**Unit-3**

**Q.1] Define Linear Regression and explain different terminologies Related to the Regression Analysis.**

* **Linear Regression:** Linear regression is a supervised learning algorithm used to predict a continuous target variable Y based on one or more predictor variables X. It assumes that there is a linear relationship between the predictor variables and the target variable. The goal is to find the best fitting line (or hyperplane in multiple dimensions) that minimizes the difference between the predicted values and the actual values. This line can then be used to make predictions on new data. Linear regression can be used for simple linear regression (one predictor variable) or multiple linear regression (more than one predictor variable).
* **Different terminologies Related to the Regression Analysis:**
* **Ependent Variable**: The main factor in Regression analysis which we want to predict or understand is called the dependent variable. It is also called target variable.
* **Independent Variable:** The factors which affect the dependent variables or which are used to predict the values of the dependent variables are called independent variable, also called as a predictor.
* **Outliers:** Outlier is an observation which contains either very low value or very high value in comparison to other observed values. An outlier may hamper the result, so it should be avoided.
* **Multicollinearity:** If the independent variables are highly correlated with each other than other variables, then such condition is called Multicollinearity. It should not be present in the dataset, because it creates problem while ranking the most affecting variable.
* **Underfitting and Overfitting:** If our algorithm works well with the training dataset but not well with test dataset, then such problem is called Overfitting. And if our algorithm does not perform well even with training dataset, then such problem is called underfitting.

**Q. 2] Describe different types of Linear Regression.**

* Based on the number of independent variables, there are two types of linear regression.

1. **Simple Linear Regression:**

|  |
| --- |
| https://www.gatevidyalay.com/wp-content/uploads/2020/02/Simple-Linear-Regression-Case-01.png |

* In simple linear regression, the dependent variable depends only on a single independent variable. For simple linear regression, the form of the model is
* **Y = β0 + β1X or Y=mx+c**

Were, Y is a dependent variable.

* X is an independent variable.
* β0 or c and β1 or m are the regression coefficients.

|  |
| --- |
| https://www.gatevidyalay.com/wp-content/uploads/2020/02/Simple-Linear-Regression-Case-02-1.png |

* β0 or c is the intercept or the bias that fixes the offset to a line.
* β1 or m is the slope or weight that specifies the factor by which X has an impact on Y.

There are following 3 cases possible

* **Case-01:** β1 < 0

|  |
| --- |
| https://www.gatevidyalay.com/wp-content/uploads/2020/02/Simple-Linear-Regression-Case-03-1.png |

• It indicates that variable X has negative impact on Y.

• If X increases, Y will decrease and vice-versa.

* **Case-02:** β1 = 0

• It indicates that variable X has no impact on variable Y.

• If X changes, there will be no change in Y.

* **Case-03:** β1 > 0

•It indicates that variable X has positive impact on Y.

•If X increases, Y will increase and vice-versa.

1. **Multiple Linear Regression:** In multiple linear regression, the dependent variable depends on more than one independent variables. For multiple linear regression, the form of the model is

**Y = β0 + β1X1 + β2X2 + β3X3 + …… + βnXn**

Here, Y is a dependent variable.

•X1 , X2 , …., Xn are independent variables.

•β0 , β1 ,…, βn are the regression coefficients.

•βj (1<=j<=n) is the slope or weight that specifies the factor by which Xj has an impact on Y

**Q.3] Explain Regression coefficients for Linear Regression**

* Regression coefficients can be defined as estimates of some unknown parameters to describe the relationship between a predictor variable and the corresponding response.
* In other words, regression coefficients are used to predict the value of an unknown variable using a known variable.
* Linear regression is used to quantify how a unit change in an independent variable causes an effect in the dependent variable by determining the equation of the bestfitted straight line.
* This process is known as regression analysis..

|  |
| --- |
|  |

* The goal of linear regression is to find the equation of the straight line that best describes the relationship between two or more variables. • For example, suppose a simple regression equation is given by y = 7x - 3, then 7 is the coefficient, x is the predictor and -3 is the constant term. • Suppose the equation of the best-fitted line is given by Y = aX + b then, the regression coefficients formula is given as follows:
* here, n refers to the number of data points in the given data sets.

**Q.4] Define Cost Function in Linear Regression.**

* While dealing with Linear Regression we can have multiple lines for different values of slopes and intercepts.
* But the main question that arises is which of those lines actually represents the right relationship between the X and Y and in order to find that we can use the Mean Squared Error or MSE as the parameter.
* For linear regression, this MSE is nothing but the Cost Function.
* Mean Squared Error is the sum of the squared differences between the prediction and true value. And the output is a single number representing the cost.
* So the line with the minimum cost function or MSE represents the relationship between X and Y in the best possible manner. And once we have the slope and intercept of the line which gives the least error, we can use that line to predict Y.
* **Mean Square Error = (1/n) \* ∑(yi – xi)** As we know that any line can be represented by two parameters- slope and Intercept.

