

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
Ans:- 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
Ans:- A
 3. A line falls from left to right if a slope is _____?
A) Positive B) Negative
C) Zero D) Undefined
Ans:- 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
Ans:- A
 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
Ans:- 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
Ans:- A
 7. Lasso and Ridge regression techniques belong to _____?
A) Cross validation B) Removing outliers
C) SMOTE D) Regularization
Ans:- D
 8. To overcome with imbalance dataset which technique can be used?
A) Cross validation B) Regularization
C) Kernel D) SMOTE
Ans:- 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
Ans:- A
 10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
A) True B) False
Ans:- A
 11. Pick the feature extraction from below:
A) Construction bag of words from a email
-

MACHINE LEARNING

- B) Apply PCA to project high dimensional data
- C) Removing stop words
- D) Forward selection

Ans:- 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.

Ans:- D and B

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

- 13. Explain the term regularization?
- 14. Which particular algorithms are used for regularization?
- 15. Explain the term error present in linear regression equation?

Qus:13- Explain the term regularization?

Ans:- Regularization is a technique used to reduce errors by fitting the function appropriately on the given training set and avoiding overfitting.

The commonly used regularization techniques are

- 1. Lasso Regularization – L1 Regularization
- 2. Ridge Regularization – L2 Regularization
- 3. Elastic Net Regularization – L1 and L2 Regularization

Qus:-14 Which particular algorithms are used for regularization?

Ans:- A regression model that uses the L2 regularization technique is called Ridge regression. Ridge regression adds the “ squared magnitude ” of the coefficient as a penalty term to the loss function (L). This model is a combination of L1 as well as L2 regularization.

Qus:-15 Explain the term error present in linear regression equation?

Ans:- In simpler terms, under the linear regression model, the error term explains why all the y values do not lie perfectly on the regression line. If the model were simply $Y = \alpha + \beta x$, the points would all lie exactly on the line. The residuals $e_i = y_i - \hat{y}_i$ are estimates of realizations of the error term for individual realizations of Y and x.
