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

Ans - c) 6

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

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 - d) All of the above

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 - cluster is a data analysis technique that find the naturally occurring groups within a data set known as clusters. Cluster analysis doesn't need to group data points into any predefined groups, which means that it is an unsupervised method.

We can use many mathodes to find cluster these are :-

Density - Based Clustering

DBSCAN (Density - Based Spatial Cluster of Applications with Noise)

HDBSCAN (Hierarchical Density - Based Spatial Cluster of Applications with Noise)

Hierarchical Clustering

Fuzzy Clustering

Grid – Based Clustering

Q 14. How is cluster quality measured?

Ans – After applying clustering method on data set measure the quality of result how good they are or not.

Methods are :-

a) Extrinsic
b) Intrinsic

Extrinsic – This method will help the other problems to be solved.

Represent images with cluster futures.

Train different classifier for each sub population

Identify and eliminate outliers / corrupted points.

Intrinsic – It will help itself to be solved.

Helps understand the makeup of your data (Quantative)

Clusters correspond to classes

Compare to human judgements

A good clustering method which produce high quality cluster with

a) High intra - class similarity inside the class or inter cluster distance is more.
b) Low inter – class similarity or intra cluster distance.

Q 15. What is cluster analysis and its types?

Ans – A statistical tool, cluster analysis is used to classify objects into groups where objects in one group are more similar to each other and different from objects in other groups. It is normally used for exploratory data analysis and as a method of discovery by solving classification issues.

Types of Clusters Analysis : -

centroid model
Distribution model

density model
connectivity model
K means clustring
Hierarchical Clustering