Machine Learnning

Q 1] A Least square method

Q 2] A) Linear regression is sensitive to outliers

Q 3] B) Negative

Q 4] B) Correlation

Q 5] C) Low bias and high variance

Q 6] B) predictive model

Q 7] D) Regularization

Q 8] D) SMOTE

Q 9] A) TPR and FPR

Q10] B) False

Q 11] D) all of the above

Q12] A) & B)

Q 13] 1) Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting. 2) 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 test or unseen data, and such model is called overfitted model. This problem can be solve with the help of a regularization technique.3) Overfitting is a phenomenon that occurs when a machine Learning model is constraint to training set and not able to perform well on unseen data.4) Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting. 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. It mainly regularizes or reduces the coefficient of features toward zero by adding penalty term to complex models.

$$\hat{Y} = \beta X + \alpha$$ Where Ŷ is our predicted value, X are the features, and β and α is the intercept of the model. In a regression model, our job is to find the best β and α.

$$\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$ Where n is the number of observations, ŷ are the predicted values and y are the actual values for the target variable.

In case of regularization we will add penalty term to this loss function . Penalty term includes

Regularization parameter lambda which which determines degree of regularization .Depending upon penalty term addition to loss function ,there 3 types of regularization

1) L1 or LASSO REGULARIZATION
2) L2 or Ridge REGULARIZATION
3) Elastic Net

Q 14) Algorithm used for regularization are depends upon penalty terms used to minimize loss function of linear regression

a) L2 Regularization or Ridge Regression :

Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as L2 regularization.In this technique, the cost function is minimized by adding the penalty term . The amount of bias added to the model is called Ridge Regression penalty. This penalty is given by multiplying with the lambda(regularization parameter) to the squared weight of each individual feature.

$$\sum_{i=1}^{M}(y_i - y'_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{n}\beta_j * x_{ij}\right)^2 + \lambda\sum_{j=0}^{n}{\beta_j}^2$$

the penalty term regularizes the coefficients of the model, and hence ridge regression the amplitudes of the coefficients that decreases the complexity of the model.It doenot make coefficients directly 0 but keep their values near to 0

b) L1 regularization or LASSO REGRESSION :

Lasso regression is another regularization technique whichstands for **Least Absolute and Selection Operator.** the cost function is minimized by adding the penalty term . The amount of bias added to the model is called LASSO Regression penalty. This penalty is given by multiplying with the lambda(regularization parameter) to the absolute weight of each individual feature.In lasso , penalty term minimizes loss function in such way that feature having less importance or contriution ,their coefficients to becomes zero i.e. these features are eliminated .

$$\sum_{i=1}^{M}(y_i - y'_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{n}\beta_j * x_{ij}\right)^2 + \lambda\sum_{j=0}^{n}|\beta_j|$$

Q15] Error term in Linear Regression

1)The linear regression involves fitting of best fit line data by evaluating coefficients and intercept Which show weightage of each feature in predicting label variable. Prediction of target variable made by regression model shows deviation from actual labels. This difference between expected values or predicted values and actual values are termed as **error term.**

2)An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. The error term is a residual variable that accounts for a lack of perfect goodness of fit. it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results.

Error= actual label - predicted label