**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?
**d. All of the above**

2. On which data type, we cannot perform cluster analysis?
**d. None**

3. Netflix's movie recommendation system uses-
**c. Reinforcement learning**

4. The final output of Hierarchical clustering is-
**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?
**d. None**

6. Which is the following is wrong?
**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

**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

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?
**a. 2**

10. For which of the following tasks might clustering be a suitable approach?
a. Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.
b. Given a database of information about your users, automatically group them into different market segments.
c. Predicting whether stock price of a company will increase tomorrow.
d. Given historical weather records, predict if tomorrow's weather will be sunny or rainy.

11.Which of the following clustering representations and dendrogram depicts the use of MIN or Single link proximity function in hierarchical clustering:

**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.
**b.**

**Q13 to Q14 are subjective answers type questions, Answers them in their own words briefly**

13. What is the importance of clustering?

1. 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.
2. This clustering analysis has been used for model analysis, vector region of attraction.
3. 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.
4. Clustering quality depends on the methods and the identification of hidden patterns.
5. They play a wide role in applications like marketing economic research and weblogs to identify similarity measures, Image processing, and spatial research.
6. They are used in outlier detections to detect credit card fraudulence.

14. How can I improve my clustering performance?

Clustering performance can be improved by k-means method. k-means is a very simple and ubiquitous clustering algorithm. But quite often it does not work on your problem, for example because the initialization is bad. I ran into a similar problem recently, where I applied k-means to a smaller number of files in my data sets and everything worked fine, but when I ran it on many more samples it just wasn’t reliably getting good results.