

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 - D

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 - C

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

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



ASSIGNMENT – 39

MACHINE LEARNING

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

13. Explain the term regularization?

Regularization is a technique used in machine learning to prevent overfitting by adding a penalty term to the loss function. The goal of regularization is to reduce the complexity of the model and simplify it by minimizing the magnitudes of the coefficients. Regularization helps to avoid overfitting by controlling the tradeoff between bias and variance. It adds a constraint to the optimization problem, which limits the range of the coefficients and ensures that the model generalizes well to unseen data.

14. Which particular algorithms are used for regularization?

Lasso and Ridge regression are two popular algorithms used for regularization. Lasso regression adds a penalty term to the loss function that is proportional to the absolute value of the coefficients, which results in sparse solutions (i.e., some coefficients are set to zero). Ridge regression adds a penalty term to the loss function that is proportional to the square of the coefficients, which shrinks the coefficients towards zero without setting them to zero. Both Lasso and Ridge regression help to reduce the variance of the model and improve its generalization performance.

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

The error present in the linear regression equation is the difference between the predicted value and the actual value of the dependent variable. In linear regression, we try to fit a line that minimizes the sum of squared errors between the predicted values and the actual values of the dependent variable. The error term represents the variability in the dependent variable that is not explained by the independent variable. The error term can be decomposed into two components: the systematic error, which is the part of the error that can be explained by the independent variable, and the random error, which is the part of the error that cannot be explained by the independent variable and is due to other factors that are not accounted for in the model.
