EXPT.5

AIM: Implementation of Clustering algorithm ( K-means in java/python).

Theory :

The algorithm works as follows:

1. First we initialize k points, called means, randomly.
2. We categorize each item to its closest mean and we update the mean’s coordinates, which are the averages of the items categorized in that mean so far.
3. We repeat the process for a given number of iterations and at the end, we have our clusters.

The “points” mentioned above are called means, because they hold the mean values of the items categorized in it. To initialize these means, we have a lot of options. An intuitive method is to initialize the means at random items in the data set. Another method is to initialize the means at random values between the boundaries of the data set (if for a feature *x* the items have values in [0,3], we will initialize the means with values for *x* at [0,3]).

The above algorithm in pseudocode:

Initialize k means with random values

For a given number of iterations:

Iterate through items:

Find the mean closest to the item

Assign item to mean

Update mean

Program code:

Post lab:

1. Differentiate K-means and K-medoide algorithms with one example
2. Explain DBScan
3. Explain Hierarchical clustering (Agglomerative and Divisive clustering)