**Q.5] What is Simple Linear Regression, Explain with Linear Regression Intuition?**

|  |
| --- |
| https://www.gatevidyalay.com/wp-content/uploads/2020/02/Simple-Linear-Regression-Case-01.png |

* In simple linear regression, the dependent variable depends only on a single independent variable. For simple linear regression, the form of the model is
* **Y = β0 + β1X or Y=mx+c**

Were, Y is a dependent variable.

* X is an independent variable.
* β0 or c and β1 or m are the regression coefficients.

|  |
| --- |
| https://www.gatevidyalay.com/wp-content/uploads/2020/02/Simple-Linear-Regression-Case-02-1.png |

* β0 or c is the intercept or the bias that fixes the offset to a line.
* β1 or m is the slope or weight that specifies the factor by which X has an impact on Y.

There are following 3 cases possible

* **Case-01:** β1 < 0

|  |
| --- |
| https://www.gatevidyalay.com/wp-content/uploads/2020/02/Simple-Linear-Regression-Case-03-1.png |

• It indicates that variable X has negative impact on Y.

• If X increases, Y will decrease and vice-versa.

* **Case-02:** β1 = 0

• It indicates that variable X has no impact on variable Y.

• If X changes, there will be no change in Y.

* **Case-03:** β1 > 0

•It indicates that variable X has positive impact on Y.

•If X increases, Y will increase and vice-versa.

**Q.6] What is Multiple Linear Regression, Explain with Linear Regression Intuition?**

* In multiple linear regression, the dependent variable depends on more than one independent variables. For multiple linear regression, the form of the model is

**Y = β0 + β1X1 + β2X2 + β3X3 + …… + βnXn**

* Here, Y is a dependent variable.

•X1 , X2 , …., Xn are independent variables.

•β0 , β1 ,…, βn are the regression coefficients.

•βj (1<=j<=n) is the slope or weight that specifies the factor by which Xj has an impact on Y

|  |
| --- |
| Gradient Descent in Machine Learning - Javatpoint |

**Q.7] Define Gradient Descent. How does Gradient Descent work?**

* **Gradient Descent** : Gradient Descent is known as one of the most commonly used optimization algorithms to train machine learning models by means of minimizing errors between actual and expected results. Further, gradient descent is also used to train Neural Networks.
* In mathematical terminology, Optimization algorithm refers to the task of minimizing/maximizing an objective function f(x) parameterized by x.
* Similarly, in machine learning, optimization is the task of minimizing the cost function parameterized by the model's parameters.
* The main objective of gradient descent is to minimize the convex function using iteration of parameter updates.
* Once these machine learning models are optimized, these models can be used as powerful tools for Artificial Intelligence and various computer science applications.
* **Gradient Descent work:** The slope of a line from linear regression **Y=mX+c** Where, **'m'** represents the slope of the line, and **'c'** represents the intercepts on the y-axis. The main objective of gradient descent is to minimize the cost function or the error between expected and actual. To minimize the cost function, two data points are required:
* **Direction & • Learning Rate :**
* It is defined as the step size taken to reach the minimum or lowest point.
* This is typically a small value that is evaluated and updated based on the behaviour of the cost function.
* If the learning rate is high, it results in larger steps but also leads to risks of overshooting the minimum.
* At the same time, a low learning rate shows the small step sizes, which compromises overall efficiency but gives the advantage of more precision.

**Q. 8] Explain polynomial Regression with mathematical intuition.**

* A Polynomial Regression algorithm is also called Polynomial Linear Regression because it does not depend on the variables, instead, it depends on the coefficients, which are arranged in a linear fashion. Equation of the Polynomial Regression Model
* Simple Linear Regression equation: y = b0+b1x .........(a)
* Multiple Linear Regression equation: y= b0+b1x+ b2x2+ b3x3+....+ bnxn .........(b)
* Polynomial Regression equation: y= b0+b1x + b2x2+ b3x3+....+ bnxn ..........(c)
* When we compare the above three equations, we can clearly see that all three equations are Polynomial equations but differ by the degree of variables. The Simple and Multiple Linear equations are also Polynomial equations with a single degree, and the Polynomial regression equation is Linear equation with the nth degree.

