Machine Learning

1. Movie Recommendation systems are an example of**:** Clustering
2. Sentiment Analysis is an example of: **Regression, Classification, Reinforcement Learning**
3. Can decision trees be used for performing clustering? **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?

**Capping and flouring of variables**

1. What is the minimum no. of variables/ features required to perform clustering? **Solution: (B) 1**
2. For two runs of K-Mean clustering is it expected to get same clustering results? **Solution: (B) No**
3. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means? **Solution: (A) Yes**
4. . 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.? **Solution: (D) All of d above**
5. 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. **Solution: (D)  1 and 3**
6. 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 **Solution: (A)**
7. 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 **Solution: (D) All above**

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

12 -Is K sensitive to outliers?

**The K-means clustering algorithm is sensitive to outliers**, because a mean is easily influenced by extreme values. K-medoids clustering is a variant of K-means that is more robust to noises and outliers.It is most sensitive to outliers as it uses the mean of cluster data points to find the cluster center.

13. Why is K means better? Better reason being **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?

of K-Means is its **non-deterministic nature**. K-Means starts with a random set of data points as initial centroids. This random selection influences the quality of the resulting clusters. Besides, each run of the algorithm for the same dataset may yield a different output

WORKSHEET 2 SQL

1. Which of the following constraint requires that there should not be duplicate entries? Unique
2. . Which of the following constraint allows null values in a column? Null
3. Which of the following statements are true regarding Primary Key? Each entry in the primary key uniquely identifies each entry or row in the table

4 Each entry in the primary key uniquely identifies each entry or row in the table? D) All of the above

5 . Which of the following is/are example of referential constraint? B) Foreign Key

6 How many foreign keys are there in the Supplier table? Answer 2

7. How many foreign keys are there in the Supplier table? Answer one to one

STATISTICS WORKSHEET-2

1 What represent a population parameter? Answers -Mean

2. What will be median of following set of scores (18,6,12,10,15)? Answers 12

3. What is standard deviation? Answer All of d above

4. The intervals should be \_\_\_\_\_\_ in a grouped frequency distribution. Answers Mutually exclusive

5. What is the goal of descriptive statistics? Answer What is the goal of descriptive statistics?

6. A set of data organized in a participant by variables format is called? Answer-Data Set

7. n multiple regression,\_\_\_\_\_\_\_ independent variables are used? Answer 1

8. Which of the following is used when you want to visually examine the relationship between 2 quantitative variables? Scatterplot

9. Two or more groups means are compared by using ? Analysis of variance

10. \_\_\_\_\_\_\_is a raw score which has been transformed into standard deviation units? Answer Z score

11. \_\_\_\_\_\_\_is the value calculated when you want the arithmetic average? Answer Mean

12. Find the mean of these set of number (4,6,7,9,2000000)? 400005.2

13. . \_\_\_\_\_\_\_ is a measure of central tendency that takes into account the magnitude of scores? Ans.Mean

14. \_\_\_\_\_\_ focuses on describing or explaining data whereas \_\_\_\_\_\_involves going beyond immediate? Answers Descriptive and inferences’

15. . What is the formula for range? Answers -H-L