

MACHINE LEARNING

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

Answer - C) Low bias and high variance

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

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

Answer - B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

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

Answer - D) Regularization

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

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

Answer - 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

Answer - A) TPR and FPR

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

- A) True B) False

Answer - B) 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

Answer - B) Apply PCA to project high dimensional data

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.

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

C) We need to iterate.

13. Explain the term regularization?

In Mathematics, Statistics, Finance, Computer Science, Particularly Machine Learning and inverse problems, Regularization is a process that changes the result answer to be simpler.

It is often used to obtain results for ill-posed problems or to prevent overfitting.

14. Which particular algorithms are used for regularization?

Regularization Techniques

Ridge Regularization

Also known as Ridge Regression, it adjusts models with overfitting or underfitting by adding a penalty equivalent to the sum of the squares of the magnitudes of the coefficients.

This means that the mathematical function representing our machine learning model is minimized and the coefficients are calculated. The size of the coefficients is multiplied and added. Ridge Regression performs regularization by reducing the coefficients present. The function shown below shows the cost function of the ridge regression.

Lasso Regularization

Modifies overfitted or under-fitted models by adding a penalty equivalent to the sum of the absolute values of the coefficients.

Lasso regression also performs coefficient minimization, but instead of squaring the magnitudes of the coefficients, it takes the actual values of the coefficients. This means that the sum of the coefficients can also be 0 because there are negative coefficients.

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

An error term represents the margin of error within a statistical model

It refers to sum of the deviations with in the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results.