Worksheet Set: 1

Machine Learning

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

= Least Sqaure Error. (A)

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

= Linear Regression is Sensitive to Outliers. (A)

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

= Positive. (A)

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

= Correlation. (B)

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

= Low Bias and High Variance. (C)

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

= Predictive Model. (B)

7. Lasso and Ridge regression techniques belong to _____?

= Regularization. (D)

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

= SMOTE. (D)

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

= TPR and FPR. (A)

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

= False (B)

11. Pick the feature extraction from below:

= Apply PCA to project high dimensional data. (B)

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

= **We don't have to choose the learning rate. (A),**

It becomes slow when number of features is very large. (B),

We need to iterate. (C)

13. Explain the term regularization?

= In Regression Model, while we train our dataset into the Model, there is a chance that maybe our model will face the overfitting problem. So, in such of the problems of overfitting we face in our model we use Regularization to solve the overfitting problem present in model.

Regularization will restrict the degree of freedom of a given equation, i.e., simply reducing the no. of degrees of a polynomial function by reducing their corresponding weight.

In other words, we can say that, when the model worked very well on the test data but doesn't perform good in real-time, then the model would be an overfitted data to resolve this we use Regularization.

14. Which particular algorithms are used for regularization?

= There are some algorithms we used for regularization:

- LASSO (L1 Form)

LASSO (Least Absolute Shrinkage & Selection Operator) regression penalizes the model based on the sum of magnitude of the coefficients. The regularization term is given by

regularization= $\lambda * \sum |\beta j|$ (Where, λ is the shrinkage factor.)

- **Lasso will make the unwanted features as zero (0)**
- **Lasso will also identify all the features which are contributing to the label, so it can also work as Features selection.**

- Ridge Regression (L2 Form)

Ridge regression penalizes the model based on the sum of squares of magnitude of the coefficients. The regularization term is given by

regularization= $\lambda * \sum |\beta j^2|$ (Where, λ is the shrinkage factor.)

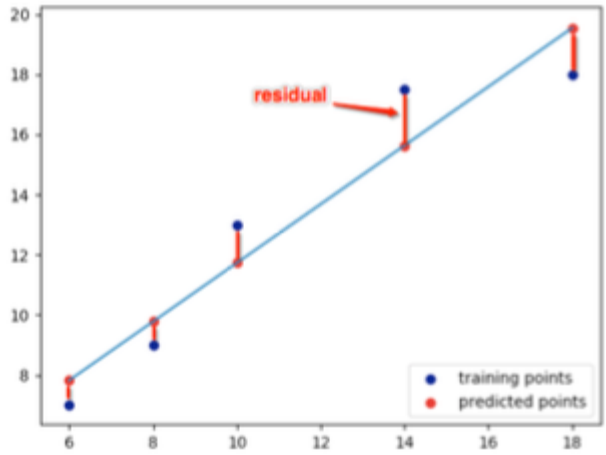
- **Ridge will not make the unwanted features as zero (0), instead it will give very less amount of importance to this kind of unwanted feature.**

- Elastic Net

Elastic Net Regression is very less popular algorithm. It uses the penalties from both the lasso and ridge techniques to regularize regression models.

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

= Error is difference between actual and predicted values. Error is also termed as Residual.



- The Red Points are Predicted data.
- Blue Points are the actual data.
- Light Blue Line is Least Square Line.
- Red Lines are residual/error, which is the distance between actual and predicted data.