MACHINE LEARNING ASSIGNMENT – 3

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

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

2. On which data type, we cannot perform cluster analysis?

a. Time series data

b. Text data

c. Multimedia data

d. None

Ans=d

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

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

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

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

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

Ans=D

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

Ans=a

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?

a. 2

b. 4

c. 3

d. 5

Ans=A

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.

Ans=B

11. Given, six points with the following attributes:

Ans=A

12.**Ans=B**

13. What is the importance of clustering?

Ans=Clustering is important in data analysis and data mining applications. It is the task of grouping a set of objects so that objects in the same group are more similar to each other than to those in other groups (clusters). Clustering can be done by the different numbers.To group items that might have same attributes together. It might be helpful to imagine that you have millions of chemical compounds that you cannot see and judge what they are trying to tell, what is similar among them. By clustering you will group those millions of clusters in lets say 5 or 10 clusters based on some similarity among them making it easier for you to analyze those 5 or 10 clusters rather than seeing each compound individually.

14. How can I improve my clustering performance?

Ans=K-means clustering algorithm can be significantly improved by using a better initialization technique, and by repeating (re-starting) the algorithm. When the data has overlapping clusters, k-means can improve the results of the initialization technique.

Graph-based clustering performance can easily be improved by applying ICA blind source separation during the graph Laplacian embedding step. Applying unsupervised feature learning to input data using either RICA or SFT, improves clustering performance.