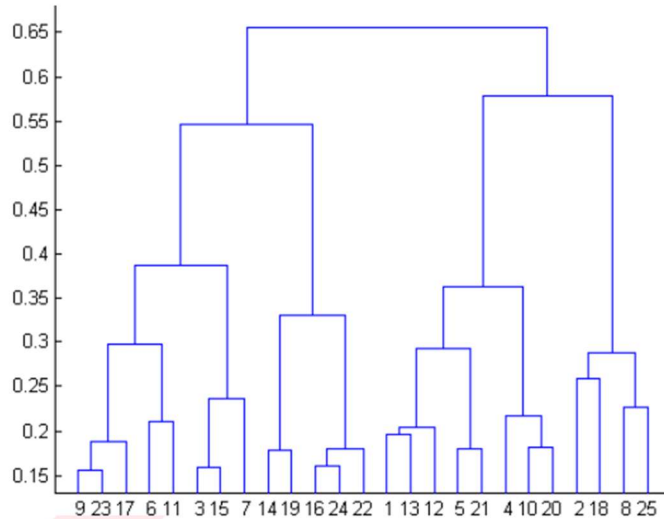


### MACHINE LEARNING

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

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



- 2
- 4
- 6
- 8

FLIP ROBO

Ans: b

- In which of the following cases will K-Means clustering fail to give good results?

- Data points with outliers
- Data points with different densities
- Data points with round shapes
- Data points with non-convex shapes

Options:

- 1 and 2
- 2 and 3
- 2 and 4
- 1, 2 and 4

Ans: a

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

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: c

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: d

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

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

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

Ans: b

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:b

FLIP ROBO

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

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

12. For clustering, we do not require-

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

Ans: a

13. How is cluster analysis is calculated

a typical clustering analysis approach via iteratively partitioning training data set to learn a partition of the given data space

learning a partition on a data set to produce several non-empty clusters (usually, the number of clusters given in advance)

in principle, optimal partition achieved via minimising the sum of squared distance to its “representative object” in each cluster

14. How cluster quality is measured

- Compare a clustering against the ground truth using certain clustering quality measure
- Ex. Purity, precision and recall metrics, normalized mutual information 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

15.What is cluster analysis and its type

The method of identifying similar instances & keeping them together is called clustering.

Clustering is a unsupervised approach which finds structure/ pattern in a collection of unlabelled data, Its type are

- 1)K mean cluster
- 2) Hierarchical cluster
- 3) DBSCAN