ASSIGNMENT-39

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: A) Cross validation

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: 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

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

13. Explain the term regularization?

Answer: Regularization is **a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting**. In other words Regularization is a technique used in regression to reduce the complexity of the model and to shrink the coefficients of the independent features. 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.

14. Which particular algorithms are used for regularization?

Answer: The commonly used regularization algorithms are as follows:

L1 Regularization or LASSO (Least Absolute Shrinkage and Selection Operator) Regularization

L2 Regularization or Ridge Regularization.

L1 and L2 regularization are the best ways to manage overfitting and perform feature selection when you've got a large set of features. When you have a large number of features in your data set, you may wish to create a less complex, more parsimonious model. L1 Regularization, also called a lasso regression, adds the "absolute value of magnitude" of the coefficient as a penalty term to the loss function. L2 Regularization, also called a ridge regression, adds the "squared magnitude" of the coefficient as the penalty term to the loss function. A regression model that uses the L1 regularization technique is called **lasso regression** and a model that uses the L2 is called **ridge regression.**

Working of Ridge and LASSO Regression:

The working of all these algorithms is quite similar to that of Linear Regression, it's just the loss function that keeps on changing!

$$Loss = \sum_{i-1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i-1}^{n} (y_i - (w_i$$

(Loss Function for Linear Regression)

Ridge Regression (also known as L2 Regularization, L2 penalty)

Ridge regression or L2 Regularization 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.

- ***λ → 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.

- ***λ → ∞,*** the impact of the shrinkage penalty grows, and the ridge regression coefficient estimates will **approach zero** (coefficients are close to zero, but not zero).

To sum up, **Ridge regression shrinks the coefficients as it helps to reduce the model complexity and multi-collinearity.**

LASSO Regression (also known as L1 Regularization, L1 penalty)

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.

To sum up, **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.**

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

Answer: The standard error of the regression (S), also known as the standard error of the estimate, represents the average distance that the observed values fall from the regression line. Conveniently, it tells you how wrong the regression model is on average using the units of the response variable. 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 appears in a statistical model, like a regression model, to indicate the uncertainty in the model. The error term is a residual variable that accounts for a lack of perfect goodness of fit. Heteroskedastic refers to a condition in which the variance of the residual term, or error term, in a regression model varies widely.