FLIP ROBO

ASSIGNMENT-39

MACHINE LEARNING

In Q1 to Q11, only one optien la 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) Lasst Square Error

B) Maximum Lihood

C) Logarithmic Loss

D) Both A and B

Ans:-

A) Lasst Square Error

2. Which of the following statement in 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 C) Both of them

B) Correlation D) None of these

Ans:-
B) Correlation

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

A) High blas 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 C) Reinforcement learning

B) Predictive modal D) All of the above

Ans:-
B) Predictive modal

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

C) Kemel

D)SMOTE

B) Regularization

Ans:-
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

B) Sensitivity and precision

C) Sensitivity and Specificity

D) Recall and precision

Ans:-
A) TPR and FPR

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) 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:- Option A,B,C all 3 are Correct.

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

13. Explain the term regularization?

Ans:-

Regularization

Regularization is a technique used in regression to reduce the complexity of the model and to shrink the coefficients of the independent features.

“Everything should be made as simple as possible, but no simpler.” -Albert Einstein

In simple words, this technique converts a complex model into a simpler one, so as to avoid the risk of overfitting and shrinks the coefficients, for lesser computational cost.

In other words we can also say that this is a form of regression, that constrains/ regularizes or shrinks the coefficient estimates towards zero. This technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting.

14. Which particular algorithms are used for regularization?

Ans:-

The different Regularization algorithms are :

Ridge Regression
LASSO (Least Absolute Shrinkage and Selection Operator) Regression
Elastic-Net Regression

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

Ans:-

The term error is the stuff that isn't explained by the model.

The term error is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis.

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.

For a very simple example, suppose you are predicting the weight of adult human males based on their height. Well, height is certainly related to weight - taller people tend to be heavier - but the model won't be perfect because there is a range of weights at each height. The error is the difference between the predicted value and the actual value.

—---END—---