ASSIGNMENT – 1

MACHINE LEARNING

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:

Answer: d) 8

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**

Answer: a) 1 and 2

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

Answer: d) formulating the clustering problem

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

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.

Answer: b) Divisive clustering

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

Answer: d) All answers are correct

7. The goal of clustering is to-

Answer: d) All of the above

8. Clustering is a

Answer: b) Unsupervised learning

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

Answer: a) K- Means clustering

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

Answer: a) K-means clustering algorithm

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

Answer: d) All of the above

12. For clustering, we do not require-

Answer: a) Labeled data

13. How is cluster analysis calculated?

Answer: The cluster analysis follows three basic steps: 1) calculate the distance between the datapoints, 2) join the clusters together, and 3) choose a solution by selecting the right number of clusters.

14. How is cluster quality measured?

Answer: Quality of cluster similarity or dissimilarity measure between cluster points on basis of distance the groups of datapoints

15. What is cluster analysis and its types?

Answer: Cluster analysis is the way to find similar datapoints and group them in order to form clusters. Types of Cluster Analysis: - a) Hierarchical Cluster Analysis, b) Centroid-based Clustering, c) Distribution-based Clustering, d) Density-based Clustering