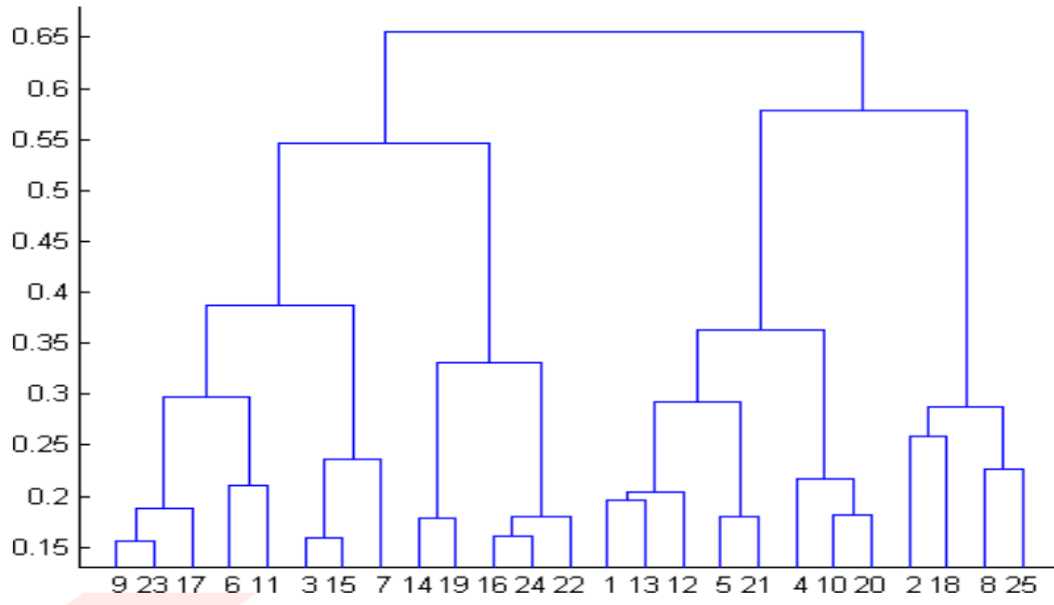


# Worksheet 1 Machine Learning

Q 1. What is the most appropriate no. of clusters for the data points represented by the following Dendrogram:



Ans: b) 4 is the most appropriate

Q 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:

Ans: d) 1, 2 and 4

Q 3. The most important part of is selecting the variables on which clustering is based.

Ans: d) formulating the clustering problem

Q 4. The most commonly used measure of similarity is the or its square.

Ans: a) Euclidean distance

Q 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.

Ans: b) Divisive clustering

Q 6. Which of the following is required by K-means clustering?

Ans: d) All answers are correct

Q 7. The goal of clustering is to?

Ans: a) Divide the data points into groups

Q 8. Clustering is a?

Ans: b) Unsupervised learning

Q 9. Which of the following clustering algorithms suffers from the problem of convergence at local optima?

Ans: a) K- Means clustering

Q 10. Which version of the clustering algorithm is most sensitive to outliers?

Ans: a) K-means clustering algorithm

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

Ans: d) All of the above

Q 12. For clustering, we do not require?

Ans: a) Labeled data

Q 13. How is cluster analysis calculated?

Ans: Calculate the distances, Link the cluster choose a solution by selecting the right number of cluster. first we have to select the variable upon which we base our clusters.

Q 14. How is cluster quality measured?

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

Q 15. What is cluster analysis and its types?

Ans: Cluster analysis is a multivariate data mining technique whose goal is to groups object (eg:-Products, respondent or other entities)

based on a set of used selected characteristics or attributes.Type of cluster Analysis :-

1. Hierarchical cluster
2. Centroid based clustering
3. Distribution based clustering
4. Density based clustering