Machine Learning

Ans 1 :- A) Least Square Error

Ans 2 :- (A) Linear regression is sensitive to outliers

Ans 3 :- B) Negative

Ans 4 :- B) Correlation

Ans 5 :- C) Low bias and high variance

Ans 6 :- B) Predictive model

Ans 7 :- D) Regularization

Ans 8 :- D) SMOTE

Ans 9 :- A) TPR and FPR

Ans 10 :- B) False

Ans 11 :- B) Apply PCA to project high dimensional data

Ans 12 :- A) We don't have to choose the learning rate.

B) It becomes slow when the number of features is very large.

Ans 13 :- **Regularization** is a technique used in machine learning and statistics to prevent overfitting of a model to the training data. It helps to improve the model's generalization to new, unseen data by adding a penalty for complexity to the model's loss function. Here's a more detailed explanation:
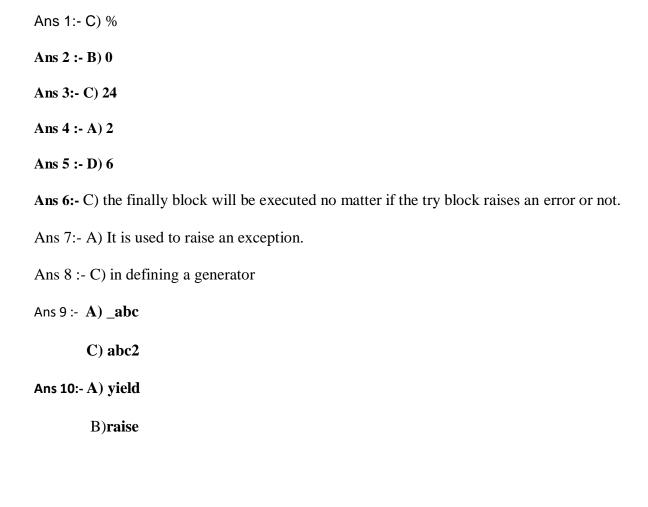
Types of Regularization:

1- L1 Regularization (Lasso):
2- L2 Regularization (Ridge):
3- Elastic Net Regularization:

Ans 14 :- Lasso Regression

Ans 15 :- An error term represents the margin of error within a statistical model; it refers to the sum of deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results. The regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent variable.

Python – worksheet

Ans 1:- C) %

Ans 2 :- B) 0

Ans 3:- C) 24

Ans 4 :- A) 2

Ans 5 :- D) 6

Ans 6:- C) the finally block will be executed no matter if the try block raises an error or not.

Ans 7:- A) It is used to raise an exception.

Ans 8 :- C) in defining a generator

Ans 9 :- **A) _abc**

C) abc2

Ans 10:- A) yield

B)**raise**

Statistics Worksheet

Ans 1:- A) True

Ans 2:- a) Central Limit Theorem

Ans 3 :- b) Modeling bounded count data

Ans 4 :- d) All of the mentioned

Ans 5 :- c) Poisson

Ans 6 :- b) False

Ans 7 :- b) Hypothesis

Ans 8 :- a) 0

Ans 9 :- d) None of the mentioned

Ans 10 :-
Normal Distribution in Statistics

The normal distribution, also known as the Gaussian distribution, is the most important probability distribution in statistics for independent, random variables. Most people recognize its familiar bell-shaped curve in statistical reports.The normal distribution is a continuous probability distribution that is symmetrical around its mean, most of the observations cluster around the central peak, and the probabilities for values further away from the mean taper off equally in both directions. Extreme values in both tails of the distribution are similarly unlikely. While the normal distribution is symmetrical, not all symmetrical distributions are normal. For example, the Student's , Cauchy, and logistic distributions are symmetric.

Ans 11:-

Ans 12 :- **A/B testing** is a statistical method used to compare two versions of a variable to determine which one performs better in a controlled experiment. It is widely used in various fields, including marketing, web design, product development, and user experience (UX) research.

Ans 13 :- The process of replacing null values in a data collection with the data's mean is known as mean imputation.

Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

Second, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

Ans 14 :- **Linear regression** is a fundamental statistical method used to model the relationship between a dependent variable and one or more independent variables. It is a type of regression analysis that assumes a linear relationship between the variables.

Ans 15 :- The two main branches of statistics are

Descriptive and inferential