

MACHINE LEARNING (With Answers)

Q1 to Q11

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- 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- 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

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.

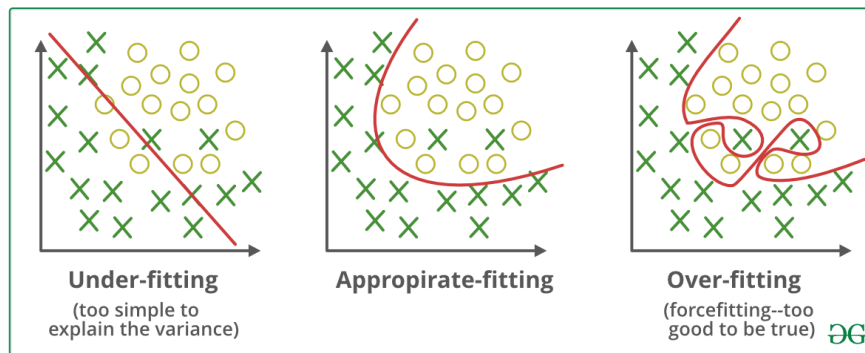
Q13 and Q15

13. Explain the term regularization?

Ans- 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.

A scenario where the machine learning model tries to learn from the details along with the noise in the data and tries to fit each data point on the curve is called **Overfitting**. Such models usually perform very well on training data but have high error rates on test data. High Variance causes overfitting in our model.

A scenario where a machine learning model can neither learn the relationship between variables in the testing data nor predict or classify a new data point is called **Underfitting**. Such models always lead to a high error on training and test data. High Bias causes underfitting in our model.



Regularization significantly reduces the variance of the model, without substantial increase in its bias. We can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

14. Which particular algorithms are used for regularization?

Ans- The commonly used regularization techniques are:

L1 regularization

L2 regularization

Dropout regularization

1. L1 regularization

A regression model which uses L1 Regularization technique is called **LASSO (Least Absolute Shrinkage and Selection Operator) regression**. It modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients. Lasso regression also performs coefficient minimization, by taking the true values of coefficients. This means that the coefficient sum can also be 0, because of the presence of negative coefficients.

2. L2 regularization

A regression model that uses L2 regularization technique is called **Ridge regression**. It modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients. This means that the mathematical function representing our machine learning model is minimized and coefficients are calculated. The magnitude of coefficients is squared and added. Ridge Regression performs regularization by shrinking the coefficients present.

3. Dropout regularization

Dropout regularization is a regularization technique used in neural networks. It prevents complex co-adaptations from other neurons. In neural nets, fully connected layers are more prone to overfit on training data. Using dropout, you can drop connections with 1-p probability for each of the specified layers. Where p is called **keep probability parameter**.

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

Ans- The error term in a regression equation represents the effect of the variables that were omitted from the equation.