Experiment 5

Shashwat Shah
60004220126
DIV B C2-2

Aim: Implementation of clustering algorithm using Kmeans and hierarchial clustering.

Theory: Clustering is a technique that involves grouping similar instances or data points into distict clusters based on certain criteria. It is an unsupervised learning algorithm. It identifies similar patterns or data.

K means clustering:
It is a partial clustering algorithm that is used to partition a dateset into predefined no. of clusters. It is a simple and efficient algorithm that works well with data that has a spherical shape. It is sensitive to choice of k values.

Hierarchial clustering.
It is an agglomerative clustering algorithm that builds a hierarchy of clusters by merging similar clusters together. It is more expensive than the K means when compared computationally.
The three main variations are:-

(1) Single linkage - Merges cluster based on the shortest distance between any 2 data points.
(2) Complete linkage - Merges clusters based on distance largest between any 2 data points.
(3) Avarge linkage: Merges clusters based on avarge distance

Conclusion: Thus we have implemented hierarchial k-means clustering