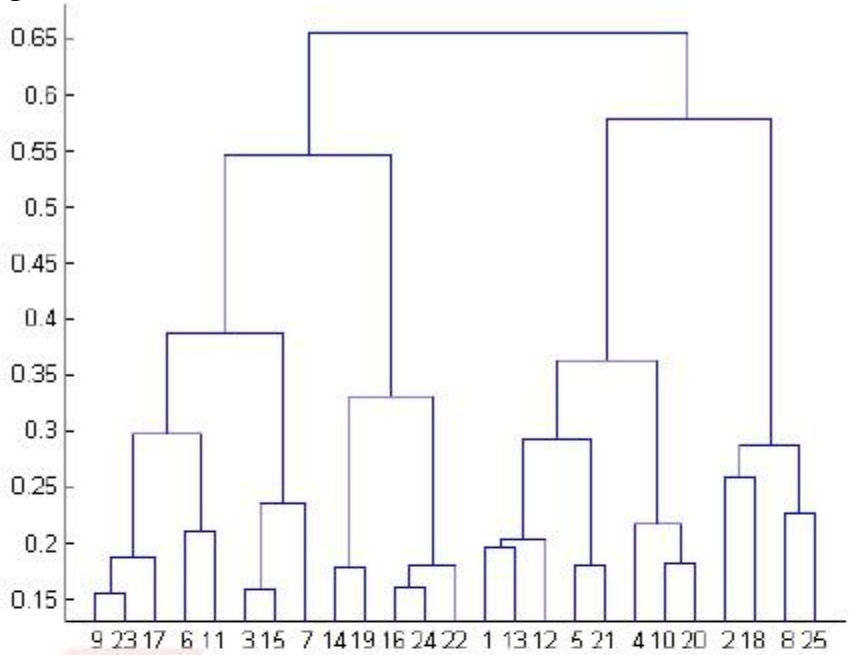
# MACHINE LEARNING

**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**

(The most appropriate no. of clusters for the data points represented by the above dendrogram:04)

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: D**
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:A**

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**
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**

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**

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

   **Answer: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

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

    **Answer:A**

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

13. How is cluster analysis calculated

Following three steps are necessary for cluster analysis:

- Copy your data into the table
- Select more than one variable
- Select the number of clusters you want to calculate

Clusters can be calculated using various grouping methods. These can be divided into

- graph-theoretical
- hierarchically
- partitioning
- optimizing

The number of clusters in the k-Means method must be determined before the start and is therefore not determined by the cluster method. The elbow method is a common way to determine the appropriate number of clusters.

14. How is cluster quality measured?

**Answer:** To measure the quality of a clustering, we can use the average silhouette coefficient value of all objects in the data set

15. What is cluster analysis and its types?

**Answer:** Cluster Analysis is the process to find similar groups of objects in order to form clusters.

It is an unsupervised machine learning-based algorithm that acts on unlabelled data.

A group of data points would comprise together to form a cluster in which all the objects would Belong to the same group.

The clustering methods can be classified into the following categories:

- Partitioning Method
- Hierarchical Method
- Density-based Method
- Grid-Based Method
- Model-Based Method
- Constraint-based Method