

4359
Page No.:
Date:
ASSIGNMENT-8

AIM: Unsupervised learning K-Means clustering & Hierarchical clustering on proper dataset of class

OBJECTIVE :- To implement concept of clustering
- To implement concept of k-means algorithm
- To find convergence.

THEORY:

• K-MEANS CLUSTERING :

It aims to position n observations into K clusters in which each observation belongs to cluster with nearest mean, serving as prototype of clusters.
i. Initially K clusters are chosen randomly.

ii. Three centroids are used to train a KNN classifier. The resulting classifier is used to classify the data & produce a randomized set of clusters.

iii. Each centroid is set to arithmetic means of clusters it defines.

iv. This process of classification & centroid adjustment is repeated until the values of centroids stabilise.

v. Final centroids will be used to produce the final classification / clustering of i/p data.

vi. x

HIERARCHICAL CLUSTERING:

In data mining, & statistics, hierarchical clustering is a method of cluster analysis which seeks to build hierarchy.

Agglomerative: This is bottom-up approach. Each observation starts in its own type & pairs of clusters are merged as one moves down the hierarchy.

Divisive: This top-down approach. All observation starts in one cluster & pairs of clusters are merged as one moves down the hierarchy.

Working:

- i. Given set of N items, to be clustered, & $N \times N$ distance matrix, the basic process of Johnson's hierarchical clustering is this:
- i. Start by assigning each item to its own cluster, so that if you have N items, you have N clusters, each containing just one of them only
- ii. Find all closest pair of clusters & merge them into a single cluster so that you have one less cluster
- iii. Compute distances between new clusters & each of old clusters
- iv. Repeat (ii) & (iii) until all items are clustered into a single cluster of size N

CONCLUSION:

Thus we learnt & implemented K-Means clustering in Python