Machine Learning Dataset 1

Q1) What is the most appropriate no. of clusters for the data points represented by the followingdendrogram:

Ans : B -> 4

Q2) 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:

A. 1 and 2
B. 2 and 3
C. 2 and 4
D. 1, 2 and 4

Ans : D

Q3) The most important part of_____is selecting the variables on which clustering is based.

A. interpreting and profiling clusters
B. selecting a clustering procedure
C. assessing the validity of clustering
D. formulating the clustering problem

Ans : D

Q4) The most commonly used measure of similarity is the_____or its square.

A. Euclidean distance
B. city-block distance
C. Chebyshev's distance
D. Manhattan distance

Ans : A

Q5) _____ is a clustering procedure where all objects start out in one giant cluster. Clusters are formed bydividing this cluster into smaller and smaller clusters.

A. Non-hierarchical clustering
B. Divisive clustering
C. Agglomerative clustering

D. K-means clustering

Ans : C

Q6) Which of the following is required by K-means clustering?

A. Defined distance metric
B. Number of clusters
C. Initial guess as to cluster centroids
D. All answers are correct

Ans : D

Q7) The goal of clustering is to-

A. Divide the data points into groups
B. Classify the data point into different classes
C. Predict the output values of input data points
D. All of the above

Ans : A

Q8) Clustering is a-

A. Supervised learning
B. Unsupervised learning
C. Reinforcement learning
D. None

Ans : B

Q9) Which of the following clustering algorithms suffers from the problem of convergence at localoptima?

A. K- Means clustering
B. Hierarchical clustering
C. Diverse clustering
D. All of the above

Ans : A

Q10) Which version of the clustering algorithm is most sensitive to outliers?

A. K-means clustering algorithm

B. K-modes clustering algorithm
C. K-medians clustering algorithm
D. None

Ans : A

Q11) Which of the following is a bad characteristic of a dataset for clustering analysis-

A. Data points with outliers
B. Data points with different densities
C. Data points with non-convex shapes
D. All of the above

Ans : D

Q12) For clustering, we do not require-

A. Labeled data
B. Unlabeled data
C. Numerical data
D. Categorical data

Ans : A

Q13) How is cluster analysis calculated?

Ans : The cluster analysis calculator use the k-means algorithm :

First, choose the number of required clusters , take centroid and take each nearest points and calculate mean of all points . Then take each nearest points to this calculated mean and thus cluster forms of similar data / nearest data . Calculate SSE(sum of squared errors) . Repeat this process until this SSE doesn't go down.

Same process goes with every cluster.

Q14) How is cluster quality measured?

Ans : We can measure the quality of Clustering by using the Dissimilarity/Similarity metric in most situations. If all the data objects in the cluster are highly similar then the cluster has high quality.

Q15) What is cluster analysis and its types?

Ans : Cluster analysis is a statistical method used to group similar objects into respective categories by identifying similar trends and patterns.

There are 3 main types of clustering :

A. K-Means clustering
B. Hierarchical clustering
C. Density-Based clustering(DBSCAN)