ASSIGNMENT – 2

MACHINE LEARNING

1. Movie Recommendation systems are an example of:
Ans. b) 1 and 2

2. Sentiment Analysis is an example of:
Ans. d) 1, 2 and 4

3. Can decision trees be used for performing clustering?
Ans. 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:
Ans. a) 1 only

5. What is the minimum no. of variables/ features required to perform clustering?
Ans. b) 1

6. For two runs of K-Mean clustering is it expected to get same clustering results?
Ans. b) No

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

8. Which of the following can act as possible termination conditions in K-Means?
Ans. d) All of the above

9. Which of the following algorithms is most sensitive to outliers?
Ans. a) K-means clustering algorithm

10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):
Ans. d) 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. d) All of the above

12. Is K sensitive to outliers?
Ans. The *k*-means algorithm is sensitive to the outliers. In this paper, we propose a robust two-stage *k*-means clustering algorithm based on the observation point mechanism, which can accurately discover the cluster centers without the disturbance of outliers.

13. Why is K means better?

Ans. Guarantees convergence. 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. The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids. Method: We propose an improved, density based version of K-Means, which involves a novel and systematic method for selecting initial centroids.