ISB46703 Principles of Artificial Intelligence
Malaysian Insititute of Information Technology (MIIT) Universiti Kuala Lumpur
July 2023
Dr. Faiz

Assignment

Total marks: 40
Weight: 10%
Due: 20 September 2023 (Wednesday)

1 Instructions

K-Means clustering is a fundamental unsupervised learning algorithm used to group data points based on their similarity. In this assignment, you'll have the opportunity to apply K-Means clustering on a simple dataset.

2 Problem Statement

You've been given a customer dataset. It has basic information (ID, age, gender, income, spending score) about the customers. Spending Score is a value assigned to the customer based on some predefined parameters such as customer behaviour and purchasing data (they are not part of the problem in this assignment). Your task is to use the K-Means algorithm to group these customers into K clusters based on their **age** and **spending score**.

2.1 Customer Dataset

Dataset can be downloaded here: https://t.ly/u3ZRP

3 Instructions

1. Load and Explore Data (5 points)
    - Load the provided dataset using Python and import the necessary libraries (e.g., NumPy, Matplotlib, and Scikit-Learn).
    - Explore the dataset briefly to understand its structure.
2. Data Preprocessing (5 points)
    - Standardise the dataset if necessary to ensure all features have a similar scale.
3. K-Means Clustering (15 points)
    - Implement the K-Means clustering algorithm using the **KMeans** class from Scikit-Learn.

- Choose an appropriate value for K based on your analysis. Provide a brief justification for your choice..

4. Cluster Visualisation (10 points)
    - Visualise the clusters in a 2D scatter plot, using **Age** on the x-axis and **Spending Score** on the y-axis.
    - Different clusters should be indicated with distinct **colours** or **markers**.
5. Cluster Analysis (5 points)
    - Analyse the results. Discuss the characteristics of each cluster, and provide insights into why the data points were grouped in this manner.

4 Submission

Submit the following materials to VLE by **20 September 2023 (Thursday) 5:00pm** :

- Fully/clearly commented Jupyter Notebook (`.ipynb`) or code files used of the clustering and via VLE. Use Markdown cells when answering essay questions.

5 Statement on collaboration and individual submission, and plagiarism policy

While collaboration is permitted, it is strictly prohibited to engage in any form of plagiarism or copying. Plagiarism involves using someone else's work, ideas, or words without proper acknowledgment. It is a serious academic offense and goes against the principles of academic integrity. Any instances of plagiarism or copying will be thoroughly investigated, and appropriate penalties will be applied as per UniKL MIIT policies. This may include receiving a failing grade on the assignment or facing further disciplinary actions.

5.1 Avoiding plagiarism

Please ensure that all the work you submit is entirely your own. Any sources used in your assignment should be properly cited and referenced. This includes both direct quotes and paraphrased information.