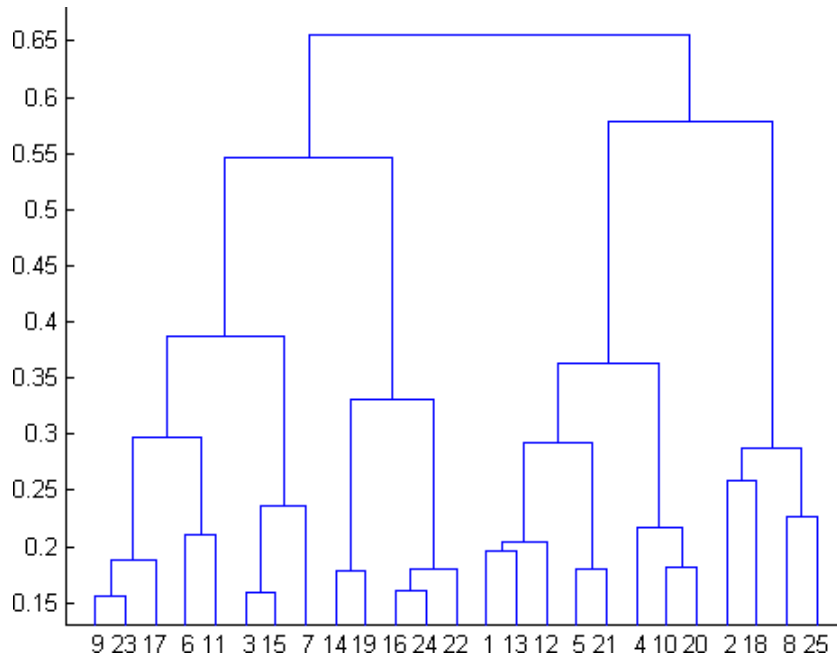


MACHINE LEARNING

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:



Answer B) 4 - Most appropriate no. of clusters for the analyzed data points is 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 is 1, 2 and 4 that is Data points with outliers, Data points with different densities & Data points with non-convex shapes.

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

Answer is 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

Answer is a) Euclidean distance

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) Non-hierarchical clustering
 - b) Divisive clustering
 - c) Agglomerative clustering
 - d) K-means clustering

Answer is a) 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

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

Answer is a) Divide the data points into groups

8. Clustering is a-
- a) Supervised learning
 - b) Unsupervised learning
 - c) Reinforcement learning
 - d) None

Answer is 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

Answer is 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

Answer is a) K-means clustering

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

FLIP ROBO

MACHINE LEARNING

- a) Data points with outliers
- b) Data points with different densities
- c) Data points with non-convex shapes
- d) All of the above

Answer is a) All of the above

12. For clustering, we do not require-

- a) Labeled data
- b) Unlabeled data
- c) Numerical data
- d) Categorical data

Answer is a) Labeled data

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

13. How is cluster analysis calculated?

It is calculated by measuring the distance between each data point and its centroid, squaring this distance, and summing these squares across one cluster.

14. How is cluster quality measured?

We can use the average coefficient value of all objects in the data set.

15. What is cluster analysis and its types?

Cluster analysis is the method of grouping the entities based on similarities. It is the process of finding similar structures in a set of unlabeled data to make it more understandable and manipulative.

Clustering generally divided into **Hard & Soft Clustering**.

Below are main methods used in Machine learning.

Partitioning Clustering
Density-Based Clustering
Distribution Model-Based Clustering
Hierarchical Clustering
Fuzzy Clustering
