# Machine Learning Assignment

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

A) Least Square Error

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

A) Linear regression is sensitive to outliers

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

B) Negative

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

D) None of these

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

C) Low bias and high variance

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

B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

D) Regularization

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

D) SMOTE

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

A) TPR and FPR

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

A) True

11. Pick the feature extraction from below:

C) Removing stop words

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

C) We need to iterate. & D) It does not make use of dependent variable.

14. Which particular algorithms are used for regularization?

**Ans.**

**There are available three main regularization techniques**-

- Ridge Regression (L2 Norm)
- Lasso (L1 Norm)
- Dropout

Mostly used algorithms is ridge regression

**Ridge Regression (L2 Regularization)-** Ridge regression is also called L2 norm or regularization.While using this technique, we add the sum of the weight squares to the loss function thus creating a new loss function. Ridge regression is a model tuning method used to analyze any data that is prone to multicollinearity. This method performs L2 regularization. When multiplexing is the issue, least squares is unbiased, and the variances are large, resulting in the predicted values being far from the true values.

The original loss function

$$\text{Loss} = \sum_{j=1}^{m}\left(Yi - Wo - \sum_{i=1}^{n} Wi\,Xji\right)^2 + \lambda \sum_{i=1}^{n} Wi^2$$

**Lasso Regression (L1 Regularization)-** This technique is different from ridge regression as it uses absolute weight values for normalization. λ is again a tuning parameter and behaves in the same as it does when using ridge regression. As loss function only considers absolute weights, optimization algorithms penalize higher weight values.

$$\text{Loss} = \sum_{j=1}^{m}\left(Yi - Wo - \sum_{i=1}^{n} Wi\,Xji\right)^2 + \lambda \sum_{i=1}^{n} |Wi|$$

**Dropout -** Dropout is a regularization technique used in neural networks. This prevents complex co-adaptation from other neurons. In neural nets, fully connected layers are more prone to overfit on training data. Using dropout, you can drop connections with 1-p probability for each of the specified layers. Where p is called keep probability parameter and which needs to be tuned.

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

Ans

The error term is also known as the residual, turbulence, or remainder term, and is represented variously by the letters E, , or U in the model.

The term an error essentially means that the model is not completely accurate and gives different results when applied to real-world applications.

The two data points with the greatest distance from the trend line should be equidistant from the trend line, which represents the largest margin of error.

**Mean absolute error-** he mean absolute error (MAE) is the simplest regression error metric to understand. We'll calculate the residual for every data point, taking only the absolute value of each so that negative and positive residuals do not cancel out. We then take the average of all these residuals. Effectively, MAE describes the typical magnitude of the residuals. If you're unfamiliar with the mean, you can refer back to this article on descriptive statistics. The formal equation is shown below –

$$\text{RMSE} = \sqrt{\frac{\sum (y_i - y_p)^2}{n}}$$

$$\text{MAE} = \frac{|(y_i - y_p)|}{n}$$

$y_i$ = actual value
$y_p$ = predicted value
$n$ = number of observations/rows

**Mean square error-** The mean square error (MSE) is just like the MAE, but squares the difference before summing them all instead of using the absolute value. We can see this difference in the equation below.

$$MSE = \frac{1}{n}\Sigma\underbrace{\left(y - \hat{y}\right)^2}_{\text{The square of the difference between actual and predicted}}$$