Machine Learning Worksheet
1. d both a and b
2. a linear regression is sensitive to outliners
3. b negative
4. c both of them
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
12. a we dont have to choose the learning rate
13. Regularization is a technique used in machine learning to prevent overfitting and improve the generalization of a model. It involves adding a penalty term to the objective function that the model is trying to minimize during training. The purpose of this penalty term is to discourage the model from fitting the training data too closely, which can lead to poor performance on new, unseen data.
14. Regularization techniques are commonly applied to various machine learning algorithms to prevent overfitting.
15. the "error" refers to the difference between the predicted values generated by the regression model and the actual observed values in the dataset. This difference is often denoted as the residual.

Python Worksheet
1. c %
2. b 0
3. c 24
4. d 0
5. d 6
6. c the finally block will be executed no matter if the try block raises an error or not
7. a it is used to raise an exception
8. c in defining a generator
9. a _abc and c abc2
10. a yield and b raise

Statistics Worksheet
1.B(False)
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)
8.0
9.C(Outliers cannot conform to the regression relationship)
10.Normal distribution, also known as Gaussian distribution or bell curve, is a continuous probability distribution that is symmetrical around its mean, exhibiting a characteristic bell-shaped curve.
11.Handling missing data is crucial in data analysis. Several imputation techniques can be used, depending on the nature of the data and the reasons for missing values.
12.A/B testing, also known as split testing, is a statistical method used in marketing, product development, and other fields to compare two versions of a variable (A and B) to determine which one performs better. The goal is to identify changes that improve a given objective, such as user engagement, click-through rates, or conversion rates.
13.Mean imputation, where missing values are replaced with the mean of the observed values, is a simple and commonly used technique. However, its acceptability depends on the context and assumptions about the data.
14.Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. The basic form of a linear regression model for a single independent variable.
15.Statistics can be broadly categorized into two main branches: Descriptive Statistics and Inferential Statistics. Each of these branches further encompasses various methods and techniques.