

## MACHINE LEARNING

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

Ans. b) 4

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

Ans. d) 1,2 and 4

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

Ans. d) formulating the clustering problem

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

Ans. 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.

Ans. b) Divisive clustering

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

Ans. d) All answers are correct

7. The goal of clustering is to-

Ans. a) Divide the data points into groups

8. Clustering is a-

Ans. b) Unsupervised learning

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

Ans. d) All of the above

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

Ans. a) K-means clustering algorithm

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

Ans. d) All of the above

12. For clustering, we do not require-

Ans. a) Labeled data

13. How is cluster analysis calculated?

Ans. It is calculated by measuring the distance between each data point and its centroid, squaring this distance, and summing these squares across one cluster. A good model is one with low inertia and a low number of clusters ( K ).

14. How is cluster quality measured?

Ans. To measure the quality of a clustering, we can use the average silhouette coefficient value of all objects in the data set.

Silhouette Method: This technique measures the reparability between clusters. First, an average distance is found between each point and all other points in a cluster. Then it measures the distance between each point and each point in other clusters.

15. What is cluster analysis and its types?

Ans. Cluster analysis is a data analysis technique that explores the naturally occurring groups within a data set known as clusters. Cluster analysis doesn't need to group data points into any predefined groups, which means that it is an unsupervised learning method.

There are 6 types of clustering algorithms in Machine learning. They are as follows –

1. Centroid-based,
2. Density-based,
3. Distribution-based,
4. Hierarchical,
5. Constraint-based,
6. Fuzzy clustering