ASSIGNMENT -3

MACHINE LEARNING

1. Which of the following is an application of clustering?

a. biological network analysis
b. Market trend prediction
c. Topic modelling
d. All of the above

Answer: - d. All of the above

2. On which data type, we cannot perform cluster analysis?
a. Time series data
b. Text data
c. Multimedia data
d. None

Answer: - b. Text data

3. Netflix's movie recommendation system uses
a. Supervised learning
b. Unsupervised learning
c. Reinforcement learning and Unsupervised learning
d. All of the above

Answer: - a. Supervised learning

4. The final output of Hierarchical clustering is
a. The number of cluster centroids
b. The tree representing how close the data points are to each other
c. A map defining the similar data points into individual groups
d. All of the above

Answer: - b. The tree representing how close the data points are to each other

5. Which of the step is not required for K-means clustering?
a. A distance metric
b. Initial number of clusters
c. Initial guess as to cluster centroids
d. None

Answer: - d. None

6. Which is the following is wrong?
a. k-means clustering is a vector quantization method
b. k-means clustering tries to group n observations into k clusters
c. k-nearest neighbour is same as k-means

d. None

Answer: - c. k-nearest neighbour is same as k-means

7. Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?
i. Single-link
ii. Complete-link
iii. Average-link Options:
a.1 and 2
b. 1 and 3
c. 2 and 3
d. 1, 2 and 3

Answer: - d. 1, 2 and 3

8. Which of the following are true?
i. Clustering analysis is negatively affected by multicollinearity of features
ii. Clustering analysis is negatively affected by heteroscedasticity
Options:
a. 1 only
b. 2 only
c. 1 and 2
d. None of them

Answer:- a. 1 only

9. Answer : - a. 2

10. . For which of the following tasks might clustering be a suitable approach?
a. Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.
b. Given a database of information about your users, automatically group them into different market segments.
c. Predicting whether stock price of a company will increase tomorrow.
d. Given historical weather records, predict if tomorrow's weather will be sunny or rainy.

Answer: - a. Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.

11. Answer: -A

12. Answer: -C

13. What is the importance of clustering?

Answer:- Clustering is the process of dividing the entire dataset into groups or clusters. Clusters are nothing but collections of objects that are similar to each other. Clustering is an unsupervised technique. Clustering is a very popular technique that is used to cluster data points. Clustering is basically done on similarity measures to group similar data objects together. This similarity measure is based on distance functions such as Euclidean distance, Manhattan distance, Minkowski distance, Cosine similarity, etc. to group objects in clusters. It tells whether two data objects are alike or different based on the distance between those two points. If the distance between two data points is less, then we can say that they are similar to each other. If the distance is large, then they are less similar to each other.

Clustering is a very popular technique that is used to cluster data points. Clustering is basically done on similarity measures to group similar data objects together. This similarity measure is based on distance functions such as Euclidean distance, Manhattan distance, Minkowski distance, Cosine similarity, etc. to group objects in clusters. It tells whether two data objects are alike or different based on the distance between those two points. If the distance between two data points is less, then we can say that they are similar to each other. If the distance is large, then they are less similar to each other.

14. How can I improve my clustering performance?

Answer:- Clustering Performance can be improved by

- Applying Unsupervised learning
- K-Means clustering algorithm can some time give surprising result when
- As a pre-processing stage of data mining and machine learning, dimension reduction not only decreases computational complexity, but also significantly improves the accuracy of the learned models from large data sets