

Machine Learning

In Q1 to Q11, only one option is correct, choose the correct option:

1. Movie Recommendation systems are an example of:

- i. Classification
- ii. Clustering
- iii. Regression

a) 2 Only

b) 1 and 2

c) 1 and 3

d) 2 and 3

Answer – a) 2 Only

2. Sentiment Analysis is an example of:

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

a) 1 Only

b) 1 and 2

c) 1 and 3

d) 1, 2 and 4

Answer – d) 1, 2 and 4

3. Can decision trees be used for performing clustering?

a) True

b) False

Answer – 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

a) 1 Only

b) 2 Only

c) 1 and 2

d) None of the above

Answer – a) 1 Only

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

a) 0

b) 1

c) 2

d) 3

Answer – b) 1

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

a) Yes

b) No

Answer – 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

Answer – 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.

a) 1, 3 and 4

b) 1, 2 and 3

c) 1, 2 and 4

d) All of the above

Answer – d) All of the above

9. 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

Answer – 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.

a) 1 Only

b) 2 Only

c) 3 and 4

d) All of the above

Answer – 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?

- a) Proximity function used
- b) of data points used
- c) of variables used
- d) All of the above

Answer – d) All of the above

12. Is K sensitive to outliers?

Yes, K means is sensitive to outliers. As it averages data points nearer to data point to find value of dependent variable.

Consider four set of values on x-y graph, such as $P_1(6, 5)$, $P_2(3, 4)$, $P_3(7, 19)$, $P_4(5, 5)$

If we plot above point and predict, for $x=3.5$, with $k=4$, our result is $(5.25, 8.25)$ lies far away from actual result is clearly and outlier. So, we can say that it is sensitive to outliers.

13. Why is K means better?

K means is better, because it takes arithmetic mean of data points in plot to find value of dependent variable. Also, it is very simple to understand and implement and can be used to large data sets. It easily generalizes to clusters of different shapes and sizes. Adding to these points, it also improves models' accuracy with less computational cost.

14. Is K means a deterministic algorithm?

It is non-deterministic algorithm. As we run this algorithm, it always gives different result. For two runs of K-Mean clustering we get different clustering results.