

MACHINE LEARNING

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

1. 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

Answer: A) Least Square Error

2. 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

Answer: A) Linear regression is sensitive to outliers

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

- A) Positive
- B) Negative
- C) Zero
- D) Undefined

Answer: B) Negative

4. 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

Answer: B) Correlation

5. 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

Answers: C) Low bias and high variance

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

- A) Descriptive model
- B) Predictive model
- C) Reinforcement learning
- D) All of the above

Answers: A) Predictive model

7. Lasso and Ridge regression techniques belong to _____?

- A) Cross validation
- B) Removing outliers
- C) SMOTE
- D) Regularization

Answers: D) Regularization

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

- A) Cross validation
- B) Regularization
- C) Kernel
- D) SMOTE

Answers: D) SMOTE

9. 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

Answers: C) Sensitivity and Specificity

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

- A) True
- B) False

Answers: A) False

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

Answers: All the above (A, B, C, D)

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.

Answers: A) and B)

MACHINE LEARNING

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

13. Explain the term regularization?

Answers: Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted coefficient function and prevent overfitting or underfitting. 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 deals with unseen data by introducing noise in the output, and hence the model are called overfitted and underfitted. This problem can be deal with the help of a regularization technique.

14. Which particular algorithms are used for regularization?

Answers: There are mainly two types of regularization techniques, which are given below:

1. Ridge Regression
2. Lasso Regression

Ridge Regression used to reduce the overfitting in the model, and it includes all the features present in the model. It reduces the complexity of the model by shrinking the 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.

Lasso Regression will reduce the coefficient value to zero those features are not informative. 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 (i.e. models with fewer parameters). 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.

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

Answers: Error in linear regression equation is the difference between the actual value and predicted value and our goal is to reduce this difference. An error term essentially means that the model is not completely accurate and results in differing results during real-world applications.

Some of the common error models are:

MAE (Mean Absolute Error)

MSE (Mean Square Error)

RMSE (Root Mean Square Error)

RAE (Relative Absolute Error)

RSE (Relative Square Error)