Machine Learning

Assignment – 2

1.(a) – 2 only.

2.(d) – 1,2 and 4 only.

3.(a) – True.

4.(a) – 1 only.

5.(b) – 1 0nly.

6.(b) – No.

7.(a) – Yes.

8.(d) - All of the above.

9.(a) – K means clustering algorithm.

10.(d) – All of the above.

11.(d) – All of the above.

12. Yes k is sensitive to the outliers. K is sensitive to the outliers because mean get impacted by extreme values thus it fails to give appropriate cluster center.

13. K mean is better because it has some advantages such as it is simple to implement , large set of data can be analyzed , convergence is guaranted , it can warm start the positions of centroids , easy adoptable to examples , generalized to clusters of different shapes and sizes like elliptical clusters.

14. No k mean is not deterministic algorithm because it has non deterministic algorithm due to this random selection of data points as initial centroids.

Worksheet 2 SQL

Assignment 2

1.(D) – Unique.

2.(A) – Primary key.

3.(C) - There can be null values in primary keys.

4.(A) – There should not be any duplicate entries.

5.(B) – Foreign key.

6.(D) - 1

7.(A) – One to many.

8.(D) – Many to many.

9.(B) – Supplier id.

10.(D) – 2.

11.(A) – One to many.

12.(C) – Table.

13.(A) – Insert in to.

14.(B) ,(C),(D) – Unique, primary key and null.

15.(A),(B) – A blood group can contain one of the following values – A , B , AB and O . And a blood can only contain characters.

Statistics Assignment 2

1.(C) - Both

2.(C) - 12

3.(D) – All of the above.

4.(C) – Both of these.

5.(B) – Summarizing and explaining a specific set data.

6.(B) – Data set.

7.(A) – 2 or more.

8.(B) – Scatterplot.

9.(D) – Analysis of variance.

10.(A) – Z score.

11.(C) – Mean.

12.(D) – 400005.2

13.(D) – Mean.

14.(D) – Central tendency.

15.(D) – H - L

4.

5.

6.

7.

8.

9.

10.

11.

12.

13.

14.

15.