MACHINE LEARNING

Q1 to Q11 have only one correct answer. Choose the correct option to answer your question.

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 : - b) 1 and 2

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:- d) 1, 2 and 4

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

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

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

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

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:
a) 1 only
b) 2 only
c) 3 and 4
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

Q12 to Q14 are subjective answers type questions, Answers them in their own words briefly

12. Is K sensitive to outliers?
Ans :- Yes K-Means algorithm is sensitive to outliers. The K-Means algorithm updates the cluster centers by taking the average of all the data points that are closer to each cluster center. When all the data points are packed nicely together, the average makes sense. However,when you have outliers, thi can affect the average calculation of the whole cluster. As a result, this will push your cluster center closer to the outlier.
So this makes K-Means sensitive to outliers.

13. Why is K means better?

Ans:- K-means is better to use cause first of all it is simple to implement. It can be used on larger data sets. By choosing the value of K we can simply form clusters and use K-means unsupervised algorithm.

Also when it comes to new data points that means when data get changed because of new examples it easily adapts the changes and means change centroid in the clusters.

K-means guarantees convergence and generalizes to clusters of different shapes and sizes, such as elliptical clusters.

14. Is K means a deterministic algorithm

Ans :- Deterministic approach of algorithm means for a given particular input, the computer will produce same output going through different states but in the case of non-deterministic algorithm, for the same input, the compiler may produce different output in different runs.

Clustering algorithms with steps involving randomness usually give different results on different executions for the same dataset. Therefore K -Means follows a non deterministic approach . The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids.