MACHINE LEARNING (With Answers)

Q1 to Q11

1. Movie Recommendation systems are an example of:

i) Classification
ii) Clustering
Iii) Regression

Ans - b) 1 and 2

2. Sentiment Analysis is an example of:
i) Regression
ii) Classification
iii) Clustering
iv) Reinforcement

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

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 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

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

12. Is K sensitive to outliers?

Ans - K-Means clustering is an unsupervised learning algorithm which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest centroid. The algorithm aims to minimize the squared Euclidean distances between the observation and the centroid of cluster to which it belongs.

 An outlier is a point which is different from the rest of data points. K-means clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values or outliers. So K-Means algorithm does not give best results with outliers.

13. Why is K means better?

Ans - K-means clustering is a very famous and powerful unsupervised machine learning algorithm. It is used to solve many complex unsupervised machine learning problems.

Advantages of k-means:
Relatively simple to implement.

Scales to large data sets.

Guarantees convergence.

Can warm-start the positions of centroids.

Easily adapts to new examples.

Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

14. Is K means a deterministic algorithm?

Ans - The basic k-means clustering is based on a **non-deterministic algorithm**. This means that running the algorithm several times on the same data, could give different results.

THE END