BT3041 – Analysis and Interpretation of Biological Data

Assignment 3

You are given a dataset of wine with 9 features. Using these features, you are supposed to predict the class of the wine.

1. Find the optimal number of clusters for the given dataset. (15 marks)

2. Predict the class of the wine using k-means clustering. (10 marks)

3. Compute different metrics such as ARI, and AMI for the predicted and actual classes. (15 marks).

4. Should scaling of data be done prior to k-means clustering? Explain based on your observations from this dataset. (10 marks)

Instructions: This is a hybrid assignment with both coding and Interpretation components. Thus, you must submit the model implementation code files with a detailed report on your interpretation of the results obtained. Discussion with fellow mates is allowed, provided you acknowledge your peer. Any similarity in the report will be treated as plagiarism.