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 - 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

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) D) None of these

Answer - B (Correlation)

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 model
C) Reinforcement learning
D) All of the above

Answer - B (Predictive model)

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 data set which technique can be used?
A) Cross validation
B) Regularization
C) Kernel

D) SMOTE

Answer - 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) 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) 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 - A) Construction bag of words from a email
B) Apply PCA to project high dimensional data
C) Removing stop words

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 in regression to reduce the complexity of the model and to shrink the coefficients of the independent features.
In simple words, 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.

This is a form of regression, that constrains/ regularizes or shrinks the coefficient estimates towards zero. In other words, this technique discourages learning a more complex or flexible model, so as to avoid the risk of over fitting.

A simple relation for linear regression looks like this. Here Y represents the learned relation and β represents the coefficient estimates for different variables or predictors(X).

$$Y \approx \beta 0 + \beta 1X1 + \beta 2X2 + \ldots + \beta pXp$$

The fitting procedure involves a loss function, known as residual sum of squares or RSS. The coefficients are chosen, such that they minimize this loss function.

- It is one of the most important concepts of machine learning. This technique prevents the model from over fitting by adding extra information to it.
- It is a form of regression that shrinks the coefficient estimates towards zero. In other words, this technique forces us not to learn a more complex or flexible model, to avoid the problem of over fitting.
- In the Regularization technique, we reduce the magnitude of the independent variables by keeping the same number of variables". It maintains accuracy as well as a generalization of the model.

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

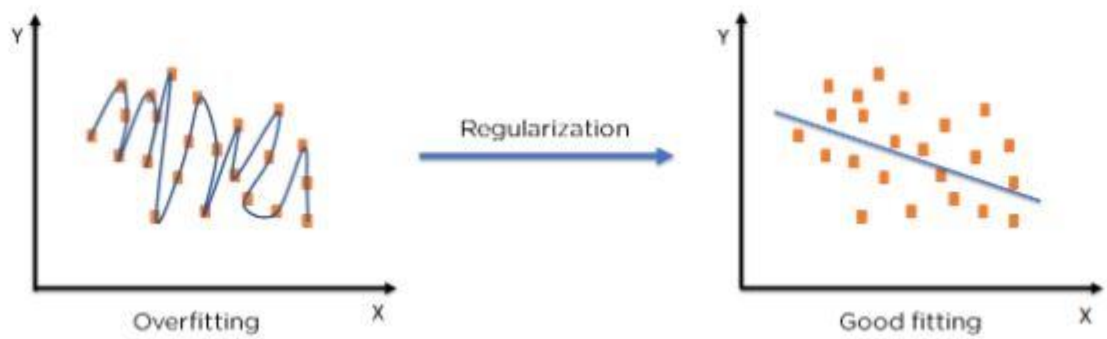
Let's consider the Simple linear regression equation:

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

Y is approximated to β0 + β1X1 + β2X2 + …+ βpXp

Here, X1, X2, …Xp are the independent features or predictors for Y, and β0, β1,…..βn 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).

- Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent over fitting or under fitting.



Using Regularization,we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

14. Which particular algorithms are used for regularization?

Answer - Basically there are mainly three algorithms are there which are used for regularization. They are follows -

1. Ridge Regression
2. LASSO (Least Absolute Shrinkage and Selection Operator) Regression
3. Elastic-Net 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!

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

Ridge Regression -

Ridge regression is a method for analyzing data that suffer from multi-col linearity.

$$Loss = \sum_{i=1}^{n} (y_i - (w_i x_i + c))^2 + \lambda \sum_{i=1}^{n} w_i^2$$

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 coeffiicient estimates will approach zero (coefficients are close to zero, but not zero).

Note: Ridge regression is also known as the L2 Regularization.

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

LASSO Regression -

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.

Note: LASSO regression is also known as the L1 Regularization (L1 penalty).

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 over fitting.

Elastic-Net Regression

Elastic-Net is a regularized regression method that linearly combines the L1 and L2 penalties of the LASSO and Ridge methods
respectively

$$Loss = \sum_{i=0}^{n} (y_i - (w_i x_i + c))^2 + \lambda_1 \sum_{i=0}^{n} |w_i| + \lambda_2 \sum_{i=0}^{n} w_i^2$$

Loss Function for Elastic-Net Regression

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

Answer - 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 regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent 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.

Formulation -
An error term essentially means that the model is not completely accurate and results in differing results during real-world applications. For example, assume there is a multiple linear regression function that takes the following form :

Y = $\alpha X + \beta p + \epsilon$

Where :

α, β = Constant parameters

X, p = Independent variables

ϵ = *Error term*

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

1. measuring the distance of the observed y-values from the predicted y-values at each value of x.
2. squaring each of these distances.
3. 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.