MACHINE LEARNING

Q1 to Q12 have only one correct answer. (Answers are marked in red)

Q1 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
Ans :- 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
Options:
a) 1 and 2
b) 2 and 3
c) 2 and 4
d) 1, 2 and 4
Ans : - 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
Ans :- 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
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.
a) Non-hierarchical clustering
b) Divisive clustering
c) Agglomerative clustering
d) K-means clustering
Ans : - 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
Ans : - 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
Ans : - a) Divide the data points into groups

8. Clustering is a-
a) Supervised learning
b) Unsupervised learning
c) Reinforcement learning
d) None
Ans ; - 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
Ans ; - a) K- Means clustering

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
Ans : - 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
Ans : - d) All of the above

12. For clustering, we do not require
a) Labeled data
b) Unlabeled data
c) Numerical data
d) Categorical data
Ans : - a) Labeled data

13. How is cluster analysis calculated?
Ans :- Cluster analysis is a process of grouping similar objects or data points into clusters or subgroups based on their characteristics or attributes. There are several methods of cluster analysis, including hierarchical clustering and k-means clustering.
1 . Hierarchical clustering:
Hierarchical clustering is a method that involves creating a tree-like diagram called a dendrogram that represents the relationships between different objects or data points.
2. K-means clustering:
K-means clustering is a method that involves partitioning the data into k clusters, where k is a predetermined number. The goal is to minimize the sum of squared distances between each data point and the centroid of its assigned cluster
The final clusters obtained from both methods depend on the choice of distance metric, linkage criterion, or number of clusters, and the interpretation of the clusters is subjective and may require domain knowledge.

14. How is cluster quality measured?
Ans :- The quality of clusters in cluster analysis can be measured using different metrics depending on the clustering algorithm used and the objectives of the analysis.
One of them is Silhouette coefficient
The silhouette coefficient measures how similar an object is to its own cluster compared to other clusters. It ranges from -1 to 1, where values closer to 1 indicate that the object is well-matched to its own cluster and poorly matched to neighboring clusters, and values closer to -1 indicate the opposite. An average silhouette coefficient is calculated for all objects in the dataset, and higher values indicate better clustering quality.

15. What is cluster analysis and its types
Ans:- Cluster analysis is a data analysis technique that is used to group similar objects or observations into clusters or subgroups based on their attributes or characteristics. The goal of cluster analysis is to identify natural groupings or patterns in the data, which can then be used for further analysis or decision-making.
Some types Of Cluster Analysis:-

1. Hierarchical Cluster Analysis
2. Centroid-based Clustering
3. Distribution-based Clustering
4. Density-based Clustering