Name: NALLA NANDAKISHORE

Batch: DS2302

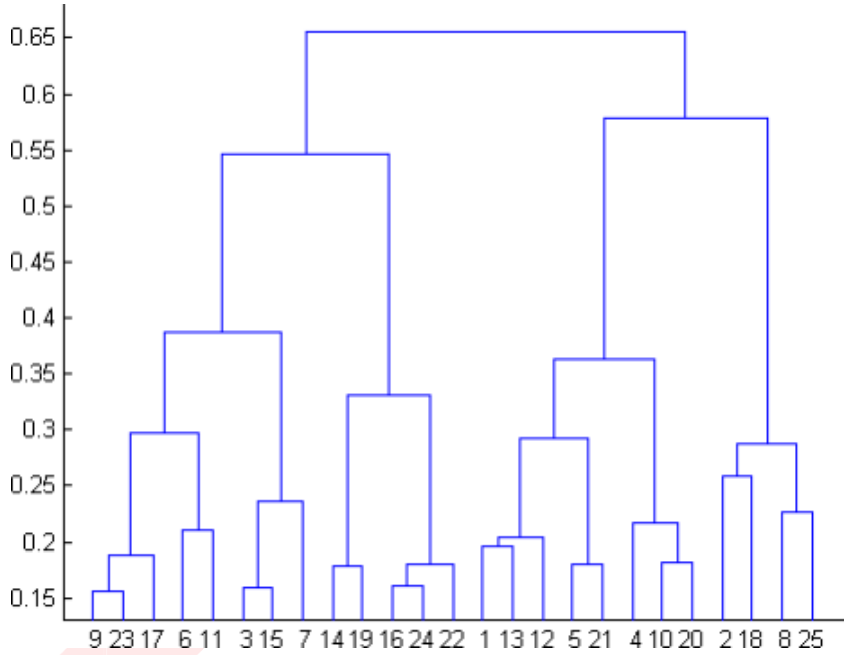
Institute: DataTrained

Internship: Machine Learning Worksheet Solutions

MACHINE LEARNING: ASSIGNMENT - I

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:



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:

d) 1, 2 and 4

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

d) Formulating the clustering problem

4. The most commonly used measure of similarity is the____or its square

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.

b) Divisive clustering

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

d) All answers are correct

7. The goal of clustering is to-

a) Divide the data points into groups

8. Clustering is a-

b) Unsupervised learning

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

d) All of the above

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

a) K-means clustering algorithm

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

d) All of the above

12.For clustering, we do not require-

a) Labelled data

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

13.How is cluster analysis calculated?

Ans : cluster analysis is casliculated by the square root of the sum of squared distances. And it is called eculidean distance

14.How is cluster quality measured?

Ans : discovering of hidden parts (all or some parts)is called high quality clustering. And is measured with high similar grouping

15.What is cluster analysis and its types?

Ans : In machine learning too, we often group examples as a first step to understand a subject (data set) in a machine learning system. Grouping unlabeled examples is called clustering. As the examples are unlabeled, clustering relies on unsupervised machine learning.