MACHINE LEARNING

WORKSHEET SET 3

SUBMITTED BY : SHAHLA M

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

Answer – (d) All of the above

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

Answer – (d) None

3. Netflix's movie recommendation system uses?

Answer – (c) Reinforcement learning and Unsupervised learning

4. The final output of Hierarchical clustering is?

Answer – (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?

Answer – (d) None

6. Which is the following is wrong?

Answer – (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

Answer – (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

Answer – (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?

Answer – (a) 2

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

Answer – (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?

Answer – (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.

Answer – (b)

13. What is the importance of clustering?

Clustering in machine learning is an essential component and makes life so much easier in creating new machine learning methods. It mainly **divides many unstructured data sets into clusters** and, according to the common attributes present in them, it helps create more and more clusters.

The two main "techniques" for developing a machine

learning algorithm are as follows:

• Supervised learning method

• Un-supervised learning method

There are two different types of clustering, which are hierarchical and non-hierarchical methods. Non-hierarchical Clustering In this method, the dataset containing N objects is divided into M clusters. In business intelligence, the most widely used non-hierarchical clustering technique is K-means.

14. How can I improve my clustering performance?

Answer – Improving clustering by focusing on tasks that are subject- and domain-specific. However, there are some actions we could do to guarantee improved performance. First of all, the outcomes are always better the more data there is. Additionally, it should go without saying that data should be as clean as possible before being used for analysis because "garbage in, trash out." Choosing the right clustering algorithm is crucial when conducting a cluster analysis. Another crucial step in the clustering process is selecting the ideal number of clusters.

 We should be very careful while selecting the algorithms for determining the ideal cluster size. such as the silhouette approach, gap statistic method, and elbow method). Dimension reduction could be one of the optimization techniques.