

MACHINE LEARNING

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

Ans-Least Square Error

2. Which of the following statement is true about outliers in linear regression?

Ans-Linear regression is sensitive to outliers

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

Ans-Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable?

Ans-Correlation

5. Which of the following is the reason for over fitting condition?

Ans-Low bias and high variance

6. If output involves label, then that model is called as:

Ans-Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

Ans-Regularization

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

Ans-SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

Ans-TPR and FPR

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

Ans-False

11. Pick the feature extraction from below:

Ans-Forward selection

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

- Ans- 1. We don't have to choose the learning rate.
2. It becomes slow when number of features is very large.
3. We need to iterate.

13. Explain the term regularization?

Ans- whenever we build the model, we don't know how it going to work. As it is seen that most of cases the model is build and trained and it is working at its best level or I can say its optimum level, but when that model is put into testing is don't perform good, the score we see in training is somehow get decreased in testing. To prevent and build the model which is neither overfitting nor the underfitting. For this we use the regularization technique.

14. Which particular algorithms are used for regularization?

Ans- The particular algorithms which are used for regularization is: -
LASSO and RIDGE.

LASSO: - or L1 form

Over fitting problem in model: -

1. Similar data at time of test which seen in training.
2. Model all the features to predict label, rather than taking the important feature which are cost effective to the model

In lasso we use to get the learning rate

For example: -

X features					Y label
Petrol, Deiseal, Colour, Seater, Brand					Milage of car
1. Yes	No	red	5	Honda	25 kmpl
2. No	Yes	blue	6	Maruti	21 kmpl
3. Yes	No	green	6	Audi	20kmpl

Here as we see the data of car and it shows the milage, but where the features, all the features are relevant or not.

There is one feature which is irrelevant in predicting the car milage which is colour. So, what we do here we just use the regularization, lasso algorithm which select the feature which is important for the prediction which affect the prediction, and just leave the irrelevant features. In nutshell we can say that lasso is work as feature selection.

RIDGE: - also called L2 form

It works on same principle as lasso, just the difference is lasso neglected the irrelevant feature while ridge just nullify it. As here in ridge algorithm gives the importance to the relevant features while it gives zero importance to irrelevant features.

To make this working we have to find the learning rate. Although both algorithm gives the learning rate.

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

Ans: -An error is a residual variable produced by the statistical or mathematical model which is created when model does not fully represent the actual relationship between the independent variable and dependent variable.

As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis. We can also say it like, an error term is the term in the model regression equation, the unexplained difference between the actual observed value of the independent variable and the result predicted by the model.