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
   ### Answer (a)2 only :Clustering

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
   ### Answer (d): 1,2 and 4 : Regression ,Classification and Reinforcement

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
   Options:
   a) 1 only
   b) 2 only
   c) 1 and 2
   d) None of the above
   ### Answer (a): Capping and flooring of variables

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

MACHINE LEARNING

c) Can't say
d) None of these

Answer(a) Yes

MACHINE LEARNING

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 witha 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
   ### 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.
       Options:
    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

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

12. Is K sensitive to outliers?

Answer: Yes k is sensitive to outliers .Because the mean , as a statistic is generally sensitive to outliers. Take for an example

The mean of 2,2,2,3,3,3,4,4,4 is 3

If we add a single 23 to that , the mean becomes 5 . which is larger than any of the other values.

Since in k-means , we will be taking the mean a lot, winding up with lot of outlier-sensitive calculations.

MACHINE LEARNING

13. Why is K means better?

Answer: **Relatively simple to implement, Scales to large data sets, Guarantees convergence, easily adapt to new examples and also generalize to clusters of different shapes and sizes .**

14. Is K means a deterministic algorithm?

Answer: **K-Means is non deterministic nature. It starts with a random set of data points as initial centroids .This random selection influences the quality of the resulting clusters**