MACHINE LEARNING

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?
Ans: 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

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

a) 0

b) 1
c) 2
d) 3
Ans: b)1

6. For two runs of K-Mean clustering is it expected to get same clustering results?
a) Yes
b) No

Ans: 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

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:
a) 1, 3 and 4
b) 1, 2 and 3
c) 1, 2 and 4
d) All of the above
Ans: 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

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:
a) 1 only
b) 2 only
c) 3 and 4
d) All of the above

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

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?
13. Why is K means better?
14. Is K means a deterministic algorithm?

Ans: K-Means clustering algorithm is most sensitive to outliers as it uses the mean of cluster data points to find the cluster centre

K Nearest Neighbor(KNN) is a very simple, easy to understand, versatile and one of the topmost machine learning algorithms. KNN used in the variety of applications such as finance, healthcare, political science, handwriting detection, image recognition and video recognition. In Credit ratings, financial institutes will predict the credit rating of customers. In loan disbursement, banking institutes will predict whether the loan is safe or risky. In political science, classifying potential voters in two classes will vote or won't vote. KNN algorithm used for both classification and regression problems. KNN algorithm based on feature similarity approach.

In KNN, K is the number of nearest neighbors. The number of neighbors is the core deciding factor. K is generally an odd number if the number of classes is 2. When K=1, then the algorithm is known as the nearest neighbor algorithm. This is the simplest case. Suppose P1 is the point, for which label needs to predict. First, you find the one closest point to P1 and then the label of the nearest point assigned to P1.