NATIONAL UNIVERSITY OF COMPUTER & EMERGING SCIENCES
FAST - PESHAWAR CAMPUS

Subject: AL 2002 - Artificial Intelligence Lab
Instructor: Muhammad Saood Sarwar

Lab Task: K-Means Clustering

Question 1: Customer Segmentation using K-Means Clustering

In this task, you are required to perform customer segmentation using the K-Means clustering algorithm. Follow the steps below to complete the task:

a) Load the customer segmentation dataset into your environment.

b) Perform data cleaning by handling missing values and removing duplicate records.

c) Normalize or standardize the features to ensure they are on the same scale.

d) Identify and select the most relevant features that contribute significantly to customer behavior.

e) Apply the K-Means clustering algorithm on the preprocessed dataset to group customers into meaningful clusters based on their behavior and demographics.

f) Visualize the resulting clusters using appropriate visualization techniques such as scatter plots or pair plots.

Question 2: Determining the Optimal Number of Clusters

After preprocessing your dataset and applying K-Means clustering, it is crucial to determine the most appropriate number of clusters (k) to ensure meaningful segmentation. Follow the steps below:

a) Apply the *Elbow Method* to analyze the Within-Cluster Sum of Squares (WCSS) and identify the point where adding more clusters yields diminishing returns.

b) Optionally, compute the *Silhouette Score* for different values of k to further validate the quality of clustering.

c) Select the optimal value of k based on the above evaluations.

d) Re-run the K-Means algorithm using the chosen value of k.

e) Visualize and interpret the final clustering results.