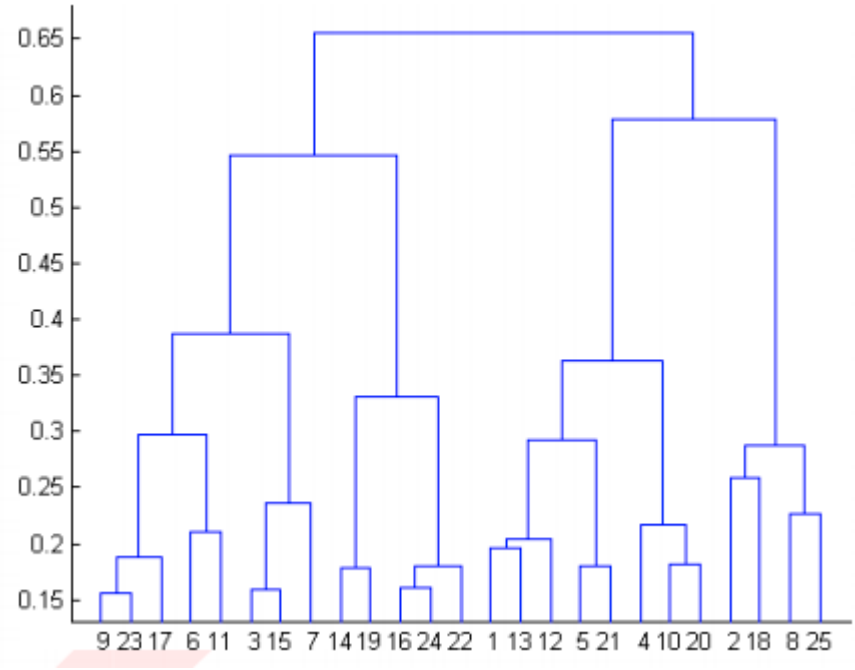
MACHINE LEARNING ASSIGNMENT – 1

Q1 to Q12 have only one correct answer. Choose the correct option to answer your question.

1. What is the most appropriate no. of clusters for the data points represented by the following dendrogram:



a) 2

b) 4

c) 6

d) 8

Answer: - b) 4 Clusters

2. In which of the following cases will K-Means clustering fail to give good results?

1. Data points with outliers

2. Data points with different densities

3. Data points with round shapes

4. Data points with non-convex shapes

Options:

a) 1 and 2

b) 2 and 3

c) 2 and 4

d) 1, 2 and 4

Answer: - d) 1,2 and 4

3. The most important part of is selecting the variables on which clustering is based.

a) interpreting and profiling clusters

b) selecting a clustering procedure

c) assessing the validity of clustering

d) formulating the clustering problem

Answer: - d) Formulating the Clustering Problem

4. The most commonly used measure of similarity is the or its square.

a) Euclidean distance

b) city-block distance

c) Chebyshev's distance

d) Manhattan distance

Answer: - a) Euclidean Distance

5. _____ is a clustering procedure where all objects start out in one giant cluster. Clusters are formed by dividing this cluster into smaller and smaller clusters.

a) Non-hierarchical clustering

b) Divisive clustering

c) Agglomerative clustering

d) K-means clustering

Answer: - b) Divisive Clustering

6. Which of the following is required by K-means clustering?

a) Defined distance metric

b) Number of clusters

c) Initial guess as to cluster centroids

d) All answers are correct

Answer: - d) All answers are Correct

7. The goal of clustering is to

a) Divide the data points into groups

b) Classify the data point into different classes

c) Predict the output values of input data points

d) All of the above

Answer: - d) All of the above

8. Clustering is a

a) Supervised learning

b) Unsupervised learning

c) Reinforcement learning

d) None

Answer: - b) Unsupervised learning

9. Which of the following clustering algorithms suffers from the problem of convergence at local optima?

a) K- Means clustering

b) Hierarchical clustering

c) Diverse clustering

d) All of the above

Answer: - d) All of the above

10. Which version of the clustering algorithm is most sensitive to outliers?

a) K-means clustering algorithm

b) K-modes clustering algorithm

c) K-medians clustering algorithm

d) None

Answer: - a) K-Means Clustering Algorithm

11. Which of the following is a bad characteristic of a dataset for clustering analysis

a) Data points with outliers

b) Data points with different densities

c) Data points with non-convex shapes

d) All of the above

Answer: - d) All of the above

12. For clustering, we do not require

a) Labeled data

b) Unlabeled data

c) Numerical data

d) Categorical data

Answer: - a) Labeled Data

13. How is cluster analysis calculated?

Answer: - Cluster Analysis is calculated by following methods:

a) Connectivity Model: In this, models are build based on distance connectivity

b) Centroid Model: In this, models are made in such a way that each cluster is represented by single mean vector.

c) Distribution Model: In this, clusters are modeled using statistical distributions such as multivariate normal distributions.

d) Density Model: DBSCAN and OPTICS defines clusters as connected dense regions in the data space.

14. How is cluster quality measured?

Answer: - It is very important to measure the cluster quality. But there is no common measure to evaluate it. We can do it by following three measures:

a) **Internal**: In this, the incites from data are generated from dataset itself.
b) **External:** It is applicable when there is some prior knowledge is available for dataset.
c) **Relative**: In this, we compare clusters by modifying the parameters of the algorithm.

15. What is cluster analysis and its types?

Answer: - It is the task of grouping a set of datapoints in such a way that they can be divided according to their similarity. In order to perform this, we have four types of clustering techniques: -

a) Centroid Clustering
b) Density Clustering
c) Distribution Clustering
d) Connectivity Clustering