

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 - 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 – 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

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 - A

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 - 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.

Answer – A, B

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

13. Explain the term regularization?

Answer - It is one of the most important concepts of machine learning. This technique prevents the model from overfitting by adding extra information to it. It is a form of regression that shrinks the coefficient estimates towards zero. In other words, this technique forces us not to learn a more complex or flexible model, to avoid the problem of overfitting.

14. Which particular algorithms are used for regularization?

Answer - Particular algorithms are used for regularization - **Ridge Regression (L2 Norm) Lasso (L1 Norm) Dropout.**

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

Answer - The error term is the difference between the expected price at a particular time and the price that was actually observed.