

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

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

3. A line falls from left to right if a slope is _____?
A) Positive B) Negative
C) Zero D) Undefined

Answer- B

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

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- A

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

7. Lasso and Ridge regression techniques belong to _____?
A) Cross validation B) Removing outliers
C) SMOTE D) Regularization

Answer- D

8. To overcome with imbalance dataset which technique can be used?
A) Cross validation B) Regularization
C) Kernel D) SMOTE

Answer- D

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

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

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- A, B, C.

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 – A, B, C

MACHINE LEARNING

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

13. Explain the term regularization?

Answer –

Regularization is a technique that prevents model from overfitting by adding some extra information to it. It is a form of regression that compress the coefficient estimates towards zero, here we reduce the magnitude of the independent variable by keeping the same number of variables. This technique is very accurate and generalized. Overfitting means sometimes the model performs well on the training data but not on testing data, hence it effects the prediction of the output. This is overfitting and to overcome this situation we use regularization, which is of three types-

- Ridge
- Lasso
- Elastic net

14. Which particular algorithms are used for regularization?

Answer-

Various Algorithm used of Regression are

- Ridge Regression- it helps us to reduce only the overfitting problem in the model and all other features remain present in the model. It reduces the complexity of the model by compressing the coefficients. Ridge regression never sets value of the coefficient to absolute zero. In this regression we reduce the sum the sum of residual (cost function) with the help of lambda. Here we reduce the slope for the generalization of the model. It is also known as L2 norm.
- Lasso Regression- it help us to reduce the overfitting as well as automatic feature selection also. It makes coefficients to absolute zero. It is similar to ridge regression, the difference here is instead of squared magnitude we use absolute value of the magnitude of slope. Coefficient of less important feature gets converted to 0, it also compresses other co-efficient to reduce complexity. It is also known as L-1 norm.
- Elastic- net regression- it is also a type of regularized technique that linearly combines the L1 and L2 penalties of ridge and lasso regression.
L1 penalty is the penalty which we add to the co-efficient in cost function. i.e. absolute value of magnitude.
L2 penalty is the penalty which we add to the co-efficient in cost function. In ridge regression i.e. squared magnitude.

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

Answer-

During plotting of best fit line in linear regression, we come across many data which are far from the line, in that case residual or error is calculated to get the best fit. Basically error is the difference between the predicted value and the actual value. It is also called residual and its formula is

$$E = Y - \bar{Y}$$
