

## MACHINE LEARNING ASSIGNMENT 2

1. Movie Recommendation systems are an example of:

Ans: Option A: 2 only

2) Sentiment Analysis is an example of:

Ans: Option D: 1,2 and 4

3) Can decision trees be used for performing clustering?

Ans: Option 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: Option A: 1only

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

Ans:Option B: 1

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

Ans:Option B: No

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

Ans: Option A: Yes

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

Ans: Option D: All of the above

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

Ans: Option A: K-means clustering algorithm

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

Ans:Option 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: Option D: All of the above

12. Is K sensitive to outliers?

Ans:Yes, K is sensitive to outliers as it uses the mean of cluster data points to find the cluster centre. Suppose there are outliers present then the cluster size changes and KMeans does not behave very well when the clusters have varying sizes, different densities and non spherical shapes.

13. Why is K means better?

Ans:K means is simple unsupervised algorithm and it works well with large datasets.It is also simple and very easy to implement.K means also produces better clusters than hierarchical clustering.

14. Is K means a deterministic algorithm?

Ans: No,K Means is a non deterministic algorithm. This means that running the algorithm several times on the same data could give different results.