

## MACHINE LEARNING

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

Answer: -A) Least Square Error

In linear regression, we try to minimize the least square errors of the model to identify the line of best fit.

2. Which of the following statement is true about outliers in linear regression?

Answer: -A) Linear regression is sensitive to outliers

The slope of the regression line will change due to outliers in most of the cases. So Linear Regression is sensitive to outliers.

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

Answer: - B) Negative

If the graph of a line rises from left to right, the slope is positive. If the graph of the line falls from left to right the slope is negative.

4. Which of the following will have symmetric relation between dependent variable and independent variable?

Answer: - B) Correlation

5. Which of the following is the reason for over fitting condition?

Answer: - C) Low bias and high variance.

6. If output involves label, then that model is called as:

Answer: - B) Predictive modal

7. Lasso and Ridge regression techniques belong to \_\_\_\_\_?

Answer: - D) Regularization

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

Answer:-D) SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses \_\_\_\_\_ to make graph?

Answer: - A) TPR and FPR

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

Answer: - A) True

11. Pick the feature extraction from below:

Answer: - B) Apply PCA to project high dimensional data.

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

Answer: - 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.

### 13. Explain the term regularization?

Answer: - 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

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?

Answer:-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.

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:

$$\text{Loss} = \sum_{j=1}^m \left( Y_i - W_0 - \sum_{i=1}^n W_i X_{ji} \right)^2 + \lambda \sum_{i=1}^n W_i^2$$

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

You may have noticed parameters  $\lambda$  along with normalized weights.  $\lambda$  is the parameter that needs to be tuned using a cross-validation dataset. When you use  $\lambda=0$ , it returns the residual sum of square as loss function which you chose initially. For a very high value of  $\lambda$ , loss will ignore core loss function and minimize weight's square and will end up taking the parameters' value as zero.

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 Regression (L1 Regularization)**

Also called lasso regression and denoted as below:

$$\text{Loss} = \sum_{j=1}^m \left( Y_i - W_o - \sum_{i=1}^n W_i X_{ji} \right)^2 + \lambda \sum_{i=1}^n |W_i|$$

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

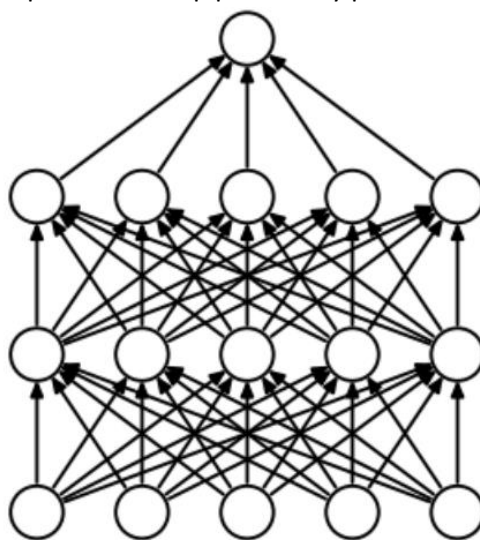
As loss function only considers absolute weights, optimization algorithms penalize higher weight values.

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.

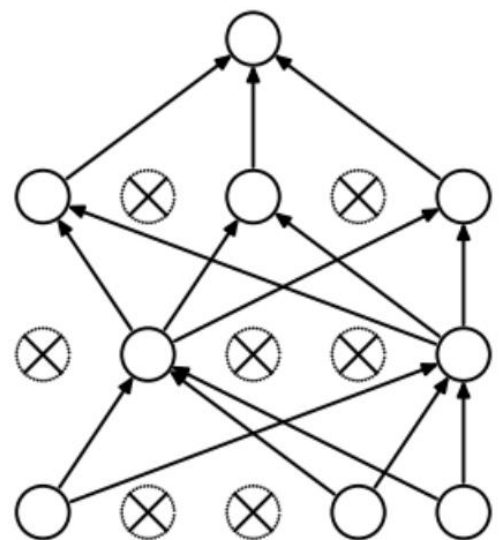
### Dropout

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.



(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.

Dropout decreases overfitting by avoiding training all the neurons on the complete training data in one go. It also improves training speed and learns more robust internal functions that generalize better on unseen data. However, it is important to note that Dropout takes more epochs to train compared to training without Dropout (If you have 10000 observations in your training data, then using 10000 examples for training is considered as 1 epoch).

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:- 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.

Points that do not fall directly on the trend line exhibit the fact that the dependent variable, in this case, the price, is influenced by more than just the independent variable, representing the passage of time. The error term stands for any influence being exerted on the price variable, such as changes in market sentiment.

The two data points with the greatest distance from the trend line should be an equal distance from the trend line, representing the largest margin of error.

If a model is heteroskedastic, a common problem in interpreting statistical models correctly, it refers to a condition in which the variance of the error term in a regression model varies widely.

**Linear Regression, Error Term, and Stock Analysis**

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.

A linear regression exhibits less delay than that experienced with a moving average, as the line is fit to the data points instead of based on the averages within the data. This allows the line to change more quickly and dramatically than a line based on numerical averaging of the available data points.