Page No.
Date

Assignment No : 2

Title : Consider a suitable dataset for clustering of data instances in different groups, apply different clustering techniques (minimum 2) visualize the clusters using suitable tool

Objectives :

1. To understand concept of clustering
2. To understand different clustering techniques [Kmeans and Kmediods]

Outcomes

1. Students will understand concept of clustering
2. They will learn 2 clustering techniques.

Theory :

Clustering :- Clustering analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups (clusters)

K-Means clustering :-

It is a method of vector quantisation, originally from signal Processing, that aims to Partition n observations into K-clusters in which each observation belong to the cluster with the nearst mean, serving as the Prototype of the cluster.

Complexity for Optimal Solution to K-means clustering Problem for Observations in d dimension

If K & d are fixed

$O(n^{dK+1})$, wher n is no of entities

KMedioids

The K-mediods or Partitioning around medoids (PAM) algorithm is a clustering technique. It chose Points that can be made centers Unlike K-means, which may not Select data Points.

Page No.
Date

Complexity

The cost function is $O(n^2 k^2)$, which can be reduced to $O(n^2)$, by Spitting the change

Algorithms:

K-Means

1. Initialize K-cluster centers by selecting Points
2. Iterate through the Point.
3. Calculate distance of each Point to the center
4. Assign the Point to cluster closest to it
5. Calculate mean of each cluster
6. Select new cluster
7. Display final cluster

K-Medoids

1. Initialize greedily select k of the n data Points as the medoids to minimize the cost
2. Assign each point to the closest medoid
3. While the cost of the Configuration decreases
   1. For each medoid m, and for each non-medoid data point.

2 Consider the swap of m & o. and compute the cost change.

3 If the cost change is the current best, remember this m and o combination.

4 Perform the best swap of Mbest & Obest, if it decrease the cost function. otherwise, the algorithm terminate.

Conclusion :- Hence, we understood and implemented K-means and K-medoids clustering techniques.