MACHINELEARNING

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
C) Zero?
B) Negative
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 model
C) Reinforcement learning
D) All of the above

7.Las 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.

MACHINELEARNING

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?

Q13. Explain the term regularization?

Ans:- The word regularize means to make things regular or acceptable. This is exactly why we use it for. Regularizations are techniques used to reduce the error by fitting a function appropriately on the given training set and avoid overfitting.

In the context of machine learning, regularization is the process which regularizes or shrinks the coefficients towards zero. In simple words, regularization discourages learning a more complex or flexible model, to prevent overfitting.

Types of regularization

L2 and L1 are the most common types of regularization. Regularization works on the premise that smaller weights lead to simpler models which in results helps in avoiding overfitting. So to obtain a smaller weight matrix, these techniques add a 'regularization term' along with the loss to obtain the cost function.

Q14. Which particular algorithms are used for regularization?

Ans:- In the context of machine learning, regularization is the process which regularizes or shrinks the coefficients towards zero. In simple words, regularization discourages learning a more complex or flexible model, to prevent overfitting.

There are three main regularization techniques, namely: Ridge Regression (L2 Norm) Lasso (L1 Norm) Dropout.

There are three main regularization techniques, namely:

1. Ridge Regression (L2 Norm)

2. Lasso (L1 Norm)

3. 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.

Ridge Regression (L2 Regularization)

Ridge regression is also called L2 norm or regularization.

When using this technique, we add the sum of weight's square to a loss function and thus create a new loss function which is denoted thus:

$$\text{Loss} = \sum_{j=1}^{m} \left(Yi - Wo - \sum_{i=1}^{n} Wi\, Xji \right)^2 + \lambda \sum_{i=1}^{n} Wi^2$$

As seen above, the original loss function is modified by adding normalized weights. Here normalized weights are in the form of squares.

Lasso Regression (L1 Regularization)

Also called lasso regression and denoted as below:

$$\text{Loss} = \sum_{j=1}^{m} \left(Yi - Wo - \sum_{i=1}^{n} Wi\, Xji \right)^2 + \lambda \sum_{i=1}^{n} |Wi|$$

This technique is different from ridge regression as it uses absolute weight values for normalization. λ is again a tuning parameter and behaves in the same as it does when using ridge regression.

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 over fit 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.

(a) Standard Neural Net   (b) After applying dropout.
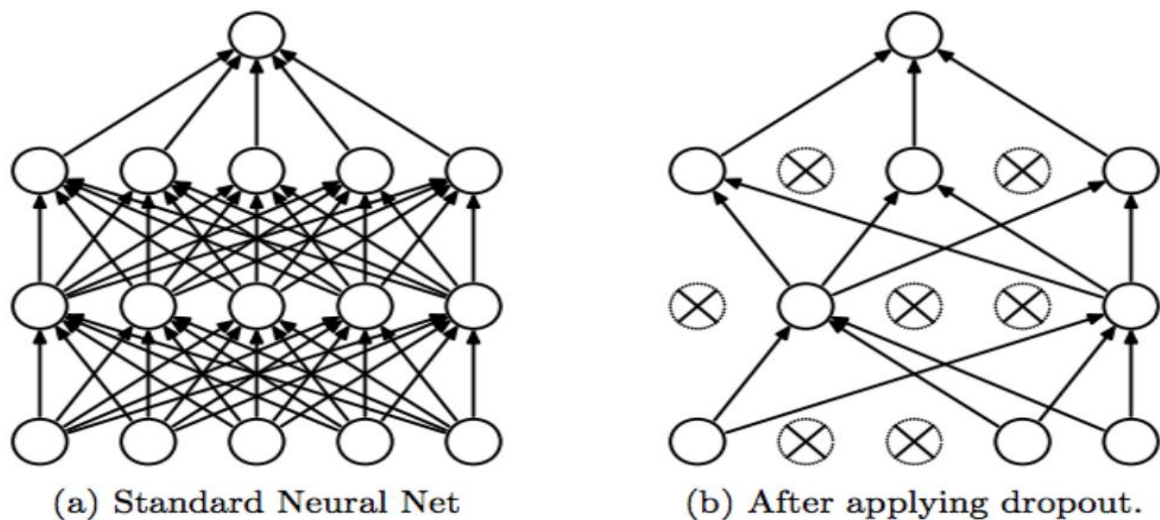
With dropout, you are left with a reduced network as dropped out neurons are left out during that training iteration.

Q15. Explain the term error present in linear regression equation?

Ans:- 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 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.

Linear Regression, Error Term
Linear regression is a form of analysis that relates to current trends experienced by a particular security or index by providing a relationship between a dependent and independent variables, such as the price of a security and the passage of time, resulting in a trend line that can be used as a predictive model.