

MACHINE LEARNING^{[1][1]}_[SEP]

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?
A. Least Square Error
2. Which of the following statement is true about outliers in linear regression?
^{[1][1]}_[SEP]A) Linear regression is sensitive to outliers
3. A line falls from left to right if a slope is _____?
^{[1][1]}_[SEP]B) Negative
4. Which of the following will have symmetric relation between dependent variable and independent variable?
D) None of these
5. Which of the following is the reason for over fitting condition?^{[1][1]}_[SEP]
C) Low bias and high variance
6. If output involves label then that model is called as:^{[1][1]}_[SEP]
B) Predictive modal
7. Lasso and Ridge regression techniques belong to _____?

D) Regularization

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

D)SMOTE

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

B) False

11. Pick the feature extraction from below:^{[1][1]}_[SEP]

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

13. Explain the term regularization?

Regularization is a method whereby we reduce the problem of over-fitting and under-fitting. When the model is not very tuned to the training data set, it leads to errors and low accuracy of the model. Underfitting is mainly caused by high biasness and low variance in the model. Whereas Over-fitting is a case where the model is tuned to all the points in the training. Its only restricted to the training data set and it doesn't work well for actual data and causes errors in prediction. Over-fitting is mainly caused by high variance and low Bias. Both underfitting and over-fitting leads to wrong prediction and to overcome this issue we use regularization. There are 3 main tools involved in this :

- 1.LASSO
- 2.RIDGE
- 3.ELASTIC NET

14. Which particular algorithms are used for regularization?

There are 3 main algorithms involved in regularization, they are as follows:

- 1.LASSO
- 2.RIDGE
- 3.ELASTIC NET

1.LASSO: This technique will try and find out the optimal values of the variables which are impacting the output. The co-efficients which have a higher degree of positivity or negativity will be omitted internally.

2.RIDGE: This algorithm will reduce the variance present in the co-efficients difference

3.ELASTIC NET: This algorithm is a combination of both lasso and Ridge. It has both the properties of reducing the variance and omitting the variables.

All 3 algorithms work with the help of alpha, but changing the figures of alpha from 0.1 to 0.001 to 0.0001 one can determine the best score in each model and use that

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

The equation of linear regression is as follows:

$$Y=A+BX+C$$

C denotes the error.

In every model, some portion of the data goes for training and remaining go for testing (the proportion is roughly 70/30 or 80/20). The difference which is calculated while checking the actual figures and the figures which are produced by the model after training is called the error. The LEAST SQUARE ERROR method is used to calculate the error present in the model.

MACHINE VALUES	ACTUAL VALUES	DIFFERENCE	SQAURE
40	40	0	0
42	44	2	4
44	45	-1	1
45	45	0	0
46	47	1	1
		SUM	6

HERE the error is 6