**MACHINE LEARNING ASSIGNMENT-2**

1. Movie Recommendation systems are an example of:

i) Classification

ii) Clustering

iii) Regression

Options:

a) 2 Only

b) 1 and 2

c) 1 and 3

d) 2 and 3

**Ans**—a) 2 only

2. Sentiment Analysis is an example of:

i) Regression

ii) Classification

iii) Clustering

iv) Reinforcement

Options:

a) 1 Only

b) 1 and 2

c) 1 and 3

d) 1, 2 and 4

**Ans**—b) 1 and 2

3. Can decision trees be used for performing clustering?

a) True

b) False

**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:

i) Capping and flooring of variables

ii) Removal of outliers

Options:

a) 1 only

b) 2 only

c) 1 and 2

d) None of the above

**Ans**-- a) 1 only

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

a) 0

b) 1

c) 2

d) 3

**Ans**-- b) 1

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

a) Yes

b) No

**Ans**-- b) No

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

a) Yes

b) No

c) Can't say

d) None of these

**Ans**-- a) 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 with a bad local minimum.

iii) Centroids do not change between successive iterations.

iv) Terminate when RSS falls below a threshold.

Options:

a) 1, 3 and 4

b) 1, 2 and 3

c) 1, 2 and 4

d) All of the above

**Ans**-- d) All of the above

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

i) K- Means clustering algorithm

ii) Agglomerative clustering algorithm

iii) Expectation-Maximization clustering algorithm

iv) Diverse clustering algorithm

Options:

a) 1 only

b) 2 and 3

c) 2 and 4

d) 1 and 3

**Ans**-- d) 1 and 3

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

a) K-means clustering algorithm

b) K-medians clustering algorithm

c) K-modes clustering algorithm

d) K-medoids clustering algorithm

**Ans**-- a) K-means clustering algorithm

11. 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:

a) 1 only

b) 2 only

c) 3 and 4

d) All of the above

**Ans**-- d) 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?

a) Proximity function used

b) of data points used

c) of variables used

d) All of the above

**Ans**-- d) All of the above

**Q13 to Q15 are subjective answers type questions, Answers them in their own words briefly**

13. Is K sensitive to outliers?

**Ans**—K –means clustering algorithm are sensitive to outliers because mean is easily influenced by extreme values. Mean is greatly influenced by outliers and thus cannot represent correct cluster center. While medoid is robust to outliers and correctly represent the cluster center.

14. Why is K means better?

**Ans**—K-means are relatively simple to implement. It scales to large datasets and also guarantees convergence. It easily adapts to new examples and can warm start the position of centroids.It also gives good results.

15. Is K means a deterministic algorithm?

**Ans**-K-means is based on a non-deterministic algorithm. The non deterministic nature of k-means is due to its random selection of datapoints as initial centroids.