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**:**

ANS. d) 2 and 3

2. Sentiment Analysis is an example of:

i) Regression

ii) Classification

iii) Clustering

iv) Reinforcement

Options:

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

Options:

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?

MACHINE LEARNING

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

witha bad local minimum.

iii) Centroids do not change between successive iterations.

iv) Terminate when RSS falls below a threshold.

Options:

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.

Options:

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

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

12. Is K sensitive to outliers?

ANS. K means is sensitive to outliers.

For e.g. Data set point are 1 2 3 7 8 80

Now 80 is outlier.

K=2

C1=1 C2=7

After first iteration

C1=2 C2=31.67

As 80 data point which is outlier comes in cluster 2.

Cluster 2 centroid changes to accommodate 80 .

Therefore K means is sensitive to outliers

13. Why is K means better?

ANS. all it truly does is figure the distances among focuses and bunch focuses, bringing about a straight intricacy O(n). This turns out entirely great with most datasets where you're not handling a huge number of pieces of information.

14. Is K means a deterministic algorithm?

ANS. The fundamental k-means grouping depends on a non-deterministic calculation. This implies that running the calculation a few times on similar information, could give various outcomes.