MACHINE LEARNING

ASSIGNMENT – 3

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?

ANS. D. All of the above

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

ANS. D. None

3. Netflix's movie recommendation system usesa. Supervised learning

ANS. C. Reinforcement learning and Unsupervised learning

4. The final output of Hierarchical clustering isa. The number of cluster centroids

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?

ANS. D. None

6. Which is the following is wrong?

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

Options:

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

Options:

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. D. 5

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

13. What is the importance of clustering?

ANS. In data analysis and data mining applications, clustering is crucial. The goal is to group the objects in a set so that they are more similar to one another than to the objects in other groups (clusters).

14. How can I improve my clustering performance?

Ans. Performing, ICA BSS after the underlying grid factorization step gave the greatest bunching execution in four out of six datasets (COIL100, CMU-PIE, MNIST, and REUTERS-10K). Applying UFL as an underlying handling part assisted with giving the most extreme execution in three out of six datasets (USPS, COIL20, and COIL100). Contrasted with cutting edge non-profound picking up bunching strategies, ICA BSS and additionally UFL with chart based grouping calculations beat any remaining techniques. As for profound learning-based bunching calculations, the new strategy introduced here got the accompanying rankings: COIL20, second out of 5; COIL100, second out of 5; CMU-PIE, second out of 5; USPS, third out of 9; MNIST, eighth out of 15; and REUTERS-10K, fourth out of 5.

By utilizing just ICA BSS and UFL utilizing RICA and SFT, grouping exactness that is better or comparable to many profound learning-based bunching calculations was accomplished. For example, by applying ICA BSS to ghostly bunching on the MNIST dataset, we acquired a precision of 0.882. This is superior to the notable Profound Implanted Grouping calculation that had gotten an exactness of 0.818 involving stacked denoising autoencoders in its model.