MACHINE LEARNING WORKSHEETS

1. A) Least Square Error
2. Linear Regression is sensitive to outliers
3. B) Negative
4. B) Correlation
5. C) Low bias and high variance
6. B) Predictive model
7. D) Regularization
8. D) SMOTE
9. A) TPR and FPR
10. B) False
11. B) Apply PCA to project high dimensional data
12. A) We don't have to choose the learning rate.
    B) It becomes slow when the number of features is very large.
    C) We need to iterate.
13. Regularization is a technique used in machine learning to prevent overfitting by adding a penalty term to the objective function. It helps in controlling the complexity of the model and reducing the impact of large coefficient values.
14. Lasso regression and Ridge regression are commonly used algorithms for regularization.
15. The error present in the linear regression equation refers to the difference between the predicted values and the actual values of the dependent variable. It is also known as the residual or the loss function, and the goal in linear regression is to minimize this error to find the best fit line.

PYTHON WORKSHEET

Q1: C) %

Q2: B) 0

Q3: C) 24

Q4: D) 0

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 and C) abc2

Q10: A) yield and B) raise

STATISTICS WORKSHEET

1. a) True
2. a) Central Limit Theorem
3. b) Modeling bounded count data
4. c) The square of a standard normal random variable follows what is called chi-squared distribution
5. c) Poisson
6. b) False
7. b) Hypothesis
8. a) 0
9. c) Outliers cannot conform to the regression relationship
10. Normal Distribution refers to a probability distribution that is symmetric and bell-shaped, characterized by its mean and standard deviation. It is commonly used in statistics to model continuous random variables.
11. Handling missing data involves various techniques, including imputation. Imputation techniques can include mean imputation, median imputation, regression imputation, or multiple imputation, depending on the nature and pattern of the missing data. The choice of technique depends on the specific situation and assumptions.
12. A/B testing is a statistical hypothesis testing method used to compare two different versions of a variable or treatment to determine which one performs better. It is commonly used in fields like marketing and web design to evaluate changes and make data-driven decisions.
13. Mean imputation of missing data is generally not considered the best practice. It assumes that the missing values are missing completely at random and can lead to biased estimates and distorted results. Alternative imputation methods that preserve the uncertainty of the missing values are often recommended.
14. Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It aims to find the best-fitting linear equation that describes the linear relationship between the variables and allows for prediction and inference.
15. Branches of statistics include:
    a. Descriptive statistics: Involves summarizing and describing data using measures such as mean, median, and standard deviation.
    b. Inferential statistics: Concerned with making inferences and drawing conclusions about a population based on a sample.
    c. Probability theory: Studies the likelihood of events and random processes.
    d. Regression analysis: Examines the relationship between dependent and independent variables.
    e. Experimental design: Focuses on designing and analyzing controlled experiments.
    f. Biostatistics: Applies statistical methods to biological and health-related data.
    g. Econometrics: Applies statistical methods to economic data and models.
    h. Time series analysis: Analyzes data collected over time to identify patterns and make predictions.