MACHINE LEARNING

ASSIGNMENT ; 1

Ques 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: a) 1 and 2

b) 2 and

3 c) 2 and 4

d) 1, 2 and 4

Answer ; D

Ques 3.. 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

Answer… D

Ques4.. 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

Answer… A

Ques5.. ……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.

a) Non-hierarchical clustering

b) Divisive clustering

c) Agglomerative clustering

d) K-means clustering

Answer… C

Ques6… 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

Answer… D

Ques7….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

Answer…. A

Ques8… Clustering is a….

a) Supervised learning

b) Unsupervised learning

c) Reinforcement learning

d) None

Answer… B

Ques9…. Which of the following clustering algorithms suffers from the problem of convergence at local optima?

a) K- Means clustering

b) Hierarchical clustering

c) Diverse clustering

d) All of the above

Answer… A

Ques10….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

Answer… A

Ques11….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…

Answer… D

Ques12….. For clustering, we do not require

a) Labeled data

b) Unlabeled data

c) Numerical data

d) Categorical data

Answer… A

Ques13….How is cluster analysis calculated?

Answer… Clustering can be done by **reordering the rows or columns of the confusion matrix so that the sum of the diagonal values is maximal….**

Ques14…. How is cluster quality measured?

Answer…. It can be done by 3 methods are following;

1. **calculate the distances**

2. link the clusters

3. choose a solution by selecting the right number of clusters

Ques15…What is cluster analysis and its types?

Answer… it is related to the based on modals of distribution…

There are 4 types of cluster analysis are as follows…

1. Centroid-based density-based,

2. Distribution-based,

3. Hierarchical, constraint-based

4. Fuzzy clustering

QUESS1 DIAGRAM..

ANWER… B Means 4 number of cluster represented in dendrogram..