Name : Apurv Sarode

UID : 2019130054

Subject : AIML

Experiment : 5

**Aim :** Given a dataset, classify the input into k categories.

**Theory:**

**K-Means Clustering Algorithm**

K-Means Clustering is an unsupervised learning algorithm that is used to solve

the clustering problems in machine learning or data science. In this topic, we

will learn what is K-means clustering algorithm, how the algorithm works,

along with the Python implementation of k-means clustering.

**What is K-Means Algorithm?**

K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of predefined 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 is an iterative algorithm that divides the unlabeled dataset into k different clusters in such a way that each dataset belongs only one group that has similar properties.

It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled 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 unlabeled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best

clusters.

**Program:**

Graphical user interface, text

Description automatically generated

Table

Description automatically generated with medium confidence

Graphical user interface, text

Description automatically generated

Chart, line chart

Description automatically generated

Graphical user interface, text

Description automatically generated

Chart, scatter chart, bubble chart

Description automatically generated

**Conclusion:**

In the above experiment of AIML Lab, I learnt about K-Means Clustering Algorithm which is a unsupervised learning.

First, we Import all the necessary modules. Then we import the Mall Customer dataset. We then try to find the optimal number of clusters for our clustering problem by the elbow method which uses the WCSS concept to draw the plot by plotting WCSS values on the Y-axis and the number of clusters on the X-axis. After that, we initialize the for loop for the iteration on a different value of k ranging from 1 to 10.

Then we see that the Elbow point is 5. We then train the model on the dataset.

The last step is to visualize the clusters. As we have 5 clusters for our model, so we will visualize each cluster one by one. Then from the graph we see that clusters are formed between two parameters of the dataset; Annual income of customer and Spending.