Practice Questions on Multi-Variate Analysis

Mentioned the topics covered in MVA slide along with the section you can refer in the book: Applied Multivariate Statistics by Johnson and Wichern, 6th Ed.

1. Section 1.3: Descriptive Statistics and Section 1.5: Distance
2. Exercises for 1.: 1.1, 1.2, 1.3, 1.4, 1.5, 1.9, 1.10, 1.13, 1.27(try inputting these points in computer to plot scatter plot)
3. Section 2.6: Mean Vectors and Covariance Matrices; Example 2.13, 2.14
4. Descriptive statistics in matrix notation: Section 3.5, Example 3.1
5. Question: Given the following data matrix $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} 12 & 17 & 29 \\ 18 & 20 & 38 \\ 14 & 16 & 30 \\ 20 & 18 & 38 \\ 16 & 19 & 35 \end{bmatrix}$$

   (a) Compute the **sample mean** vector $\bar{\mathbf{x}}$. What are the dimensions of $\bar{\mathbf{x}}$ in terms of samples and features?

   (b) Compute the **sample covariance** matrix $\mathbf{S}$ of the data matrix.

   (c) Compute the Mahalanobis distance between the first two samples in the dataset, using the covariance matrix.

   (d) Try to create a scatter plot for dataset X. Observe any visible patterns between the features based on the scatter plot. You can develop a small code that will plot X using *matplotlib* and *seaborn*. Is there any visible relationship between features?

   (e) Standardize the dataset : standardize X by subtracting the mean and dividing by the standard deviation for each feature. Then compute the sample covariance matrix of the standardized dataset. Discuss the difference between the covariance matrices of the original and standardized datasets. [Optional: you could use the numpy to calculate covariance with 'np.cov']

6. Multivariate Normal distribution; Section 4.1, 4.2
7. Exercises for 6.: 4.1, 4.2, 4.3