Exercise Sheet 2

Let $Y \in \{\text{Bus}, \text{Train}, \text{Car}\}$ be choices of mode of transportation chosen by an individual. Let $X \in \{\text{Time}, \text{Cost}\}$ be predictors for their choice.

Assume that

$$\mathbf{X}_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}) \quad \text{where} \quad \boldsymbol{\mu}_B = \begin{bmatrix} 30 \\ 2 \end{bmatrix}, \quad \boldsymbol{\mu}_T = \begin{bmatrix} 20 \\ 5 \end{bmatrix}, \quad \boldsymbol{\mu}_C = \begin{bmatrix} 10 \\ 7 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 100 & 15 \\ 15 & 9 \end{bmatrix}.$$

1. Calculate the correlation coefficient of $\mathbf{X}_k$.
2. You only have information of one individual which can be described by the following variables $\mathbf{X}_{i,B} = \begin{bmatrix} 20 \\ 3 \end{bmatrix}$, $\mathbf{X}_{i,T} = \begin{bmatrix} 25 \\ 7 \end{bmatrix}$, $\mathbf{X}_{i,C} = \begin{bmatrix} 20 \\ 10 \end{bmatrix}$. Calculate the individual's likelihood $f_k(x)$. Hint: you can use R and the package `mvtnorm`.
3. Ignoring the marginal probability for $\mathbf{X}$, calculate the posterior probability of this individual for all choices. Assume that you have have a sample with $(n_{\text{Bus}}, n_{\text{Train}}, n_{\text{Car}}) = (20, 40, 40)$.
4. Derive the risk function.
5. Which decision minimises the risk?
6. Find the Bayes Decision Boundaries.
7. Do you think the assumptions are appropriate in this case. Argue why or why not.