WORKSHEET SET 1 : SOLUTIONS

MACHINE LEARNING WORKSHEET

Q1. (A) Least square error

Q2. (A) Linear regression is sensitive to outliers

Q3. (B) Negative

Q4. (A) Regression

Q5. (C) Low bias and high variance

Q6. (B) Predictive model

Q7. (D) Regularization

Q8. (D) SMOTE

Q9. (A) TPR & FPR

Q10. (B) False

Q11. (D) Forward selection

Q12. (A) We don't have to choose the learning rate.
(B) It becomes slow when number of features is very large.
(C) We need to iterate

Q13. Regularization in Machine Learning, is a technique used to regularize the data in order to reduce the number of errors in a particular dataset. It helps in optimizing a dataset so that it produces the best results after training and testing. It helps solve the problem of over fitting and under fitting.

Q14. Two main algorithms are used in Regularization:

Ridge Regression

Lasso Regression

Ridge Regression is also called L2 Regularization. It uses normalized weights in the form of squares. On the other hand, Lasso regression or L1 regularization, only uses absolute weights.

Q15. Error present in a Linear Regression equation refers to the points away from the line of regression. It represents the margin of error which must be accounted for.

STATISTICS WORKSHEET

Q1. (A) True

Q2. (A) Central Limit Theorem

Q3. (B) Modeling bounded count data

Q4. (D) All of the mentioned

Q5. (C) Poisson

Q6. (B) False

Q7. (B) Hypothesis

Q8. (A) 0

Q9. (B) Outliers can be the result of spurious or real processes

Q10. A Normal Distribution is a distribution in which the mean, median and mode are same. Data in a normal distribution is distributed symmetrically. On a graph, a normal distribution shows a bell-shaped curve. It is also called a Gaussian Distribution.

Q11. Missing Data can cause errors and lead to distortions in datasets. Imputation of missing data is necessary for smooth functioning. Some imputation techniques include: Complete Case Analysis (CCA), Mode Imputation.

Q12. A/B testing refers to a test in which two versions of the same variable are compared to see which version performs better. It is also called split testing.

Q13. It is not an acceptable practice as mean imputation preserves the mean of observed data which leads to an underestimate of the standard deviation.

Q14. Linear Regression in statistics, is the relationship between two variables. One variable is dependent on the other. There is a Dependent Variable and an Independent Variable. A change in the Independent variable leads to a change in the Dependent variable. A straight line is used to show the relationship between these variables on a graph.

Q15. Statistics is divided into two branches:

Descriptive Statistics

Inferential Statistics

PYTHON WORKSHEET

Q1. (C) %

Q2. (B) 0

Q3. (C) 24

Q4. (A) 2

Q5. (D) 6

Q6. (C) The finally block will be executed no matter if the try block raises an error or not.

Q7. (A) it is used to raise an exception

Q8. (C) in defining a generator

Q9. (A) _abc & (C) abc2

Q10. (A) yield & (B) raise