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

Ans: b:4 (Hierarchical Clustering)

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: d) formulating the clustering problem

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: b) Divisive 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

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

Ans: a) K-means clustering algorithm

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

Ans: b) Data points with different densities

12. For clustering, we do not require

Ans: a) Labeled data

13. How is cluster analysis calculated?

Ans: we have a millinions of data and data sorting a data to each group called cluster how group we need we have few technique so which group we create.when cluster creation the data is randomaly choose and create a seprated groups and this called centroids each centroids connect every centroids and divide and this divide data called centroids 1,centroids2 whichever centroids is create we called that and each cluster find the distance and calculate all datapoints and find the average of distance. after calulated average distance the centroids possition turn to different possintion and again calculate the distance of datapoints and find average of distance this process again and again when the centroids possition is not change and average distance is same.

14. How is cluster quality measured?

Ans: we have to average the coefficient the cluster value to to all object in data it means we

have to standardise data

15. What is cluster analysis and its types?

Ans: cluster analysis is basically is find out the similar data and create a seprate groups

For ex. Bank have a data so cluster analysis is creating a group which have low balance mid level balance and high level balance.and according to group we can target each group to grow revenue.

Types: 1.Centroid-based Clustering

2. Density-based Clustering.
3. Distribution-based Clustering.
4. Hierarchical Clustering.