120A3051
Shreya Idate
Batch: E3

EXPERIMENT NO. 9

Aim: Clustering:

      a) Clustering algorithms for unsupervised classification.
      b) Plot the cluster data.

Theory:

Unsupervised learning is a machine learning technique in which models are not supervised using training dataset. Instead, models itself find the hidden patterns and insights from the given data. It can be compared to learning which takes place in the human brain while learning new things. The goal of unsupervised learning is to find the underlying structure of dataset, group that data according to similarities, and represent that dataset in a compressed format.

K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. It groups the unlabelled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on. It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabelled dataset on its own without the need for any training. It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters. The algorithm takes the unlabelled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.
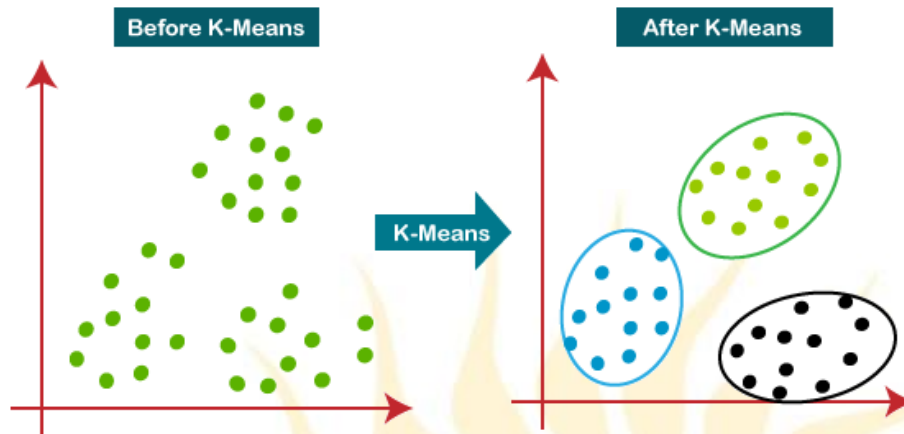
The k-means clustering algorithm mainly performs two tasks:

Determines the best value for K center points or centroids by an iterative process.

Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.

Hence each cluster has datapoints with some commonalities, and it is away from other clusters.

The below figure explains the working of the K-means Clustering Algorithm:

Before K-Means   K-Means   **After K-Means**

Program:

```
[1]  # Importing required libraries
     import numpy as np
     import matplotlib.pyplot as plt
     from sklearn.cluster import KMeans


[2]  # Generating sample data
     np.random.seed(0)
     X = np.random.randn(100, 2)


[3]  # Initializing KMeans algorithm with 2 clusters
     kmeans = KMeans(n_clusters=2, random_state=0)
```
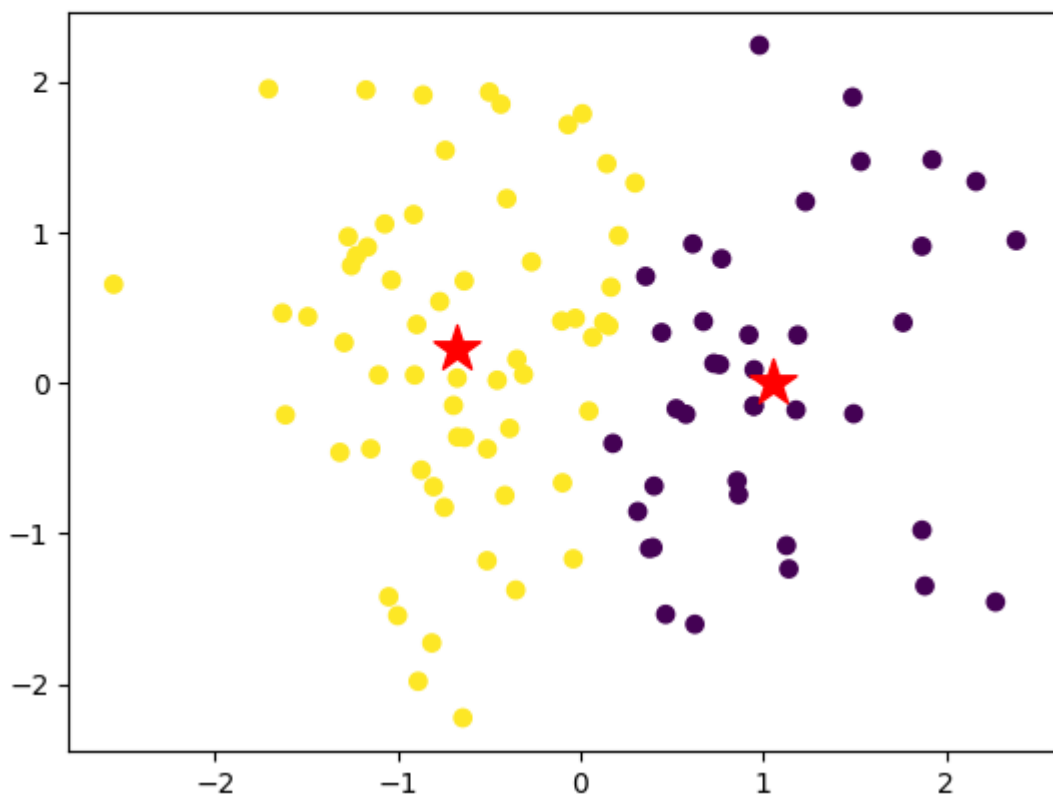
```
[4]  # Fitting the data
     kmeans.fit(X)

     /usr/local/lib/python3.9/dist-packages/sk
       warnings.warn(

              ▼          KMeans
       KMeans(n_clusters=2, random_state=0)
```

```
[5]  # Getting the cluster labels for each data point
     labels = kmeans.labels_
```

```
[6]  # Getting the coordinates of the cluster centers
     centroids = kmeans.cluster_centers_
```

```
[7]  # Plotting the data points and centroids
     plt.scatter(X[:, 0], X[:, 1], c=labels)
     plt.scatter(centroids[:, 0], centroids[:, 1], marker='*', s=300, c='r')
     plt.show()
```

Conclusion: Successfully implemented K-mean clustering algorithm using Python.