**Machine learning assignment**

**Ans .1 Least Square Error**

**Ans. 2 Linear regression is sensitive to outliers**

**Ans.3 Positive**

**Ans.4 Correlation**

**Ans.5 Low bias and high variance**

**Ans .6 Predictive model**

**Ans.7 Regularization**

**Ans .8 Smote**

**Ans. 9 TPR and FPR**

**Ans .10 True**

**Ans.11 Apply PCA to project high dimensional data**

**Ans.12 We need to iterate**

**Ans 13. Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.**

**Ans 14. Algorithm use for Regularization are as follows :**

**Ridge Regression:**

**It modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients.**

**This means that the mathematical function representing our machine learning model is minimized and coefficients are calculated. The magnitude of coefficients is squared and added. Ridge Regression performs regularization by shrinking the coefficients present.**

**Lasso Regression:**

**It modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients.**

**Lasso regression also performs coefficient minimization, but instead of squaring the magnitudes of the coefficients, it takes the true values of coefficients. This means that the coefficient sum can also be 0, because of the presence of negative coefficients.**

**Ans 15. The term error present in linear regression equation is difference between the actual value and predicted value.**

**Python learning assignment**

**Ans .1 %**

**Ans. 2 1**

**Ans.3 1**

**Ans.4 2**

**Ans.5 0**

**Ans .6 None of the above**

**Ans.7 It is used to raise an exception.**

**Ans .8 in defining a generator**

**Ans. 9 None of the above**

**Ans .10 Yeild ,Raise**

**STATISTICS LEARNING ASSIGNMENTS**

**Ans .1True**

**Ans. 2 Central Limit Theorem**

**Ans.3 Modelling bounded count data**

**Ans.4 All of the mentioned above**

**Ans.5 Poisson**

**Ans .6 False**

**Ans.7 Hypothesis**

**Ans .8 0**

**Ans. 9 Outliers cannot conform to the regression relationship**

**Ans 10. a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.**

**Ans11. When dealing with missing data, data scientists can use two primary methods to solve the error: imputation or the removal of data. The imputation method develops reasonable guesses for missing data. It's most useful when the percentage of missing data is low.**

**Imputation using the statistics (mean, median, mode) and K-Nearest Neighbor Imputation.**

**Ans.12 An AB test is an example of statistical hypothesis testing, a process whereby a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not.**

**Ans 13 Mean imputation does not preserve the relationships among variables. True, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased. That's a good thing.**

**Ans 14 Linear regression is an attempt to model the relationship between two variables by fitting a linear equation to observed data, where one variable is considered to be an explanatory variable and the other as a dependent variable.**

**Ans 15. Two types of statistical methods are used in analysing data: descriptive statistics and inferential statistics.**

**Ans.11**