Statistics Worksheet 3

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

Ans : All of the above

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

Ans : None

3) Netflix's movie recommendation system uses-

Ans : Reinforcement learning and Unsupervised learning

4) The final output of Hierarchical clustering is-

Ans : 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 : None

6) Which is the following is wrong?

Ans : 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?

Ans : 1,2&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

Ans : 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 : 2

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

Ans : Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.

11) Given, six points with the following attributes:

Ans : option 1

12) Which of the following clustering representations and dendrogram depicts the use of MAX or Complete link proximity function in hierarchical clustering

Ans : option b

13) What is the importance of clustering?

Ans : Clustering methods (like Hierarchical method, Partitioning, Density-based method, Model-based clustering, and Grid-based model) help in grouping the data points into clusters, using the different techniques are used to pick the appropriate result for the problem, these clustering techniques helps in grouping the data points into similar categories, and each of these subcategories is further divided into subcategories to assist the exploration of the queries output. Importance of clustering are as below: 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

14) How can I improve my clustering performance?