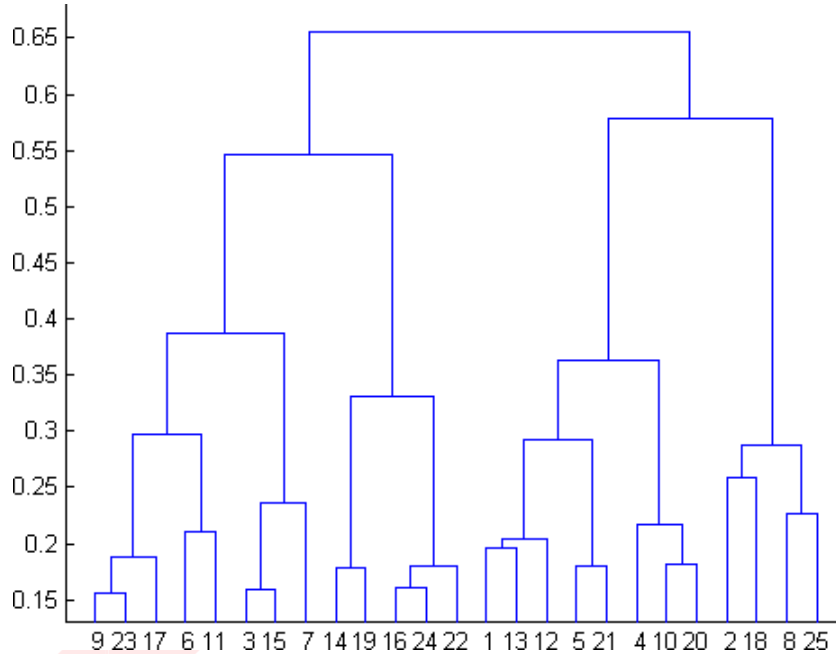


## MACHINE LEARNING

**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:



- a) 2  
b) 4  
c) 6  
d) 8

**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

Options:

- a) 1 and 2  
b) 2 and 3  
c) 2 and 4  
d) 1, 2 and 4

**Answer:c]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

**Answer:d]Formulating the clustering problem**

**MACHINE LEARNING**

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: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: 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

**Answer: d] All answer 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: a] Divide the data points into groups**

8. Clustering is a-

- a) Supervised learning
- b) Unsupervised learning
- c) Reinforcement learning
- d) None

**Answer: 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: d] All 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

**Answer: 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

**Answer: d] All of the above**

12. For clustering, we do not require-

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

**Answer: a] Labeled data**

## MACHINE LEARNING

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

13. How is cluster analysis calculated?

**Answer:**1]Calculate the distance

2]Link the clusters

3]Choose a solution by selecting the right number of clusters

14. How is cluster quality measured?

**Answer:**There are 2 methods they are extrinsic and intrinsic

1] Extrinsic:supervised,i.e,the growth truth(ideal clustering,e.g built by domain experts)is available.

Compare a clustering against the ground truth using certain clustering quality measure.Ex bcubed precision and recall metrics

2] Intrinsic:unsupervised,i.e,the ground truth is unavailable

Evaluate the goodness of a clustering by considering how well the clusters are separated,and how compact the clusters are.EX Silhouette coefficient

15. What is cluster analysis and its types?

**Answer:**Cluster analysis is a data analysis technique that explores the naturally occurring groups within a data set known as clusters.Cluster analysis doesn't need to group data points into any predefined groups which means that is an unsupervised learning method.ex: Streaming service often use this analysis to identify viewers who have similar behaviour and also they can collect data about how much a individual have watched this streaming for minutes or for an hour.

### **TYPES**

1] Centroid based clustering

2] Density based clustering

3] Distribution based clustering

4] Hierarchical clustering

---