MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:

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

Ans: A) Least Square Error

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

Ans: A) Linear regression is sensitive to outliers

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

Ans: B) Negative

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

Ans: B) Correlation

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

Ans: A) High bias and high variance

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

Ans: B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

Ans:D) Regularization

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

 Ans: D) SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary

classification problems. It uses _____ to make graph?

Ans: A) TPR and FPR

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

Ans: B) False

11. Pick the feature extraction from below:

Ans: B) Apply PCA to project high dimensional data

In Q12, more than one options are correct, choose all the correct options:

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

Ans:

A) We don't have to choose the learning rate.

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

ASSIGNMENT – 39

MACHINE LEARNING

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?

Ans: When training a machine learning model, the model can be easily overfitted or under fitted. To avoid this, we use regularization in machine learning to properly fit the model to our test set. Regularization techniques help reduce the possibility of overfitting and help us obtain an optimal model. Regularization refers to techniques used to calibrate machine learning models to minimize the adjusted loss function and avoid overfitting or underfitting.
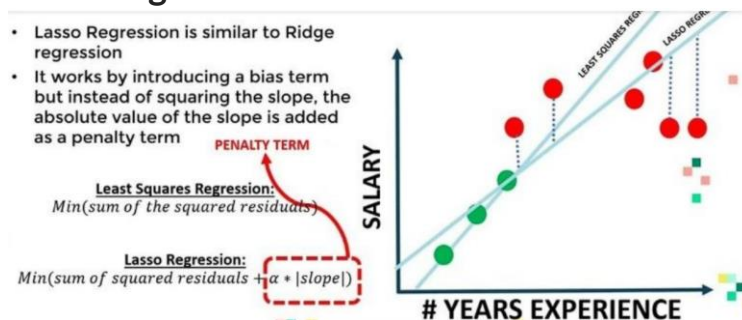
14. Which particular algorithms are used for regularization?

Regularization Techniques

1. Ridge Regularization
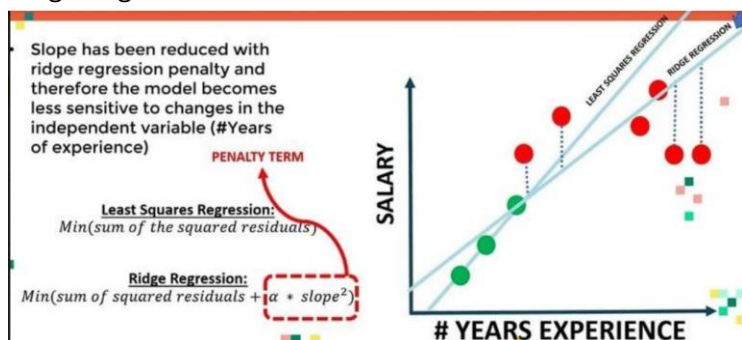
2. Lasso Regularization

3. Elastic Net

- Elastic net combines L1 and L2 with the addition of an alpha parameter deciding the ratio between them:

$$\frac{\sum_{i=1}^{n}(y_i - x_i^J \hat{\beta})^2}{2n} + \lambda \left(\frac{1-\alpha}{2} \sum_{j=1}^{m} \hat{\beta}_j^2 + \alpha \sum_{j=1}^{m} |\hat{\beta}_j| \right)$$

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

Ans: An error term in statistics is a value which represents how observed data differs from actual population data. It can also be a variable which represents how a given statistical model differs from reality. The error term is often written ε.

Properties of the Error Term

The error term includes everything that separates your model from actual reality. This means that it will reflect nonlinearities, unpredictable effects, measurement errors, and omitted variables.

Errors and Residuals

Although the terms error and residual are often interchanged, there is an important formal difference. While an error term represents the way observed data differs from the actual population, a residual represents the way observed data differs from sample population data. This means that a residual is often much easier to quantify. Although an error is generally unobservable, a residual is observable.

The residual can be considered an estimate of the true error term.