WORKSHEET-2

MACHINE LEARNING

1. d)
2. d)
3. a)
4. a)
5. b)
6. a)
7. a)
8. d)
9. a)
10. a)
11. a)
12. Now coming to your question, K-means++ is still sensitive to the outliers. One workaround could be removing outliers using techniques like LOF, RANSAC, simple univariate box-plots, etc. before clustering. The other I think could be reinitialize the Centroids in case you are getting sub-optimal performance in the first attempt.
13. K-Means Advantages : 1) If variables are huge, then K-Means most of the times computationally faster than hierarchical clustering, if we keep k smalls. 2) K-Means produce tighter clusters than hierarchical clustering, especially if the clusters are globular. How does K mean?
14. Non-deterministic algorithm

STATISTICS WORKSHEET-1

1. a)

2. a)

3. a)

4. d)

5. c)

6. b)

7. b)

8. d)

9. c)

10. Normal Distribution is also known as Gaussian distribution is special type of distribution that looks like bell-shaped curve also the mean of that curved 0 and the standard deviation is 1

11. For The missing data i would suggest the following technique

Drop all the null values (if we have larage dataset)

Fill with mean or Median

Fill backward

Fill forwrad

Regression imputation

12. A/B testing is a user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics.

13. Mean imputation is one of the popular and simple to use but it's not very accurate so Regression imputation could be a better practice

14. Linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables. The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear

regression

15. The two main branches of statistics are descriptive statistics and inferential statistics

Descriptive statistics - A descriptive statistic is a summary statistic that quantitatively describes or summarizes features from a collection of information, while descriptive statistics is the process of using and analysing those statistics.

Inferential statistics - Statistical inference is the process of using data analysis to infer properties of an underlying distribution of probability. Inferential statistical analysis infers properties of a population, for example by testing hypotheses and deriving estimates.

WORKSHEET -2 SQL

1. d)
2. c)
3. b)
4. d)
5. d)
6. c)
7. c)
8. d)
9. b)
10. d)
11. b)
12. c)
13. d)
14. b) c)
15. a) b)