MACHINE LEARNING

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

   Answer: 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

   Answer: D. 1,2,4

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

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

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

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

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

8. Clustering is a-

a) Supervised learning

b) Unsupervised learning

c) Reinforcement learning

d) None

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

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

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

12. For clustering, we do not require

a) Labeled data

b) Unlabeled data

c) Numerical data

d) Categorical data

13. How is cluster analysis calculated?

In cluster analysis the data or information is grouped of cluster analysis is known as segmentation analysis. It tries to make homogenous groups so that it helps to analysis data more efficiently. It is often Used in conjunction with other analysis. for eg. Discriminant analysis The cluster analyse does not make any distinction between. independent and dependent variables

14. How is cluster quality measured

Calculation steps

Step 1: Choose the number of clusters

step 2: We have to make selection of centroid

step 3: Assign each data element to its nearest centroid

step 4: New selection of centroid should be mode for each cluster.

Steps 5. Repeat the process from steps 3, until the centroids don't change

15. What is cluster analysis and its types?

Cluster analysis is a multivariate data mining technique whose goal is to groups objects based on a set of user selected characteristics or attributes. It is the basic and most important step of data mining and a common technique for statistical data analysis, and it is used in many fields such as data compression, machine learning, pattern recognition, information retrieval etc

Types of Cluster Analysis

A. **Hierarchical Cluster Analysis**
B. **Centroid-based Clustering**
C. **Distribution-based Clustering**
D. **Density-based Clustering**