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:
b) 4

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
d) 1, 2 and 4

3. The most important part of is selecting the variables on which clustering is based.
b) selecting a clustering procedure

4. The most commonly used measure of similarity is the or its square.
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.
b) Divisive clustering

6. Which of the following is required by K-means clustering?
a) Defined distance metric

7. The goal of clustering is to
d) All of the above

8. Clustering is a
b) Unsupervised learning

9. Which of the following clustering algorithms suffers from the problem of convergence at local optima?
a) K- Means clustering

10. Which version of the clustering algorithm is most sensitive to outliers?
a) K-means clustering algorithm

11. Which of the following is a bad characteristic of a dataset for clustering analysis
b) Data points with different densities

12. For clustering, we do not require
a) Labeled data

Q13 to Q15 are subjective answers type questions, Answers them in their own words briefly

13. How is cluster analysis calculated?
Ans: Steps involved in grid-based clustering algorithm are:

1.Divide data space into a finite number of cells.

2.Randomly select a cell 'c', where c should not be traversed beforehand.
3.Calculate the density of 'c'
4.If the density of 'c' greater than threshold density. ...
5.Repeat steps 2,3 and 4 till all the cells are traversed.
6.Stop.

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 d ata set.

15. What is cluster analysis and its types
Ans:Agglomerative clustering starts with single objects and starts grouping them into clusters.
The divisive method is another kind of Hierarchical method in which clustering starts with the complete data set and then starts dividing into partitions.