MATH5855: Multivariate Analysis

Assignment 1

Due data: **5 pm on Tuesday October 4, 2022**

Instructions:

- The assignment 1 contains 3 questions and worth a total of 70 points which will count towards 15% of the final mark for the subject.
- Use tables, graphs and concise text explanations to support your answers. Unclear answers may not be marked at your own cost. All tables and graphs must be clearly commented and identified.
- You need to submit **TWO** files, the pdf file of the answers and the R markdown file, containing the R codes.

Questions

Question 1. [20 Marks] Let $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N_3\left(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 11 & -6 & 2 \\ -6 & 10 & -4 \\ 2 & -4 & 6 \end{pmatrix} \right)$.

(a) Find the best linear approximation of X_3 by a linear function of X_1 and X_2. [5 Marks]

(b) Using R, simulate $n = 1000$ samples from $\boldsymbol{X}$. Transform the data to $\boldsymbol{Z} = \begin{pmatrix} Z_1 & Z_2 & Z_3 \end{pmatrix}$, where $Z_1 = X_2 - X_3$, $Z_2 = X_2 + X_1$ and $Z_3 = Z_1 + Z_2$. Plot the scatter-plots of pairs of observations and do the test to confirm the multivariate normality of $\boldsymbol{Z}$. To generate the data, set the seed equal to the last 3 digits of your student zID; i.e., if your student zID is 1234567, you need to use ”‘set.seed(567)”’. Interpret the result of the test. [15 Marks]

Question 2. [25 Marks]

(a) State, explicitly, all possible values that a and b can take in order for the following matrix to be a covariance matrix. Give arguments that justify your answer [5 Marks]

$$\Sigma = \begin{pmatrix} 1 & 2 \\ a & b \end{pmatrix}.$$

(b) Withou using using R, compute the eigenvalues and the eigenvectors of the matrix [10 Marks]

$$\Sigma = \begin{pmatrix} 13 & -4 \\ -4 & 7 \end{pmatrix}$$

(c) Using R, confirm the eigenvalues and the eigenvectors of Σ obtained in (b) and define the matrix P and the diagonal matrix Λ such that $\Sigma = P\Lambda P^T$. [10 Marks]

Question 3. [25 Marks] Consider $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$. Let $\boldsymbol{a} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

(a) Show that $\boldsymbol{a}^T\boldsymbol{X}$ and $\boldsymbol{b}^T\boldsymbol{X}$ are independent. [5 Marks]

(b) Using R, simulate $n = 1000$ samples from $\boldsymbol{X}$ and use the appropriate plot to confirm independence of $\boldsymbol{a}^T\boldsymbol{X}$ and $\boldsymbol{b}^T\boldsymbol{X}$, visually. Make sure to set the seed equal to the last 3 digits of your student ID, as in Question 1. [10 Marks]

(c) Assume that the mean and covariance matrix of the simulated data is unknown. Find the ML estimators for $\boldsymbol{\mu}$ and Σ based on the sample. [10 Marks]