

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 -> D) None of these

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

Ans -> B) Low bias and low 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

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

Ans -> D) It does not make use of dependent variable.

13. Explain the term regularization?

Ans -> Regularization is a set of methods for reducing overfitting in machine learning models. Typically, regularization trades a marginal decrease in training accuracy for an increase in

generalizability. Regularization encompasses a range of techniques to correct for overfitting in machine learning models.

14. Which particular algorithms are used for regularization?

Ans -> Regularization in machine learning is a set of techniques used to ensure that a machine learning model can generalize to new data within the same data set. These techniques can help reduce the impact of noisy data that falls outside the expected range of patterns. Regularization can also improve the model by making it easier to detect relevant edge cases within a classification task.

Consider an algorithm specifically trained to identify spam emails. In this scenario, the algorithm is trained to classify emails that appear to be from a well-known U.S. drugstore chain and contain only a single image as likely to be spam. However, this narrow approach runs the risk of disappointing loyal customers of the chain, who were looking forward to being notified about the store's latest sales. A more effective algorithm would consider other factors, such as the timing of the emails, the use of images and the types of links embedded in the emails to accurately label the emails as spam.

This more complex model, however, would also have to account for the impact that each of these measures added to the algorithm. Without regularization, the new algorithm risks being overly complex, subject to bias and unable to detect variance.

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

Ans -> An error term represents the margin of error within a statistical model; 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.

To find the MSE, take the observed value, subtract the predicted value, and square that difference. Repeat that for all observations. Then, sum all of those squared values and divide by the number of observations. Notice that the numerator is the sum of the squared errors (SSE), which linear regression minimizes.