ASSIGNMENT – 3

MACHINE LEARNING

Q1 to Q12 have only one correct answer. Choose the correct option to answer your question.

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

a. Biological network analysis

b. Market trend prediction

c. Topic modeling

d. All of the above

Ans: 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

Ans: d) None

3. Netflix's movie recommendation system uses

a. Supervised learning

b. Unsupervised learning

c. Reinforcement learning and Unsupervised learning

d. All of the above

Ans: c) Reinforcement learning and Unsupervised 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

Ans: 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

Ans: 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

Ans: 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

Ans: 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

Ans a. 1 only

9. In the figure above, if you draw a horizontal line on y-axis for y=2. What will be the number of clusters formed?

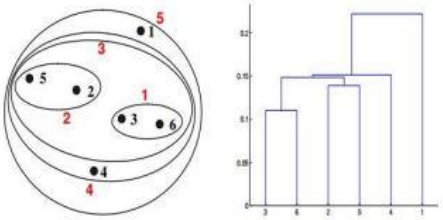
Ans : a) 2

10. For which of the following tasks might clustering be a suitable approach?

Ans : b) Given a database of information about your users, automatically group them into different market segments.

11. Given, six points with the following attributes:

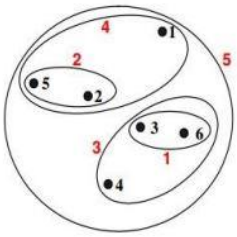
Which of the following clustering representations and dendrogram depicts the use of MIN or Single link proximity function in hierarchical clustering:

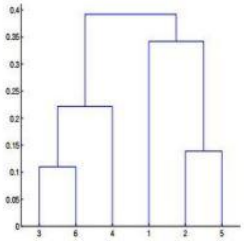


Ans :a)

12. Given, six points with the following attributes:

Which of the following clustering representations and dendrogram depicts the use of MAX or Complete link proximity function in hierarchical clustering.





Ans :b)

Q13 to Q14 are subjective answers type questions, Answers them in their own words briefly

13. What is the importance of clustering?

Ans :

Clustering in general term is a method of grouping a given data into clusters of data points with similar characteristics . Clustering technique helps in understanding natural grouping of data set. It helps to create logical partition between given data set, which in turn optimize the solution.

14. How can I improve my clustering performance?

Ans :

By selecting the proper method and understanding the hidden pattern ,clustering can be done in more significant way