

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: C

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: 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

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,C

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

13. Explain the term regularization?

ANS: Regularization is important concept in machine learning , as it avoid to prevent the model from overfitting by adding some penalties or we can say by adding some information to it. it is the technique to fit the model by reducing their errors.

There are 2 commonly used techniques in regularization:

- 1- lasso regularization (L1)
- 2- ridge regularization (L2)

Lasso and ridge can be used for any algorithms involving weight parameters including neural nets.

14. Which particular algorithms are used for regularization?

ANS: There are three main regularizations techniques:

- 1- Lasso (L1)
- 2- Ridge (L2)
- 3- Dropout

Lasso (L1) regularization used as it eliminates or treat as zero values for features which is least predictive to label. And minimises the loss to high treating values features which is most predictive to label. It uses absolute weights values for normalization .

Ridge (L2) regularization used as it gives minimum importance to those features which is least predictive to label. And minimises the loss to high treating values features which is most predictive to label. In this we add sum of weight's squares to a loss function and thus create a new loss function.

Dropout is primarily used in any kind of neural networks, eg-ANN, DNN, CNN,RNN to moderate learning . with dropouts you are left with a reduced network as dropped out neurons are left out during that training iteration. It decreased overfitting by avoiding training all the neurons on complete training data in one go. It also improves training speed

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

ANS: The term error is defined as the difference between expected things and actually observed things. And error in linear regression equation, I am defining as taking the example of model tracking stock's price over time. In this error is defined as expected price at a particular time and the price that was actually observed.

In this as where the price is exactly what was anticipated at a particular time, the price will fall on the trend line and the error term will be zero.