Machine Learning Assignment Answers :-

Ans.1 :- Least Square Error

Ans.2 :- Linear regression is sensitive to outliers

Ans.3 :- Negative

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 don't have to choose the learning rate.

It becomes slow when number of features is very large.

We need to iterate. We need to iterate.

Ans.13 :-

Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it.

Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

Sometimes the machine learning model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when deals with unseen data by introducing noise in the output, and hence the model is called overfitted. This problem can be deal with the help of a regularization technique.

This technique can be used in such a way that it will allow to maintain all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy as well as a generalization of the model.

There are mainly two types of regularization :-

- **Ridge Regression**
     - It is also called as L2 regularization.
- **Lasso Regression**
     - It is also called as L1 regularization.

In simple words, *"In regularization technique, we reduce the magnitude of the features by keeping the same number of features."*

Ans.14 :-

The most popular regularization algorithm are used given below :-

- Ridge Regression
- Lasso Regression
- Elastic Net

1) **Ridge Regression** :-

   Ridge Regression is also called as L2 regularization.

   Ridge regression is a regularization technique, which is used to reduce the complexity of the model.

   Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions.

   In this technique, the cost function is altered by adding the penalty term to it. The amount of bias added to the model is called Ridge Regression penalty.

2) **Lasso Regression** :-

Lasso regression is another regularization technique to reduce the complexity of the model. It stands for Least Absolute and Selection Operator.

Lasso Regression is also called as L1 regularization.

It is similar to the Ridge Regression except that the penalty term contains only the absolute weights instead of a square of weights.

Since it takes absolute values, hence, it can shrink the slope to 0, whereas Ridge Regression can only shrink it near to 0.

3) **Elastic Net** :-

When we are working with high-dimensional data sets with a large number of independent variables, correlations (relationships) amid the variables can be often result in multicollinearity.

These correlated variables which are strictly related can sometimes form groups or clusters called as an elastic net of correlated variables.

We would want to include the complete group in the model selection even if just one variable has been selected.

Ans.15 :-

Error is the difference between the actual value and predicted value and the goal is to reduce this difference.

The vertical distance between the data point and the regression line is known as error or residual.

Each data point has one residual and the sum of all the differences is known as the sum of residuals or error.

An error term essentially means that the model is not completely accurate and results in differing results during real-world applications.

The error term used in linear regression is :- $Y = (a + bX) + \varepsilon$

Here, the Y is the dependent variable, X is the independent variable, a is the intercept, b is the coefficient and ε is the error term.

This error is called marginal or residual error.