<u>**Worksheet 2**</u>

1. <u>**Machine Learning Assignment 2**</u>

Q1. Movie Recommendation systems are an example of:
Ans. © 1 and 3

Q2. Sentiment Analysis is an example of:
Ans. (d) 1, 2 and 4

Q3. Can decision trees be used for performing clustering?
Ans. (a) True

Q4. 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
Ans. (a) 1 only

Q5. What is the minimum no. of variables/ features required to perform clustering?
Ans. (b) 1

Q6. For two runs of K-Mean clustering is it expected to get same clustering results?
Ans. (a) Yes

Q7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?
Ans. (a) Yes

Q8. 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.
Ans. (d) All of the above

Q9. Which of the following algorithms is most sensitive to outliers?
Ans. (a) K-means clustering algorithm

Q10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):
Ans. (a) 1 only

Q11. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset?
Ans. (a) All of the above

Q12. Is K sensitive to outliers?
Ans. Since mean is generally infleucned by the outliers due to extreme values. Hence K is very sensitive to the outliers.

Q13. Why is K means better?
Ans. It guarantees the convergence in the distribution.

Q14. Is K means a deterministic algorithm?
Ans. No, K means is non-deterministic in nature. As initial centroids, it uses random data points. Each time you run the algorithm, you will get the different answer.

2. SQL Worksheet 2

Q1. Which of the following constraint requires that there should not be duplicate entries?
Ans. (d) Unique

Q2. Which of the following constraint allows null values in a column?
Ans. © Null

Q3. Which of the following statements are true regarding Primary Key?
Ans. (a) Each entry in the primary key uniquely identifies each entry or row in the table

Q4. Which of the following statements are true regarding Unique Key?
Ans. (d) All of the above

Q5. Which of the following is/are example of referential constraint?
Ans. (b) Foreign Key

Q6. How many foreign keys are there in the Supplier table?
Ans. (d) 1

Q7. The type of relationship between Supplier table and Product table is:
Ans. (b) Many to one

Q8. The type of relationship between Order table and Headquarter table is:
Ans. (b) Many to one

Q9. Which of the following is a foreign key in Delivery table?
Ans. (d) None of them

Q10. The number of foreign keys in order details is:
Ans. (b) 1

Q11. The type of relationship between Order Detail table and Product table is:

Ans. (b) Many to one

Q12. DDL statements perform operation on which of the following database objects?
Ans. (a) Rows of table

Q13. Which of the following statement is used to enter rows in a table?
Ans. (a) Insert in to

Q14. Which of the following is/are entity constraints in SQL?
Ans. B) Unique C) Primary Key D) Null

Q15. Which of the following statements is an example of semantic Constraint?
Ans. (d) Two or more donors can have same blood group

3. Statistics Worksheet 2

Q1. What represent a population parameter?
Ans. © Both

Q2. What will be median of following set of scores (18,6,12,10,15)?
Ans. © 12

Q3. What is standard deviation?
Ans. (d) All of the above

Q4. The intervals should be _____ in a grouped frequency distribution?
Ans. © Both of above

Q5. What is the goal of descriptive statistics?
Ans. (b) Summarizing and explaining a specific set of data

Q6. A set of data organized in a participant by variables format is called
Ans. (b) Data set

Q7. In multiple regression,_____ independent variables are used
Ans. (a) 2 or more

Q8. Which of the following is used when you want to visually examine the relationship between 2 quantitative variables?
Ans. (b) Scatterplot

Q9. Two or more groups means are compared by using
Ans. (d) Analysis of variance

Q10. _____is a raw score which has been transformed into standard deviation units?
Ans. (a) Z-Score

Q11. _____is the value calculated when you want the arithmetic average?
Ans. © mean

Q12. Find the mean of these set of number (4,6,7,9,2000000)?
Ans. (d) 400005.2

Q13. _____ is a measure of central tendency that takes into account the magnitude of scores?
Ans. (d) Mean

Q14. _____ focuses on describing or explaining data whereas _____involves going beyond immediate data and making inferences.
Ans. (a) Descriptive and inferences

Q15. What is the formula for range?
Ans. (d) H-L