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

Ans: 4

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

Ans: d) 1, 2 and 4

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

Ans: a) interpreting and profiling clusters

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

Ans: 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.

Ans: c) Agglomerative clustering

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

Ans: d) All answers are correct

7. The goal of clustering is to

Ans : a) Divide the data points into groups

8. Clustering is a

Ans: b) Unsupervised learning

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

Ans : a) K-means clustering algorithm

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

Ans:a ) K-means clustering algorithm

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

Ans:d) ) All of the above

12. For clustering, we do not require

Ans : a) Labeled data

13. How is cluster analysis calculated:

Cluster is basically calculating the distance of data points, linking the cluster, and selecting right number of clusters

14. How is cluster quality measured?

To check the quality of cluster we use silhouette, In which we calculate a(i)- which is calculating the distance within cluster data points and take the average, calculate b(i)-selecting the closest

cluster and calculating the distance between nearby cluster's data point distance. Then we take mean of them .When the mean is towards to 1 which means is good model. Mean lies between -1 to 15. What is cluster analysis and its types:

Cluster analysis is basically a method, grouping the data set based on the distance of the data points or it will create groups with in the data set as clusters based on distance values.