Worksheet-3 Machine Learning

1. Which of the following is an application of clustering?
   d. All the above

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

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

Options:
   a. 1 only

9. In the above figure, if you draw 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?
   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 dendrograms depicts the use of MAX or complete link proximity function in hierarchical clustering.

Ans: b

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

13. What is the importance of clustering?

Ans:

1. Having clustering methods helps in restarting the local search procedure and remove the inefficiency. In addition, clustering helps to determine the internal structure 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 identification of hidden patterns.
5. The 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?

Ans:

Clustering performance can be improved by-

1. Merging neighbouring clusters if the resulting cluster's variance is below the threshold.
2. Isolating elements that are "far" if a cluster's variance is above the threshold.
3. Moving some elements that are between neighbouring clusters if it decreases the sum of squared errors. (This evolution acts as a global optimization procedure, and prevents the bad consequences of initial assignment of cluster means in k-means)