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**

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**

3. A line falls from left to right if a slope is _____?

✓A) Positive   B) Negative
C) Zero   D) Undefined   **ANSWER -A**

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- C**

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**

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**

7. Lasso and Ridge regression techniques belong to _____?

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

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

A) Cross validation   B) Regularization
C) Kernel   ✓ D) SMOTE   **ANSWER- D**

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**

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**

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**

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,B & C

13. **Explain the term regularization**?

ANSWER

 It is one of the most important concepts of machine learning. This technique prevents 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. 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.

Regularization woking process

Regularization works by adding a penalty or complexity term or shrinkage term with Residual Sum of Squares (RSS) to the complex model.

Let's consider the Simple linear regression equation:

Here Y represents the dependent feature or response which is the learned relation. Then,

Y is approximated to $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$

Here, X1, X2, …Xp are the independent features or predictors for Y, and

$\beta_0, \beta_1, \ldots \beta_n$ represents the coefficients estimates for different variables or predictors(X), which describes the weights or magnitude attached to the features, respectively.

In simple linear regression, our optimization function or loss function is known as the residual sum of squares (RSS).

Techniques of Regularization

there are two types of regularization techniques

- Ridge Regression
- Lasso Regression

Ridge Regression - Ridge regression is one of the types of linear regression in which we introduce a small amount of bias, known as Ridge regression penalty so that we can get better long-term predictions.

Lasso Regression- Lasso regression is another variant of the regularization technique used to reduce the complexity of the model. It stands for Least Absolute and Selection Operator.

14. Which particular algorithms are used for regularization?

ANSWER

The term 'regularization' refers to a set of techniques that regularizes learning from particular features for traditional algorithms or neurons in the case of neural network algorithms.

There are three main regularization techniques, namely:

- Ridge Regression (L2 Norm)
- Lasso (L1 Norm)
- Dropout

Ridge and Lasso can be used for any algorithms involving weight parameters, including neural nets. Dropout is primarily used in any kind of neural networks e.g. ANN, DNN, CNN or RNN to moderate the learning. Let's take a closer look at each of the techniques

Ridge Regression (L2 Regularization)

Ridge regression is also called L2 norm or regularization.

Now the parameters are learned using a modified loss function. To minimize the above function, parameters need to be as small as possible. Thus, L2 norm prevents weights from rising too high.

Lasso(L1 Norm)

In ridge regression, loss function along with the optimization algorithm brings parameters near to zero but not actually zero, while lasso eliminates less important features and sets respective weight values to zero. Thus, lasso also performs feature selection along with regularization.

Dropout

Dropout is a regularization technique used in neural networks. It prevents complex co-adaptations from other neurons.

In neural nets, fully connected layers are more prone to overfit on training data. Using dropout, you can drop connections with 1-p probability for each of the specified layers. Where p is called keep probability parameter and which needs to be tuned.

Along with Dropout, neural networks can be regularized also using L1 and L2 norms. Apart from that, if you are working on an image dataset, image augmentation can also be used as a regularization method.

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

ANSWER

In almost all the data mining techniques, we try to find out the closer estimates of a variable we want to predict. In regression models, we assume that the relation between the response variable and predictors to be linear and we find out a straight line that fits the relation well.

The error term in regression is a catch-all for what we miss out with this model, because in reality

- The true relation may not be linear

- There may be other variables not included in the model that cause variation in response variable

- There may be measurement errors in the observations

The error is calculated as the difference between actual and estimated value of the response.

Linear regression most often uses mean-square error (MSE) to calculate the error of the model. MSE is calculated by:

- measuring the distance of the observed y-values from the predicted y-values at each value of x;
- squaring each of these distances;
- calculating the mean of each of the squared distances.

Linear regression fits a line to the data by finding the regression coefficient that results in the smallest MSE.