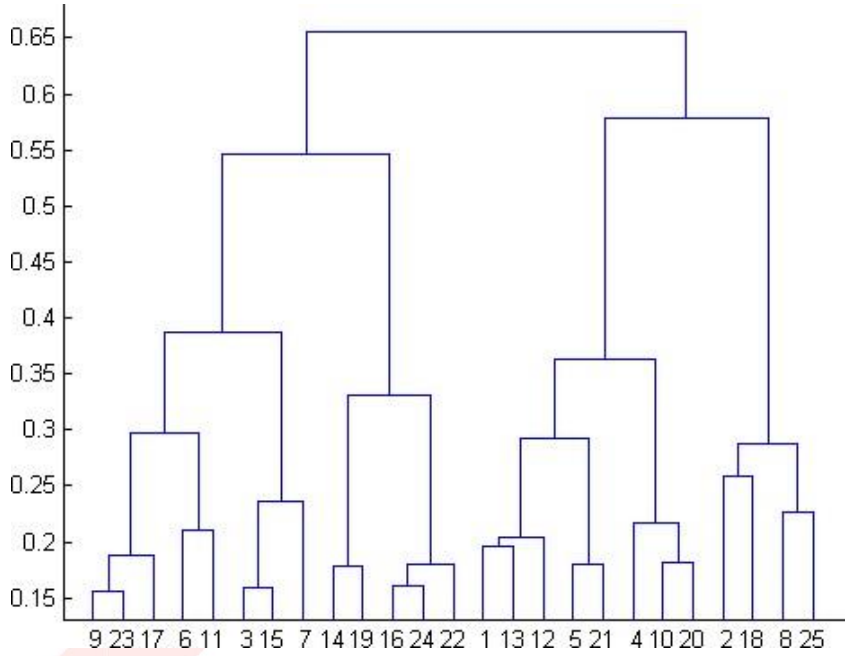
MACHINE LEARNING

Q1 to Q12 have only one correct answer. Choose the correct option to answer your question.

Answers: 1 to 12

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



a) 2
b)
c)
d)

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)
   c)
   d)

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

4. The most commonly used measure of similarity is the____or its square.
   a) Euclidean distance
   b)
   c)
   d)

MACHINE LEARNING

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)
   b) Divisive clustering
   c)
   d)

6. Which of the following is required by K-means clustering?
   a)
   b)
   c)
   d) All answers are correct

7. The goal of clustering is to-
   a) Divide the data points into groups
   b)
   c)
   d)

8. Clustering is a-
   a)
   b) Unsupervised leaming
   c)
   d)

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

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

11. Which of the following is a bad characteristic of a dataset for clustering analysis-
   a)
   b)
   c)
   d) All of the above

12. For clustering, we do not require-
   a) Labeled data
   b)
   c)
   d)

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

Answers: 13 to 15

13. How is cluster analysis calculated?

Ans. Cluster analysis is an unsupervised machine learning technique used to group similar data points together based on some similarity or distance measure. Cluster analysis can be calculated using a variety of algorithms, including K-means clustering, hierarchical clustering, density-based clustering, and others.

14. How is cluster quality measured?

Ans. Cluster quality can be measured using various metrics that evaluate the performance and accuracy of the clustering algorithm. some commonly used metrics for evaluating cluster quality are silhouette score, davies-Bouldin index, calinski-harabasz index, rand index and, jaccard index.

15. What is cluster analysis and its types?

Ans. Cluster analysis is an unsupervised machine learning technique used to group similar data points together based on some similarity or distance measure. The goal of cluster analysis is to identify patterns or structure in the data and group the data points into subsets or clusters that are similar to each other and dissimilar to the data points in other clusters.
There are several types of cluster analysis algorithms, including hierarchical clustering, K-means clustering, density-based clustering, model-based clustering and fuzzy clustering.