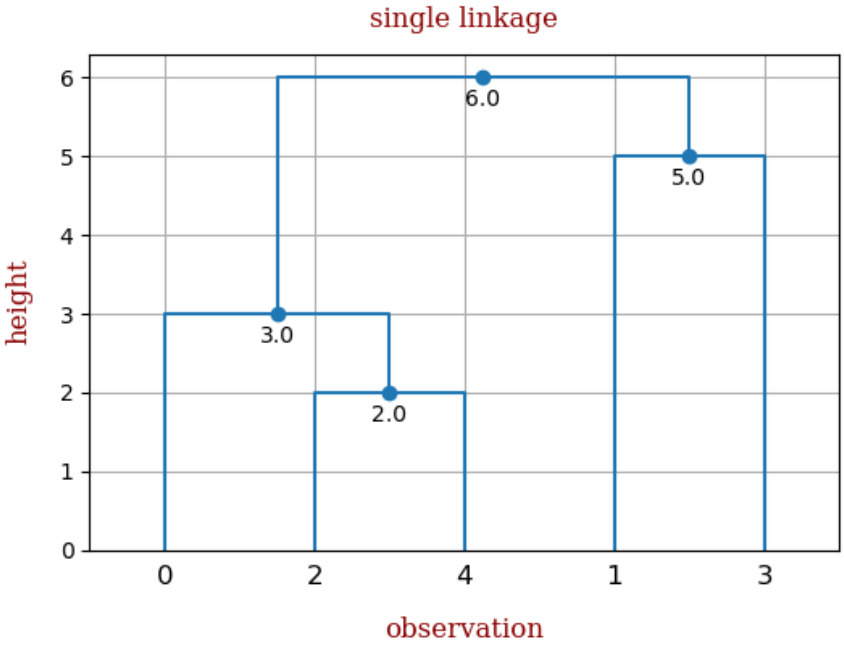

Figure 3: Dendogram using single linkage hierarchical clustering.

Problem 5

$$
\begin{aligned}
\Sigma_{11}^{-1/2} &= \begin{pmatrix} 0.366666 & -0.066666 \\ -0.066666 & 0.466666 \end{pmatrix} \\
\Sigma_{22}^{-1/2} &= \begin{pmatrix} 0.424328 & 0.064465 \\ 0.064465 & 0.392095 \end{pmatrix} \\
\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2} &= \begin{pmatrix} 0.275584 & -0.032163 \\ -0.032163 & 0.269005 \end{pmatrix} \\
\operatorname{diag}(\rho)\,\rho &= \begin{pmatrix} 0.304626 \\ 0.239963 \end{pmatrix} \\
e &= \begin{pmatrix} 0.742206 & 0.670170 \\ -0.670170 & 0.742206 \end{pmatrix} \\
a &= \Sigma_{11}^{-1/2} e \\
&= \begin{pmatrix} 0.316820 & 0.196248 \\ -0.362226 & 0.301685 \end{pmatrix} \\
\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2} &= \begin{pmatrix} 0.294609 & -0.023396 \\ -0.023396 & 0.249981 \end{pmatrix} \\
f &= \begin{pmatrix} 0.919285 & 0.393591 \\ -0.393591 & 0.919285 \end{pmatrix} \\
b &= \Sigma_{22}^{-1/2} f \\
&= \begin{pmatrix} 0.364705 & 0.226274 \\ -0.095062 & 0.385820 \end{pmatrix}
\end{aligned}
$$

a) $\rho_1 = \sqrt{0.304626} = 0.551930$ and $\rho_2 = \sqrt{0.239963} = 0.489861$.

b) $(U_1, V_1) = \left(\begin{pmatrix} 0.316820 & -0.362226 \end{pmatrix} X^{(1)}, \begin{pmatrix} 0.364705 & -0.095062 \end{pmatrix} X^{(2)}\right)$
$(U_2, V_2) = \left(\begin{pmatrix} 0.196248 & 0.301685 \end{pmatrix} X^{(1)}, \begin{pmatrix} 0.226274 & 0.385820 \end{pmatrix} X^{(2)}\right)$

c)

$$
\begin{aligned}
TS &= -\left(n - 1 - \frac{p+q+1}{2}\right) \sum_{i=1}^{p} \log\left(1 - \rho_i^2\right) \\
&= -\left(100 - 1 - \frac{2+2+1}{2}\right) \left(\log\left(1 - 0.304626\right) + \log\left(1 - 0.239963\right)\right) \\
&= 61.537652 \\
&> 9.487729 \\
&= \chi^2_{0.95}\,[4] \\
&= \chi^2_{\alpha}\,[pq]
\end{aligned}
$$

Hence, we reject H_0.