

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
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
3. A line falls from left to right if a slope is _____?
A) Positive B) **Negative**
C) Zero D) Undefined
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
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
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
7. Lasso and Ridge regression techniques belong to _____?
A) Cross validation B) Removing outliers
C) SMOTE D) **Regularization**
8. To overcome with imbalance dataset which technique can be used?
A) **Cross validation** B) Regularization
C) Kernel 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
10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
A) **True** 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**

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.
-

MACHINE LEARNING

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

13. Explain the term regularization?
14. Which particular algorithms are used for regularization?
15. Explain the term error present in linear regression equation?

13. Explain the term regularization?

Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it. Sometimes the machine learning model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when deals with unseen data by introducing noise in the output, and hence the model is called Overfitted. This problem can be deal with the help of a regularization technique.

This technique can be used in such a way that it will allow to maintain all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy as well as a generalization of the model. It mainly regularizes or reduces the coefficient of features toward zero. In simple words, “In regularization technique, we reduce the magnitude of the features by keeping the same number of features.”

14. Which particular algorithms are used for regularization?

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

1. **Ridge Regression**
2. **Lasso Regression**

Ridge Regression:-

- ✓ Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can 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.
- ✓ 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.
-

MACHINE LEARNING

Lasso Regression:-

- ✓ 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:

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

- ✓ 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.

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

An error term 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 .

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.

An error term essentially means that the model is not completely accurate and results in differing results during real-world applications. For example, assume there is a multiple linear regression function that takes the following form:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Dependent Variable
Population Y intercept
Population Slope Coefficient
Independent Variable
Random Error term

Linear component
Random Error component

Within a linear regression model tracking a stock's price over time, the error term is the difference between the expected price at a particular time and the price that was actually observed. In instances where the price is exactly what was anticipated at a particular time, the price will fall on the trend line and the error term will be zero.