**FLIP ROBO ASSIGNMENT**

**WORKSHEET 2**

**MACHINE LEARNING**

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

Ans- 2 and 3(D)

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

Ans- 1, 2 and 4(D)

3. Can decision trees be used for performing clustering?

Ans- True(A)

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:

Ans- 1 only(A)

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

Ans- 1 (B)

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

Ans- No(B)

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

Ans- Yes (A)

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. Options:

Ans- All of the above(D)

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

Ans- K-medoids clustering algorithm (D)

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. Options:

Ans- All of the above(D)

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(D)

12. Is K sensitive to outliers?

Ans- The *k*-means algorithm is sensitive to the outliers. Unsupervised learning algorithm such as K-Means clustering divides n observations into k clusters which each observation belonging to the cluster with the nearest centroid. The algorithm aims to minimize the squared Euclidean distances between the observation and the centroid of cluster to which it belongs.

13. Why is K means better?

Ans- K-Means for Clustering is one of the popular algorithms for the approach where K means the number of clustering and means implies the statistics mean a problem.  K-means is very simple and easy to implement. However, it is unlikely to be the state-of-the-art, but for straightforward clustering, it is also a part of a larger data-processing pipeline and is a reasonable default choice, at least until you figure out that the clustering step is your bottleneck in terms of overall performance. For example, clustering can be applied to MP3 files, cellular phones are the general areas that use this technique.

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.