Date :

Practical No. 5

Aim: To implement k-means algorithm.

Descriptions :

Clustering :→

Clustering is the method of dividing a set of abstract objects into groups. points to keep in mind A set of data object can be viewed as a single entity. when performing cluster analysis, we divide the data set into groups based on data similarity, then assign labels to the groups.

Simple K means clustering :→

K means clustering is a simple unsupervised learning algorithm. In this, the data objects ('n') are grouped into a total 'k' clusters, with each observation belonging to the cluster with closest mean. It defines 'k' sets, one for each clusters k n . The clusters are separated by a large distance.

The data is then organized into acceptable data sets and linked to the nearest collection. If no data is pending, the first stage is more difficult to complete; in the case, an early grouping is performed. The 'k' new set must be recalculated as the barycenters of the clusters from the previous stage.

The same data set points and the nearest new sets are bound together after these 'k' new sets have been created. After that, a loop is created. The 'k' sets change their position step by step until no further changes are made as a result of this loop.

k-means clustering algo. computers the centroids and iterates until we it finds optimal centroid. It assumes that the no. of clusters are already known. It is also called flat clustering algo. The no. of clusters identified from data by algo. is represented by 'k' in k-means.

In this algo. the data points are assigned to a cluster in such a manner that the sum of the squared distance between the

Date :

data points and centroid would be minimum. It is to be understood that less variation within the clusters will leads to more similar data points within same clusters.

working of k-means Algorithm: →

We can understand the working of k-means clustering algo. with the following steps:

Step 1) First, we need to specify the number of clusters, K, need to be generated by this algorithm.

Step 2: Next, randomly select K data points and assign each data point to a cluster.

Step 3: Now it will compute the cluster centroids

Step 4: Next, keep iterating the following until we find optimal centroid which is the assignment of data points to the clusters that are not changing any more -

Result:

This program has been successfully executed.

Viva Questions:

① What is the purpose of k-means algorithm?

→ It is used to find group which have not been explicitly labeled in the data. It is unsupervised learning algo., which groups the unlabeled datasets into different clusters.

② How do we use k means clustering algo. in weka?

→ step 1: open weka and load dataset.

step 2: find cluster tab in explorer and choose button to execute clustering and select simple - k-means algo.

step 3: Then to the right of choose icon, press text- button and enter three for number of clusters and leave the seed value alone.

Date :

Step 4 : The choice to use training set is selected and then 'start' button is pressed.

Step 5 : The centroid of each cluster is shown in result window and right click the result set on result, selecting to visualize cluster assignment from list column.

③ what does k-mean in K-mean algorithm?

→ The number of clusters found from data by the method is denoted by the letter 'k' in K-means.

④ what type of algorithm is K-means?

→ It is simple unsupervised learning algorithm. It is widely used centroid-based clustering algorithm.