ASSIGNMENT – 2

MACHINE LEARNING

Q1 to Q11 have only one correct answer. Choose the correct option to answer your question.

1. Movie Recommendation systems are an example of:

i) Classification

ii) Clustering

iii) Regression

Options: b) 1 and 2 , is correct.

2. Sentiment Analysis is an example of:

i) Regression

ii) Classification

iii) Clustering

iv) Reinforcement

Options: d) 1, 2 and 4 , is correct.

3. Can decision trees be used for performing clustering?

a) True

4. Which of the following is the most appropriate strategy for data cleaning before performing clustering analysis, given less than desirable number of data points:

i) Capping and flooring of variables

ii) Removal of outliers

Options: a) 1 only , is correct

5. What is the minimum no. of variables/ features required to perform clustering?

Options: b) 1 , is correct.

6. For two runs of K-Mean clustering is it expected to get same clustering results?

Options: b) No , is correct.

ASSIGNMENT – 2

MACHINE LEARNING

7. Is it possible that Assignment of observations to clusters does not change between successive

iterations in K-Means?

Ans :- Options: a) Yes , is correct.

8. Which of the following can act as possible termination conditions in K-Means?

i) For a fixed number of iterations.

ii) Assignment of observations to clusters does not change between iterations. Except for cases

witha bad local minimum.

iii) Centroids do not change between successive iterations.

iv) Terminate when RSS falls below a threshold.

Ans : - Options: d) All of the above , is correct.

9. Which of the following algorithms is most sensitive to outliers?

Ans : - Options: a) K-means clustering algorithm ,is correct.

10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression

Ans : - Options: d) All of the above , is correct

11. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset?

Ans : - Options: d) All of the above

Q12 to Q14 are subjective answers type questions, Answers them in their own words briefly

12. Is K sensitive to outliers?

Ans : - **The K-means clustering algorithm is sensitive to outliers**, because a mean is easily influenced by extreme values. K-medoids clustering is a variant of K-means that is more robust to noises and outliers.

13. Why is K means better?

Ans : - **Advantages of k-means –**

- Can warm-start the positions of centroids.
- Easily adapts to new examples.
- Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

14. Is K means a deterministic algorithm?

Ans : - One of the significant drawbacks of K-Means is its non-deterministic nature. K-Means starts with a random set of data points as initial centroids. This random selection influences the quality of the resulting clusters. Besides, each run of the algorithm for the same dataset may yield a different output.