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 1 : 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

Answer 2 : 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

Answer 3: 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

Answer 4: 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

Answer 5: 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

Answer 6: Predictive modal

7. Lasso and Ridge regression techniques belong to_____?
   A) Cross validation B) Removing outliers
   C) SMOTE D) Regularization

Answer 7: Regularization

MACHINE LEARNING

8. To overcome with imbalance dataset which technique can be used?
   A) Cross validation
   B) Regularization
   C) Kernel
   D) SMOTE

Answer 8: 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

Answer 9: Sensitivity and Specificity

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

Answer 10: 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

Answer 11: Forward selection

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 12: (A) We don't have to choose the learning rate.
(B) It becomes slow when number of features is very large.

MACHINE LEARNING

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

13. Explain the term regularization?

Answer 13 : Regularization is a process of using techniques that are used to calibrate machine learning models to prevent overfitting or underfitting.

Regularization works by adding a penalty or complexity term to the complex model. Let's consider the simple linear regression equation:

$$y= a+b1x1+b2x2+b3x3+\cdots+bnxn + c$$

In the above equation, Y represents the value to be predicted

X1, X2, …Xn are the features for Y.

b1,…..bn are the weights or magnitude attached to the features, respectively. Here represents the bias of the model, and b represents the intercept.

Linear regression models try to optimize the a and c to minimize the cost function.

In this function we add a loss function and optimize parameter to make the model that can predict the accurate value of Y. The loss function for the linear regression is called as **RSS or Residual sum of squares.**

There are mainly two types of regularization techniques, which are given below:

- **Ridge Regression**
- **Lasso Regression**

14. Which particular algorithms are used for regularization?

Answer 14 . Mainly following two algorithms are used for regularization:

- **Ridge Regression**
- **Lasso Regression**

- Ridge regression is linear regression in which a small amount of bias is added to get better long-term predictions.
- Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as **L2 regularization**.

MACHINE LEARNING

- In this technique, the cost function is altered by adding the penalty term to it. The amount of bias added to the model is called **Ridge Regression penalty**. We can calculate it by multiplying with the lambda to the squared weight of each individual feature.
- The equation for the cost function in ridge regression will be:

$$\sum_{i=1}^{M}(y_i - y'_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{n}\beta_j * x_{ij}\right)^2 + \lambda\sum_{j=0}^{n}{\beta_j}^2$$

- In the above equation, the penalty term regularizes the coefficients of the model, and hence ridge regression reduces the amplitudes of the coefficients that decreases the complexity of the model.
- As we can see from the above equation, if the values of **λ tend to zero, the equation becomes the cost function of the linear regression model.** Hence, for the minimum value of λ, the model will resemble the linear regression model.
- A general linear or polynomial regression will fail if there is high collinearity between the independent variables, so to solve such problems, Ridge regression can be used.
- It helps to solve the problems if we have more parameters than samples.

- Lasso regression is another regularization technique to reduce the complexity of the model. It stands for **Least Absolute and Selection Operator.**
- It is similar to the Ridge Regression except that the penalty term contains only the absolute weights instead of a square of weights.
- Since it takes absolute values, hence, it can shrink the slope to 0, whereas Ridge Regression can only shrink it near to 0.
- It is also called as **L1 regularization.** The equation for the cost function of Lasso regression will be:

$$\sum_{i=1}^{M}(y_i - y'_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{n}\beta_j * x_{ij}\right)^2 + \lambda\sum_{j=0}^{n}|\beta_j|^{\square}$$

- Some of the features in this technique are completely neglected for model evaluation.
- Hence, the Lasso regression can help us to reduce the overfitting in the model as well as the feature selection.

MACHINE LEARNING

The main difference between the above two is : Ridge regression is mostly used to reduce the overfitting in the model, and it includes all the features present in the model. It reduces the complexity of the model by shrinking the coefficients.

Where as Lasso regression helps to reduce the overfitting in the model as well as feature selection.

Both the algorithms are very useful and can be used based on their model score on different data sets.

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

Answer 15: Basic equation of Linear Regression is as under :

Y=a+bx+e

Here y is the target variable for which the value need to be predicted.

a is intercept which specifies from which point the regression line starts.

b is coefficient which specifies the amount what gets increased in the value for every unit increased in x or input variable.

x is input variable for which value of y is to be predicted.

e is the error which comes in the prediction.

Error is the calculated value which may be either positive or negative and the same value is added or deduction from the result of the equation to get best prediction.

The common method used to get the value of the error is **Least Squares Error (Method).**

Following is the formula of Least Squares Error :

$$S = \sum_{i=1}^{n} {d_i}^2$$

$$S = \sum_{i=1}^{n} [y_i - f_{x_i}]^2$$

$$S = {d_1}^2 + {d_2}^2 + {d_3}^2 + \cdots + {d_n}^2$$

It is basically the sum of squares of all the differences of actual and predicted values of a particular data set.

Submitted by : Amit Puri Goswami
Batch No. : 1843
Project: Internship 30