



National University



of Computer and Emerging Sciences, CFD Campus

AI 2002 - Artificial Intelligence (Spring 2024) Assignment # 6

<u>Topics Covered:</u> <ul style="list-style-type: none">• K-means clustering• K-medoid clustering	<u>Submission Deadline:</u> Sunday- May 5, 2024, by 23:00 sharp
<u>Submission Guidelines:</u> <ul style="list-style-type: none">• Submit PDF file for Problem #1 and Problem #4.• Submit PDF file along Zip file of codes for Problem #2 and Problem #3.• Submit your files in given naming conventions only. 21F_XXXX_Assignment_6_Sec• You are expected to submit your own code and answers, any help from internet/chatgpt/fellows is not allowed.	

Basic overview:

In this This assignment delves into the practical application of two fundamental clustering techniques: K-means clustering and K-medoid clustering. You will explore the principles and implementation of these algorithms to segment datasets into distinct groups based on similarity. Through hands-on experience, you will gain proficiency in utilizing these clustering techniques to analyze datasets and extract meaningful insights. You will also compare the performance and characteristics of K-means and K-medoid clustering, enhancing your understanding of clustering algorithms.

Problem # 1: Hand Written Task

Given the following dataset:

Point	X	Y	Z
A1	2	3	1
A2	3	4	7
A3	5	6	5
A4	6	7	9
A5	8	9	10
A6	7	2	6
A7	1	3	11



National University



of Computer and Emerging Sciences, CFD Campus

Perform **K-means clustering** with **K = 2**. Show the initial centroids, the assignment of points to clusters, and the final centroids.

Additionally, perform **K-medoid clustering** with **K = 2** using the same dataset. Show the initial medoids, the assignment of points to clusters, and the final medoids.

Problem # 2: Coding Task

Perform K-means clustering on the provided dataset of customer demographics. The dataset contains information about customers including their Gender, Age, Annual Income (in thousands), and Spending Score (ranging from 1 to 100). Your task is to segment these customers into distinct groups based on their characteristics. **Built in libraries are not allowed you can use NumPy, pandas and matplotlib**

1. Load the dataset into your preferred programming environment.
2. Preprocess the data as necessary. You may need to encode categorical variables like Gender.
3. Choose appropriate features for clustering. Consider using Age, Annual Income, and Spending Score for segmentation.
4. Try different number of K values i.e, K=2,3,4.
5. Perform K-means clustering on the selected features.
6. Visualize the clusters to gain insights into customer segmentation.
7. Interpret the results and provide recommendations for marketing strategies based on the identified customer segments.

Problem # 3: Coding Task

Perform K-medoids clustering on the provided dataset of customer demographics. The dataset contains information about customers including their Gender, Age, Annual Income (in thousands), and Spending Score (ranging from 1 to 100). Your task is to segment these customers into distinct groups based on their characteristics. **Built in libraries are not allowed you can use NumPy, pandas and matplotlib.**

1. Load the dataset into your preferred programming environment.
2. Preprocess the data as necessary. You may need to encode categorical variables like Gender.
3. Choose appropriate features for clustering. Consider using Age, Annual Income, and Spending Score for segmentation.
4. Try different number of K values i.e, K=2,3,4.
5. Perform K-medoids clustering on the selected features.
6. Visualize the clusters to gain insights into customer segmentation.

7. Interpret the results and provide recommendations for marketing strategies based on the identified customer segments.

Page 2



National University



of Computer and Emerging Sciences, CFD Campus

Problem # 4: Hand-Written Task

Compare the **K-means** and **K-medoids** clustering algorithms in terms of their characteristics, strengths, and weaknesses.

- Compare the computational complexity of the K-means and K-medoid algorithms. Which algorithm tends to be more computationally efficient for large datasets, and why?
- Explain the sensitivity of K-means and K-medoid clustering to initialization. How does the initial selection of centroids or medoids affect the final clustering outcome?
- Which clustering algorithm, K-means or K-medoid, is more robust to outliers? Why? How do outliers affect the clustering results in K-means and K-medoid algorithms differently?