MACHINE LEARNING

1. Movie Recommendation systems are an example of: i) Classification ii) Clustering iii) Regression

Ans. 2 and 3

2. Sentiment Analysis is an example of: i) Regression ii) Classification iii) Clustering iv) Reinforcement

Ans. 1, 2 and 4

3. Can decision trees be used for performing clustering? a) True b) False

Ans. 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

Ans. 1 only

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

Ans. 1

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

Ans. No

7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?

Ans. Yes

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. All of the above

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

Ans. K-means clustering algorithm

10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning): i) Creating different models for different cluster groups. ii) Creating an input feature for cluster ids as an ordinal variable. iii) Creating an input feature for cluster centroids as a continuous variable. iv) Creating an input feature for cluster size as a continuous variable.

Ans. All of the above

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

Ans. All of the above

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

Ans. All of the above

13. Is K sensitive to outliers?

Ans. K-Means clustering algorithm is most sensitive to outliers as it uses the mean of cluster data points to find the cluster center.

14. Why is K means better?

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

15. 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