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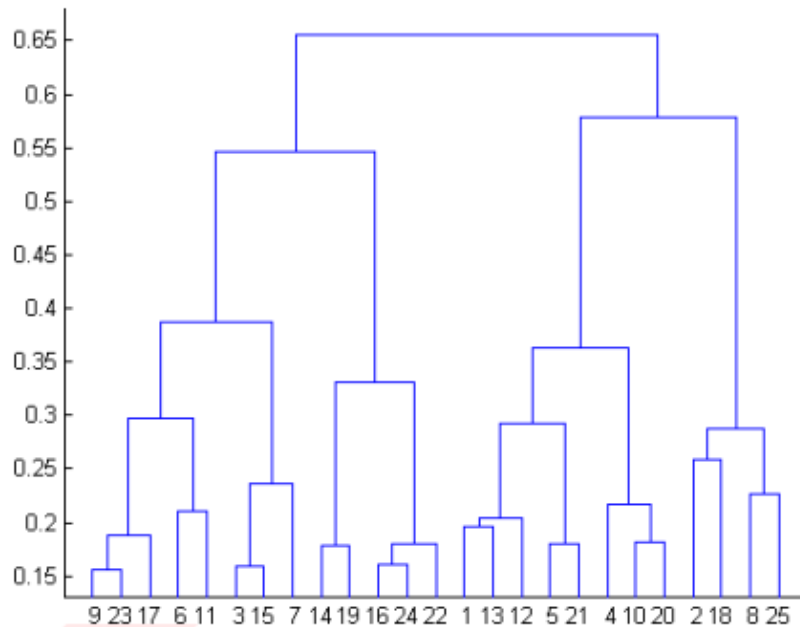
1. What is the most appropriate no. of clusters for the data points represented by the following dendrogram:

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

Answer (b) 4 : The most appropriate no. of clusters for the data points represented by the given dendrogram 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
   Options:
   a) 1 and 2
   b) 2 and 3
   c) 2 and 4
   d) 1, 2 and 4

Answer: d) 1,2 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

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

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 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: 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 of 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

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

13. How is cluster analysis calculated?

Answer: The hierarchical cluster analysis follows three basic steps: 1) calculate the distances, 2) link the clusters, and 3) choose a solution by selecting the right number of clusters.

14. How is cluster quality measured?

Answer: There are some methods to measure the Qualities of Good Clustering

1. Dissimilarity/Similarity metric
2. Cluster completeness
3. Ragbag
4. Small cluster preservation

15. What is cluster analysis and its type?

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.

Clustering Types:
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