Date / /
Page No.

ASSIGNMENT 4

Title: Implement K-Means algorithm for clustering to create a cluster on the given data (using Python) (use Iris dataset)

Theory:

K-Means clustering: One of the simplest and most widely used unsupervised learning algorithm. It involves a simple way to classify the data set into fixed no. of K clusters. The idea is to define K centroids one for each cluster. The final clusters depend on the initial configuration of centroids. So they should be initialized as far from each other as possible. K-means is iterative in nature and easy to implement.

1) Let there be N data points. At first, K centroids are initialized in our dataset representing K different clusters.
2) Now, each of the N data points are reassigned to closest centroid in the dataset and merged with that centroid as a single cluster. In this way every data point is assigned to one of the centroids.
3) Then, K cluster centroids are calculated and again, each of the N data points are reassigned to the nearest centroid.

Date / /
Page No.

4) Step 3 is repeated until no further improvement can be made.

5) This algorithm aims at minimising the objective function

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \| x_i^{(j)} - c_j \|^2$$

It represents the sum of Euclidean distance of all the data points from the cluster its centroids which is minimized

Conclusion: Thus we have performed clustering using K-means algorithm.

7)