|	Fall 2024 Assignment 1
Student Name:	
Student Number:	Aug 20, 2024

Assignment 1: Unsupervised Learning

Clustering is an unsupervised method for grouping patterns based on similarity such that the patterns in the same group are like one another, and those in different groups are less like one another. Clustering has many uses, including outlier detection and unsupervised image segmentation. The K-Means clustering algorithm is one of the most widely used clustering algorithms. In this assignment, you are going to implement basic k-means and analyze the effect of normalization.

Implement basic K-means algorithm to cluster the sample data set. Do not use built-in function such as kmeans for clustering.
Two sample data sets (kmtest and iris) are provided. Make sure your program works with both data sets.

A. What to DO:

1. Clustering with K-means algorithm for kmtest dataset.
   a. Without normalization, cluster the dataset by choosing the K value as 2, 3, 4, 5. Plot results for each K values by showing each cluster with different color and cluster centers.
   b. With normalization, cluster the dataset by choosing the K values as 2, 3, 4, 5. You should create clustering centers and clustering input for normalized data. Use z-score normalization as the normalization method. First normalize the data and apply clustering on the normalized data. Plot results for each K values by showing each cluster with different color and cluster centers.

2. Clustering with K-means algorithm for iris dataset only using column 1, 2, 3, and 4. Do not use label information in column 5.
   a. Cluster the dataset by choosing the K value as 3 without normalization. Run kmeans algorithm at least five times to get different results.
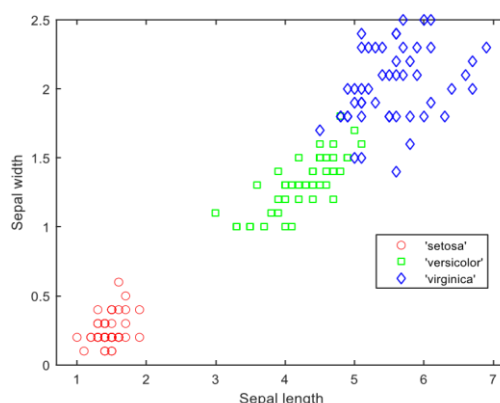
b. Using attribute 3 and 4, plot your best and worst clustering results by showing each cluster with different color and cluster centers.
c. Plot the original result using attribute 3 and 4 and compare it with your best result.
d. Calculate the distance between centers of your best result and original centers.

C. What to turn in:
- A zip file with all the necessary SOURCE code
- A written report (in pdf format) using the template including the following contents:
    - cover page with your name, class title, class number, date, etc.
    - an "Abstract" (no more than 100 words) summarizing what this project is about (objective), what you did, and what you found out in this project.
    - a "Result and Discussion" section showing your results, discussing them, and summarizing lessons learned, your experience working on the project, potential future work if given time, etc.
    - source code printout.
- Upload the zip file and project report to D2L as separate files.

D. Due Date: It is announced on D2L.