

MACHINE LEARNING

1. Movie Recommendation systems are an example of:

- i) Classification
- ii) Clustering
- iii) Regression

Ans: d) 2 and 3

2. Sentiment Analysis is an example of:

- i) Regression
- ii) Classification
- iii) Clustering
- iv) Reinforcement

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:

- i) Capping and flooring of variables
- ii) Removal of outliers

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?

- i) For a fixed number of iterations.
- ii) Assignment of observations to clusters does not change between iterations. Except for cases with bad local minimum.
- iii) Centroids do not change between successive iterations.
- iv) Terminate when RSS falls below a threshold.

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):

- 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: 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 clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values. K-medoids clustering is a variant of K-means that is more robust to noises and outliers.

Instead of using the mean point as the centre of a cluster, *K-medoids* uses an actual point in the cluster to represent it. Medoid is the most centrally located object of the cluster, with minimum sum of distances to other points.

13. Why is K means better?

Ans: K-Means for Clustering is one of the popular algorithms for this approach. Where K means the number of clustering and means implies the statistics mean a problem. It is used to calculate code-vectors. K means algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible

14. Is K means a deterministic algorithm?

Ans: K-Means has many drawbacks too. 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.

