

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) Least Square Error

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) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is _____?

- A) Positive B) Negative
- C) Zero D) Undefined

Ans. B) Negative

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) Correlation

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) Low bias and high variance

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) Descriptive model

7. Lasso and Ridge regression techniques belong to _____?

- A) Cross validation B) Removing outliers
- C) SMOTE D) Regularization

Ans. D) Regularization

8. To overcome with imbalance dataset which technique can be used?

- A) Cross validation B) Regularization
- C) Kernel D) SMOTE

Ans. A) Cross validation

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) Sensitivity and Specificity

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the

curve should be less.

True B) False

Ans. B) False

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) Apply PCA to project high dimensional data

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, C

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

13.Explain the term regularization?

Ans. Regularization is an application of Occam's Razor. It is one of the key concepts in Machine learning as it helps choose a simple model rather than a complex one. Regularization refers to the modifications that can be made to a learning algorithm that helps to reduce this generalization error and not the training error. It reduces by ignoring the less important features. It also helps prevent overfitting, making the model more robust and decreasing the complexity of a model.

The regularization techniques in machine learning are:

Lasso regression: having the L1 norm

Ridge regression: with the L2 norm

Elastic net regression: It is a combination of Ridge and Lasso regression.

14. Which particular algorithms are used for regularization?

Ans. The term 'regularization' refers to a set of techniques that regularizes learning from particular features for traditional algorithms or neurons in the case of neural network algorithms.

It normalizes and moderates weights attached to a feature or a neuron so that algorithms do not rely on just a few features or neurons to predict the result. This technique helps to avoid the problem of overfitting.

There are three main regularization techniques, namely:

Ridge Regression (L2 Norm)

Lasso (L1 Norm)

Dropout

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

Ans. The error term is also known as the residual, disturbance, or remainder term, and is variously represented in models by the letters e , ϵ , or u . An error term represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results. The regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent variable.