**MACHINE LEARNING WORKSHEET 3**

Q1. Which of the following is an application of clustering?

Ans: D) None of the Above

Q2. On which data type, we cannot perform cluster analysis?

Ans: D) None

Q3. Netflix’s movie recommendation system uses-

Ans: C) Reinforcement learning and Unsupervised learning

Q4. The final output of Hierarchical clustering is

Ans: B) The tree representing how close the data points are to each other

Q5. Which of the step is not required for K-means clustering?

Ans: D) None

Q6. Which is the following is wrong?

Ans: C) k-nearest neighbour is same as k-means

Q7. Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?

Ans: C) 2 and 3

Q8. Which of the following are true?

Ans: A) 1 Only

Q9. 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

Q10. 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.

Q11. 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)

Q12. 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. What is the importance of clustering?

Ans: Clustering is a Machine Learning technique where, the way of grouping the data points into different clusters, consisting of similar data points. The objects with the possible similarities remain in a group that has less or no similarities with another group**.**

Importance of clustering:

1. It is normally used for exploratory data analysis and as a method of discovery by solving classification issues.
2. Clustering helps in understanding the natural grouping in a dataset.
3. Clustering helps to determine the internal structure of the data.
4. It is used in outlier detection to detect credit fraudulence

Q14. 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

**The End**