Assignment No.:- 04

- **Aim :-** Implement k-Means algorithm

- **Problem statement :-**

We have given a collection of 8 points.
$P_1 = [0.1, 0.6]$, $P_2 = [0.15, 0.71]$, $P_3 = [0.08, 0.9]$, $P_4 = [0.16, 0.85]$
$P_5 = [0.2, 0.3]$, $P_6 = [0.25, 0.5]$, $P_7 = [0.24, 0.1]$, $P_8 = [0.3, 0.2]$.
Perform the k-mean clustering with initial centroids as $m_1 = P_1$ = Cluster#1 = C_1 & $m_2 = P_8$ = Cluster# 2 = C_2.
Answer the following.
1] Which cluster does P_6 belongs to?
2] What is the population of cluster around m_2?
3] What is updated value of m_1 & m_2?

- **Prerequisite :-**
  - Basic of Python
  - Data mining Algorithm
  - Concept of k-mean Clustering

- **Theory :-**

A Hospital care chain wants to open a series of Emergency-care wards within a region. We assume that the hospital knows the location of all the maximum accident prone areas in the region. They have to decide the number of emergency units to be opened & the location of these emergency units, so that all the accident-prone zones is covered in the vicinity of emergency units.

SAMAR

The challenge is to decide the location of these emergency units so that the whole region is covered. Here is when K-means clustering comes to rescue.

- What is clustering?
  - A cluster refers to a small group of objects. Clustering is grouping those objects into clusters.
  - In order to learn clustering, it is important to understand the scenarios that lead to cluster different objects.
  - Clustering is dividing data points into homogeneous classes or clusters.
  - Point in the same group are as similar as possible.
  - Point in different group are as dissimilar as possible.
  - When a collection of objects is given, we put objects into group based on similarity.

- Application of Clustering :-
  - Clustering helps marketers improve their customer base & work on the target areas. It helps group people (according to different criteria's such as willingness, purchasing power, etc.) based on their similarity in many ways related to the product under consideration.
  - Clustering helps in identification of groups of houses on the basis of their value, type & geographical locations.

SAMAR

- Clustering is used to study earth-quake. Based on the areas hit by an earthquake in a region, clustering can help analyse the next probable location where earthquake can occur.

- What is K-means clustering:-
  K-means (Macqueen, 1967) is one of the simplest unsupervisted learning algorithms that solve the well known clustering problem. K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining.

- Example:-
  A pizza chain wants to open its delivery centers across a city. What do you think would be the possible challenges:
  - They need to analyse the areas from where the pizza is being ordered frequently.
  - They need to understand as to how many pizza store has to be opened to ~~cover~~ delivery in the area.
  - They need to figure out the locations for the pizza stores within all these areas in order to keep the distance between the store & delivery points minimum.

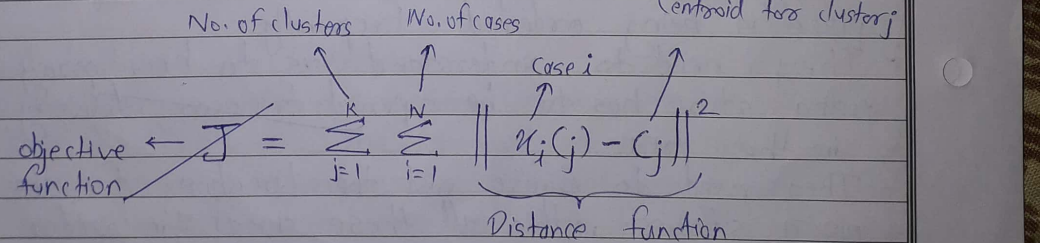
  Resolving these challenges includes a lot of analysis of mathematics. We would now learn about how clustering can provide a meaningful

SAMAR

and easy method of sorting out such real life challenges.

- k-means clustering method:
  If k is given, the k-means algorithm can be executed in the following steps:
  - Partition of objects into k non-empty subsets.
  - Identifying the cluster centroids (mean point) of the current partition.
  - Assigning each point to a specific cluster.
  - Compute the distances from each point & allot points to the cluster where the distance from the centroids is minimum.
  - After re-allotting the points, find the centroids of the new cluster formed.

- Mathematical Formulation of k-means algorithm:



SAMAR

Page 5

Date / /20

- Algorithm :-
1) Import the required packages.
2) Create dataset using dataframe.
3) Find centroids points
4) Plot the given points
5) For i in centroids():
6) Plot given elements with centroids elements.
7) Import kmeans class & create object of it.
8) Using labels find populations around centroids.
9) Find new centroids.

- Conclusion :- I understand how to implement k-means clustering algorithm successfully.

- Reference :-
  - Class notes.