

MACHINE LEARNING

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

- Which of the following methods do we use to find the best fit line for data in Linear Regression?
A) Least Square Error
B) Maximum Likelihood
C) Logarithmic Loss
D) Both A and B
➤ Ans :- (A)
- Which of the following statement is true about outliers in linear regression?
A) Linear regression is sensitive to outliers
B) linear regression is not sensitive to outliers
C) Can't say
D) none of these
➤ Ans :- (A)
- A line falls from left to right if a slope is _____?
A) Positive
B) Negative
C) Zero
D) Undefined
➤ Ans :- (a)
- Which of the following will have symmetric relation between dependent variable and independent variable?
A) Regression
B) Correlation
C) Both of them
D) None of these
➤ Ans:- (B)
- Which of the following is the reason for over fitting condition?
A) High bias and high variance
B) Low bias and low variance
C) Low bias and high variance
D) none of these
➤ Ans :- (C)
- If output involves label then that model is called as:
A) Descriptive model
B) Predictive modal
C) Reinforcement learning
D) All of the above
➤ Ans :- (B)
- Lasso and Ridge regression techniques belong to _____?
A) Cross validation
B) Removing outliers
C) SMOTE
D) Regularization
➤ Ans :- (D)
- To overcome with imbalance dataset which technique can be used?
A) Cross validation
B) Regularization
C) Kernel
D) SMOTE
➤ Ans :- (D)
- The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?
A) TPR and FPR
B) Sensitivity and precision
C) Sensitivity and Specificity
D) Recall and precision
➤ Ans :- (D)
- In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
A) True
B) False
➤ Ans :- (A)

MACHINE LEARNING

11. Pick the feature extraction from below:

- A) Construction bag of words from a email
- B) Apply PCA to project high dimensional data
- C) Removing stop words
- D) Forward selection

➤ Ans :- (B)

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?

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

➤ Ans :- (A) , (D)

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

13. Explain the term regularization?

- Ans :- in case of regularization, The model will itself adjust the values of the variable coefficients up and down so that the model could learn the things well. Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

There are two techniques :

- L1 regularization (LASSO regularization)
- L2 regularization (RIDGE regularization)

L1 regularization :- This technique is used to overcome overfitting for a regression model. Regularization is one of the methods widely used to make your model more generalized. Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models. This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.

L2 regularization :- Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values being far away from the actual values.

14. Which particular algorithms are used for regularization?

- Ans :- Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it. Sometimes the machine learning model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when it deals with unseen data by introducing noise in the output, and hence the model is called overfitted. This problem can be dealt with the help of a regularization technique. This technique can be used in such a way that it will allow to maintain all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy as well as a generalization of the model.
-

MACHINE LEARNING

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

- Ans :-Within a linear regression model tracking a stock's price over time, the error term is the difference between the expected price at a particular time and the price that was actually observed. In instances where the price is exactly what was anticipated at a particular time, the price will fall on the trend line and the error term will be zero. Points that do not fall directly on the trend line exhibit the fact that the dependent variable, in this case, the price, is influenced by more than just the independent variable, representing the passage of time. The error term stands for any influence being exerted on the price variable, such as changes in market sentiment. The two data points with the greatest distance from the trend line should be an equal distance from the trend line, representing the largest margin of error. If a model is heteroskedastic, a common problem in interpreting statistical models correctly, it refers to a condition in which the variance of the error term in a regression model varies widely.
-