

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: A) 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: 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

Answer: B) 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: 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

Answer: C) 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: B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

- A) Cross validation B) Removing outliers
- C) SMOTE D) Regularization

Answer: D) Regularization

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

- A) Cross validation B) Regularization
- C) Kernel D) SMOTE

Answer: 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

Answer: 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

Answer: A) True

11. Pick the feature extraction from below:

- A) Construction bag of words from an email
B) Apply PCA to project high dimensional data
C) Removing stop words
D) Forward selection

Answer: 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.

Answer: A) We don't have to choose the learning rate.

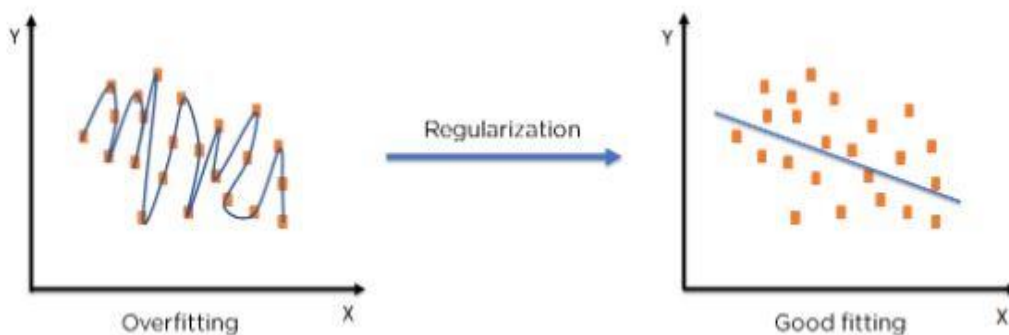
B) It becomes slow when number of features is very large.

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

13. Explain the term regularization?

Answer:

Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or under fitting.



Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

Regularization Techniques

There are two main types of regularization techniques:

Ridge Regularization and Lasso Regularization.

Ridge Regularization:

Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values being far away from the actual values.

Advantages:

Ridge Regression solves the problem of overfitting, as just regular squared error regression fails to recognize the less important features and uses all of them, leading to overfitting. Ridge regression adds a slight bias, to fit the model according to the true values of the data.

Lasso Regularization:

Lasso regression is also called Penalized regression method. This method is usually used in machine learning for the selection of the subset of variables. It provides greater prediction accuracy as compared to other regression models. Lasso Regularization helps to increase model interpretation.

Advantages:

In statistics and machine learning, lasso is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model.

14. Which particular algorithms are used for regularization?

Answer:

Regularization Algorithms:

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

Ridge Regression:

Ridge regression is a method for analyzing data that suffer from multi-collinearity.

$$Loss = \sum_{i=1}^n (y_i - (w_i x_i + c))^2 + \lambda \sum_{i=1}^n w_i^2$$

Loss Function for Ridge Regression

Ridge regression adds a penalty (*L2 penalty*) to the loss function that is equivalent to the square of the magnitude of the coefficients.

The regularization parameter (λ) regularizes the coefficients such that if the coefficients take large values, the loss function is penalized.

- $\lambda \rightarrow 0$, the penalty term has no effect, and the estimates produced by ridge regression will be equal to least-squares i.e. the loss function resembles the loss function of the Linear Regression algorithm. Hence, a lower value of λ will resemble a model close to the linear regression model.
- $\lambda \rightarrow \infty$, the impact of the shrinkage penalty grows, and the ridge regression coefficient estimates will approach zero (coefficients are close to zero, but not zero).

LASSO Regression:

LASSO is a regression analysis method that performs both feature selection and regularization in order to enhance the prediction accuracy of the model.

$$Loss = \sum_{i=1}^n (y_i - (w_i x_i + c))^2 + \lambda \sum_{i=1}^n |w_i|$$

Loss Function for LASSO Regression

LASSO regression adds a penalty (*L1 penalty*) to the loss function that is equivalent to the magnitude of the coefficients.

In LASSO regression, the penalty has the effect of forcing some of the coefficient estimates to be exactly equal to zero when the regularization parameter λ is sufficiently large.

LASSO regression converts coefficients of less important features to zero, which indeed helps in feature selection, and it shrinks the coefficients of remaining features to reduce the model complexity, hence avoiding overfitting.

Elastic-Net Regression:

Elastic-Net is a regularized regression method that linearly combines the L1 and L2 penalties of the LASSO and Ridge methods respectively.

$$Loss = \sum_{i=0}^n (y_i - (w_i x_i + c))^2 + \lambda_1 \sum_{i=0}^n |w_i| + \lambda_2 \sum_{i=0}^n w_i^2$$

Loss Function for Elastic-Net Regression

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

Answer:

Regression analysis is a statistical technique that can test the hypothesis that a variable is dependent upon one or more other variables. Further, regression analysis can provide an estimate of the magnitude of the impact of a change in one variable on another. This last feature, of course, is all important in predicting future values.

Regression analysis is based upon a functional relationship among variables and further, assumes that the relationship is linear. This linearity assumption is required because, for the most part, the theoretical statistical properties of non-linear estimation are not well worked.

The general linear regression model can be stated by the equation:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \epsilon_i$$

Where β_0 is the intercept, β_i 's are the slope between Y and the appropriate X_i , and ϵ (pronounced epsilon), is the error term that captures errors in measurement of Y and the effect on Y of any variables missing from the equation that would contribute to explaining variations in Y.

This equation is the theoretical population equation and therefore uses Greek letters. The equation we will estimate will have the Roman equivalent symbols. This is parallel to how we kept track of the population parameters and sample parameters before. The symbol for the population mean was μ and for the sample mean \bar{X} and for the population standard deviation was σ and for the sample standard deviation was S. The equation that will be estimated with a sample of data for two independent variables will thus be:

$$Y_i = b_0 + b_1x_{1i} + b_2x_{2i} + e_i$$

These are that the Y is normally distributed, the errors are also normally distributed with a mean of zero and a constant standard deviation, and that the error terms are independent of the size of X and independent of each other.