*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*

d) All of the above

*8. Clustering is a*

b) Unsupervised learning

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

*optima?*

a) K- Means clustering

*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

*13. How is cluster analysis calculated?*

* First, an initial partition with k clusters (given number of clusters) is created.
* Then, starting with the first object in the first cluster, Euclidean distances of all objects to all cluster foci are calculated.
* If an object is detected whose distance to the centre of gravity of the own cluster is greater than the distance to the centre of gravity (centroid) of another cluster, this object is shifted to the other cluster.
* Finally, the centroids of the two changed clusters are calculated again, since the compositions have changed here.
* These steps are repeated until each object is located in a cluster with the smallest distance to its centroid (centre of the cluster) (optimal solution).

*14. How is cluster quality measured?*

The methods can be categorized into two groups according to whether condition is true.

**Extrinsic methods -** When the ground truth is available, we can use it to test the clustering by comparing it to a clustering. As a result, extrinsic methods' main purpose is to give a score.

extrinsic methods are also known as supervised methods

**Intrinsic methods -** If the ground truth isn't available, which evaluates a clustering's goodness by looking at how well the clusters are separated

intrinsic methods are unsupervised methods.

*15. What is cluster analysis and its types.*

Cluster analysis is the process of grouping a series of data points such that their relevance to one another can be determined. These methods produce clusters that help us understand how our data is connected.

**Connectivity models** - These models are based on the idea that data points that are closer together in data space are more identical to data points that are further apart. These models will take one of two directions. They begin by classifying all data points into different clusters and then aggregating them as the distance between them decreases in the first method.

The second method classifies all data points into a single cluster, which is then partitioned as the distance between them expands. Furthermore, the preference of distance function is a personal one.

These models are simple to understand, but they lack the scalability needed to handle large datasets.

**Centroid models** - These are iterative clustering algorithms in which similarity is determined by a data point's proximity to the cluster's centroid. The K-Means clustering algorithm is a common example of this kind of algorithm.

**Distribution models** - This clustering models are based on the premise that all data points in a cluster are likely to belong to the same distribution.

**Density Models -** This models look for locations in the data space with varying densities of data points. It isolates multiple density regions and groups the data points within these regions into clusters.