

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
Ans) A) D) none of these
 3. A line falls from left to right if a slope is _____?
A) Positive B) Negative
C) Zero
Ans) A) D) Undefined
 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) 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
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) B)
 7. Lasso and Ridge regression techniques belong to _____?
A) Cross validation B) Removing outliers
C) SMOTE D) Regularization
Ans) B)
 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) C)
 10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
A) True B) False
Ans) B)
 11. Pick the feature extraction from below:
A) Construction bag of words from a email
B) Apply PCA to project high dimensional data
-

MACHINE LEARNING

- C) Removing stop words
- D) Forward selection

Ans)B)

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) A), B) and C

MACHINE LEARNING

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

Explain the term regularization?

Ans) **Regularization:** Regularization is a technique used in regression to reduce the complexity of the model and to shrink the coefficient of the independent features. It is used to calibrate machine learning model in order to minimize the adjusted loss function and prevent overfitting or underfitting.

Q14 Which particular algorithms are used for regularization?

Ans) **There are three main regularization techniques, namely:**

Ridge Regression (L2 Norm)

Lasso (L1 Norm)

Dropout.

Be it an over-fitting or under-fitting problem, it will lower down the overall performance of a machine learning model. To get the best out of machine learning models, you must optimize and tune them well.

Q15 Explain the term error present in linear regression equation?

Linear regression most often uses mean-square error (MSE) to calculate the error of the model. MSE is calculated by:

1-measuring the distance of the observed y-values from the predicted y-values at each value of x;

2-squaring each of these distances;

3-calculating the mean of each of the squared distances.

Linear regression fits a line to the data by finding the regression coefficient that results in the smallest MSE.
