Machine Learning

Q.1 Which of the following methods do we use to find the best fit line for data in Linear Regression?

Ans- Least Square Error

Q.2 Which of the following statement is true about outliers in linear regression?

Ans- Linear regression is sensitive to outliers

Q.3 A line falls from left to right if a slope is _____?

Ans- Negative

Q.4 Which of the following will have symmetric relation between dependent variable and independent variable?
Ans- Correlation

Q.5 Which of the following is the reason for over fitting condition?

Ans- Low Bias and High Variance

Q.6 If output involves label then that model is called as:

Ans- Predictive Model

Q.7 Lasso and Ridge regression techniques belong to _____?

Ans- Regularization

Q.8 To overcome with imbalance dataset which technique can be used?

Ans- SMOTE

Q.9 The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

Ans – Sensitivity and Specificity

Q.10 In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
Ans- False

Q.11 Pick the feature extraction from below:

Ans- Construction bag of words from an email

Q.12 Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
Ans- 1. We don't have to choose the learning rate.
2. It becomes slow when number of features is very large.

Q.13 Regularization

Regularization is a technique in machine learning to prevent the model from over fitting. In some cases, the model works perfectly fine with training data however it doesn't perform well with 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 techniques, which are given below:

1. **Ridge Regression**
2. **Lasso Regression**

Q.14 Which particular algorithms are used for regularization?

1. **Ridge Regression:**

Ridge regression is a method for analyzing data that suffer from multi-collinearity. Ridge regression adds a penalty to the loss function that is equivalent to the square of the magnitude of the coefficients.

The regularization parameter *(λ)* regularizes the coefficients such that if the coefficients take large values, the loss function is penalized.

- $\lambda \rightarrow 0$, the penalty term has no effect, and the estimates produced by ridge regression will be equal to least-squares i.e. the loss function resembles the loss function of the Linear Regression algorithm. Hence, a lower value of λ will resemble a model close to the Linear regression model.
- $\lambda \rightarrow \infty$, the impact of the shrinkage penalty grows, and the ridge regression coefficient estimates will **approach zero** (coefficients are close to zero, but not zero).
- Ridge regression shrinks the coefficients as it helps to reduce the model complexity and multi-collinearity.

2.LASSO Regression: LASSO is a regression analysis method that performs both feature selection and regularization in order to enhance the prediction accuracy of the model. LASSO regression adds a penalty *(L1 penalty)* to the loss function that is equivalent to the magnitude of the coefficients**.** In LASSO regression, the penalty has the effect of forcing some of the coefficient estimates to be exactly equal to zero when the regularization parameter λ is sufficiently large. *Note: LASSO regression is also known as the L1 Regularization (L1 penalty).* To sum up, LASSO regression converts coefficients of less important features to zero, which indeed helps in feature selection, and it shrinks the coefficients of remaining features to reduce the model complexity, hence avoiding overfitting**.**

Q.15 Explain the term error present in linear regression equation?

Linear regression most often uses mean-square error (MSE) to calculate the error of the model. MSE is calculated by:

1. measuring the distance of the observed y-values from the predicted y-values at each value of x;
2. squaring each of these distances;
3. calculating the mean of each of the squared distances.

Linear regression fits a line to the data by finding the regression coefficient that results in the smallest MS