Worksheet-3

Machine learning

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

Answer- D) All of the above

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

a. Time series data b. Text data

c. Multimedia data d. None

Answer- D) None

Question-3 . Netflix's movie recommendation system usesa.

Supervised learning b. Unsupervised learning

c. Reinforcement learning and Unsupervised learning d. All of the above

Answer- C) Reinforcement learning and Unsupervised learning

Question-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

Question-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

Question-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

Question-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

Question – 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

Question-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

Answer- A) 2

Question-10

Answer-B) Given a database of information about your users, automatically group them into different market segments.

Question 11- Given, six points with the following attributes

Answer- a

Question 12- Which of the following clustering representations and dendrogram depicts the use of MIN or Single link proximity function in hierarchical clustering:

Answer- b

Question 13- What is the importance of clustering?

Answer- Having clustering methods helps in restarting the local search procedure and remove the inefficiency. In addition, clustering helps to determine the internal structure of the data.

This clustering analysis has been used for model analysis, vector region of attraction.

Clustering helps in understanding the natural grouping in a dataset. Their purpose is to make sense to partition the data into some group of logical groupings.

Clustering quality depends on the methods and the identification of hidden patterns.

They play a wide role in applications like marketing economic research and weblogs to identify similarity measures, Image processing, and spatial research.

They are used in outlier detections to detect credit card fraudulence.

Question – 14 How can I improve my clustering performance?

Answer- Applying unsupervised feature learning to input data using either RICA or SFT, improves clustering performance. Surprisingly for some cases, high clustering performance can be achieved by simply performing K-means clustering on the ICA components after PCA dimension reduction on the input data.