

## MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:

**Correct answer- dark black**

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
- B) Negative**
- C) Zero
- 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 modal**
- C) Reinforcement learning
- D) All of the above

7. 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) Kerne
- 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.**

13. Explain the term regularization?

**Ans-**

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.

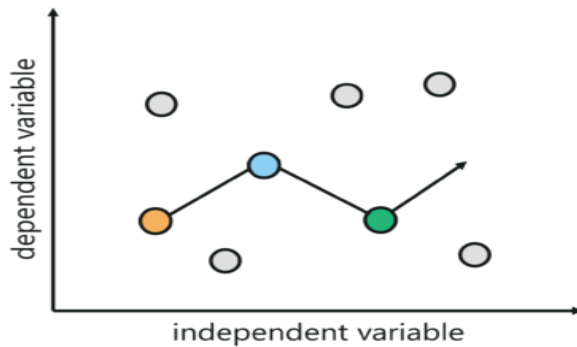
It normalizes and moderates weights attached to a feature or a neuron so that algorithms do not rely on just a few features or neurons to predict the result. This technique helps to avoid the problem of overfitting.

To understand regularization, let's consider a simple case of linear regression. Mathematically, linear regression is stated as below:

$$y = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n$$

There are two main regularization techniques,

- 1 Ridge Regression
- 2 Lasso Regression



14. Which particular algorithms are used for regularization?

**Ans-**

There are three main regularization techniques, namely:

1. Ridge Regression (L2 Norm)
2. Lasso (L1 Norm)
3. Dropout

## 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 this.

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

## Lasso Regression (L1 Regularization)

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.

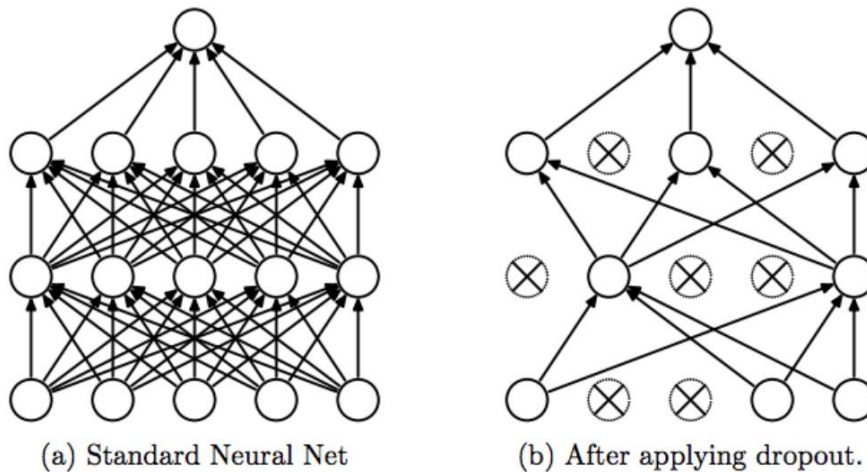
In ridge regression, loss function along with the optimization algorithm brings parameters near to zero but not actually zero, while lasso eliminates less important features and sets respective weight values to zero. Thus, lasso also performs feature selection along with regularization.

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

## 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.

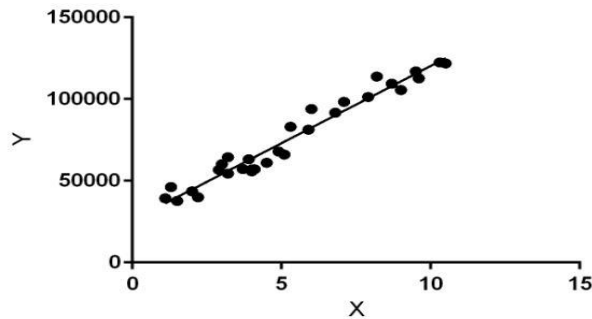


15 Explain the term error present in linear regression equation.

It is often said that the error term in a regression equation represents the effect of the variables that were omitted from the equation. This is unsatisfactory, even in simple contexts, as the following discussion should indicate. Suppose subjects are IID, and all variables are jointly normal with expectation 0. Suppose the explanatory variables have variance 1. The explanatory variables may be correlated amongst themselves, but anyp of them have a non-singularp-dimensional distribution. The parameters  $\alpha_j$  are real.

$$Y_i = \sum_{j=1}^{\infty} \alpha_j X_{ij}$$

For each  $p = 1, 2, \dots$ , consider the regression model



Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression. In the figure above, X (input) is the work experience and Y (output) is the salary of a person. The regression line is the best fit line for our model.

Residual/Error = Actual values – Predicted Values

Sum of Residuals/Errors = Sum(Actual- Predicted Values)

Square of Sum of Residuals/Errors = (Sum(Actual- Predicted Values))<sup>2</sup>

$$\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$$