

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

Sol. A) Least Square Error

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

Sol. A) Linear regression is sensitive to outliers

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

Sol. B) Negative

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

Sol. C) Both of them

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

Sol. C) Low bias and high variance

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

Sol. B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

Sol. D) Regularization

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

Sol. D) SMOTE

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

Sol. A) TPR and FPR

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

Sol. B) False

11. Pick the feature extraction from below:

Sol. 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?

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

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

C) We need to iterate.

13. Explain the term regularization?

Sol. Regularization is a technique used in machine learning and statistics to prevent overfitting in a model. It involves adding a penalty term to the loss function during training to discourage the model from fitting the training data too closely. Regularization methods, such as L1 (Lasso) and L2 (Ridge) regularization, control the complexity of a model by either adding absolute values of coefficients (L1) or squared values of coefficients (L2) to the loss function. This helps to constrain the model's coefficients and prevent them from becoming excessively large, which can lead to instability and poor generalization on unseen data.

14. Which particular algorithms are used for regularization?

Sol. Regularization techniques are often used with various machine learning algorithms to prevent overfitting. Some algorithms that commonly use regularization include:

1. Linear Regression with Lasso (L1) and Ridge (L2) regularization.
2. Logistic Regression with L1 and L2 regularization.
3. Support Vector Machines (SVM) with L2 regularization

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

Sol. In the context of linear regression, the term "error" refers to the difference between the actual observed values of the dependent variable and the predicted values generated by the linear regression model. This error is also known as the residual. Mathematically, for each data point (i), the error (ϵ_i) is calculated as: $\epsilon_i = y_i - \hat{y}_i$, where y_i is the actual observed value and \hat{y}_i is the predicted value for the i -th data point. The goal of linear regression is to minimize these errors (residuals) by finding the best-fitting line that minimizes the sum of squared errors, known as the Least Squares Error (LSE). This line represents the linear relationship between the independent variable(s) and the dependent variable.