

Machine learning

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 root

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

Ans. A. Linear regression is sensitive to outliers

Explanation:

The slope of the regression line will change due to outliers in most of the cases. So Linear Regression is sensitive to outliers.

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

Ans. A. Positive

Explanation:

If the graph of a line rises from left to right, the slope is positive. If the graph of the line falls from left to right the slope is negative.

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

Ans: C. Both of them

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

Ans: C. Low bias and high variance

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

Ans: B. Predictive model

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: C. Sensitivity and Specificity

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

Ans: A. True

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

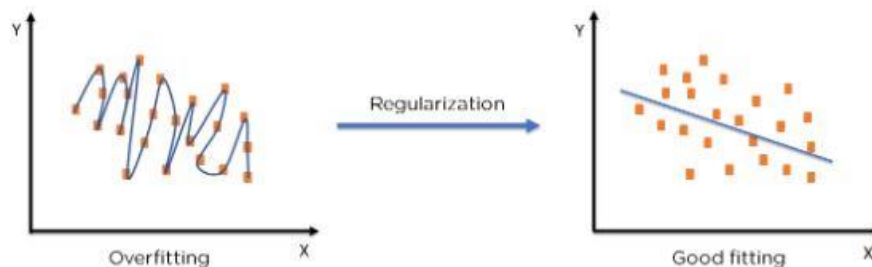
B. It becomes slow when number of features is very large.

C. We need to iterate.

Answer them briefly.

13. Explain the term regularization?

Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.



14. Which particular algorithms are used for regularization?

There are three main regularization techniques, namely:

Ridge Regression (L2 Norm)

Lasso (L1 Norm)

Dropout

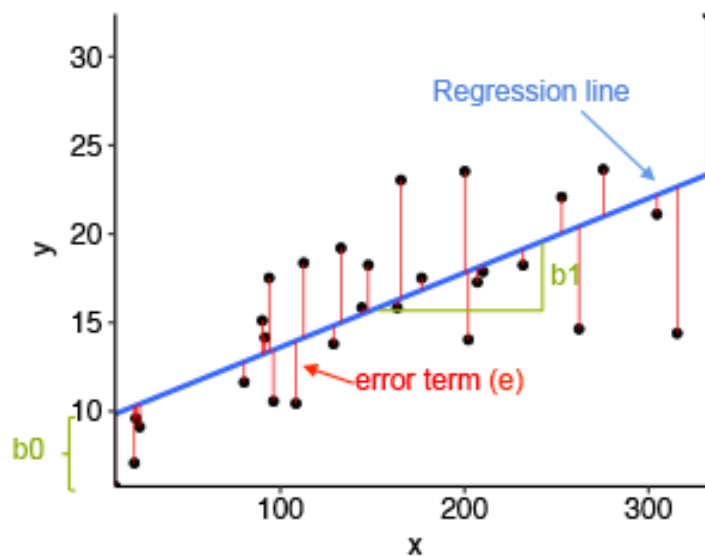
Ridge and Lasso can be used for any algorithms involving weight parameters, including neural nets. Dropout is primarily used in any kind of neural networks e.g. ANN, DNN, CNN or RNN to moderate the learning.

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

A Linear Regression model's main aim is to find the best fit linear line and the optimal values of intercept and coefficients such that the error is minimized.

Error is the difference between the actual value and Predicted value and the goal is to reduce this difference.

Example:



In the above diagram,

- x is our independent variable which is plotted on the x -axis and y is the dependent variable which is plotted on the y -axis.
- Black dots are the data points i.e the actual values.
- b_0 is the intercept which is 10 and b_1 is the slope of the x variable.
- The blue line is the best fit line predicted by the model i.e the predicted values lie on the blue line.

The vertical distance between the data point and the regression line is known as error or residual. Each data point has one residual and the sum of all the differences is known as the Sum of Residuals/Errors.

Mathematical Approach:

Residual/Error = Actual values – Predicted Values

Sum of Residuals/Errors = Sum(Actual- Predicted Values)

Square of Sum of Residuals/Errors = (Sum(Actual- Predicted Values))²

i.e

$$\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$$