**Q. 9] Explain Regularization Techniques with suitable example.**

* Regularization When it comes to training models, there are two major problems one can encounter: Overfitting and Underfitting.
* Overfitting happens when the model performs well on the training set but not so well on unseen (test) data. Underfitting happens when it neither performs well on the train set nor on the test set.
* Particularly, regularization is implemented to avoid overfitting of the data, especially when there is a large variance between train and test set performances.
* With regularization, the number of features used in training is kept constant, yet the magnitude of the coefficients (w) is reduced.
* While there are quite a number of predictors, RM and RAD have the largest coefficients. The implication of this will be that housing prices will be driven more significantly by these two features leading to overfitting, where generalizable patterns have not been learned.
* There are different ways of reducing model complexity and preventing overfitting in linear models. This includes ridge and lasso regression models.

**Q.10] Differentiate between Ridge and Lasso Regularization.**

* Ridge and Lasso regularization are both methods to prevent overfitting in linear regression. They work by adding a penalty term to the cost function that the model optimizes.
* **Ridge regularization**, also known as L2 regularization, adds a term to the cost function that is proportional to the square of the magnitude of the coefficients.
* The term is represented by lambda **\*** the sum of the square of the coefficients. This has the effect of shrinking the coefficients towards zero, but not setting any of them exactly to zero.
* This regularization term is added to the cost function to minimize the effect of large coefficients on the model, which in turn reduces the variance of the model.
* Ridge regularization can be useful for models with a large number of features, where some features may be correlated.
* In Ridge regularization will shrink all coefficients but none will be exactly zero, Lasso regularization will shrink some coefficients to zero, making it useful for feature selection.
* L2 regularization is generally considered a smooth regularization, while L1 regularization is considered as a non-smooth regularization. L2 regularization is sensitive to all features, while L1 regularization is only sensitive to the most important features.
* **Lasso regularization,** also known as L1 regularization, adds a term to the cost function that is proportional to the absolute value of the coefficients.
* The term is represented by lambda **\*** the sum of the absolute value of the coefficients.
* This has the effect of shrinking some coefficients all the way to zero, effectively eliminating them from the model.
* As a result, Lasso regularization can also be used for feature selection. The L1 regularization term in Lasso causes some weights to be exactly zero.
* This can be useful in situations where there are many correlated predictors, as it will automatically select the most important predictors and remove the redundant predictors.
* Lasso regularization is also useful in situations where we have a large number of features and we want to select a subset of the most important features for building our model.

**Q.11] Describe Logistic Regression with mathematical intuition.**

* **Logistic Regression:** Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.
* The logistic function is defined as:
* **p(x) = 1 / (1 + e^(-wx))**

Where, **x** is the input vector (consisting of one or more features),

**w**, is the weight vector (containing the model parameters),

**e**, is the natural exponential constant.

* The logistic regression model is trained by adjusting the weight vector w to maximize the likelihood of the observed data. The likelihood is a function of the parameters of the model, and it measures how well the model fits the data. To maximize the likelihood, we use a optimization algorithm such as gradient descent.
* Once the model is trained, we can use it to make predictions on new data by plugging in the input features and computing the probability that the outcome is 1. A threshold is used to decide whether the output of the model is 1 or 0.
* **For example,** if the probability is greater than 0.5, we predict 1, otherwise we predict 0.
* Logistic Regression is a simple yet powerful method for binary classification, and it is widely used in many real-world applications, such as image classification, natural language processing, and medical diagnosis.
* Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.
* Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.
* In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1). The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.
* Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.
* Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification.
* It can also be extended to multi-class classification problems by using one-vs-all or softmax regression.

**Q.12] Describe Cost Function in Logistic Regression.**

* The cost function is defined as the measurement of difference or error between actual values and expected values at the current position and present in the form of a single real number.
* It helps to increase and improve machine learning efficiency by providing feedback to this model so that it can minimize error and find the local or global minimum.
* Further, it continuously iterates along the direction of the negative gradient until the cost function approaches zero.
* At this steepest descent point, the model will stop learning further. Although cost function and loss function are considered synonymous, also there is a minor difference between them.
* The slight difference between the loss function and the cost function is about the error within the training of machine learning models, as loss function refers to the error of one training example, while a cost function calculates the average error across an entire training set.
* The cost function is defined as the negative log-likelihood of the data given the model:

**Cost(w) = -(1/m) \* Σ(y log(p(x)) + (1-y) log(1-p(x)))**

* where ,**p(x)** is the probability that the outcome is 1 given the input features x,

**y** is the true label (0 or 1),

**m** is the number of training examples, and the summation is taken over all the training examples.

* The cost function is a non-convex function and in general has multiple local minima. But if the data is linearly separable and the learning rate is selected appropriately, the optimization algorithm like gradient descent will converge to global minima. Also, in case of non-linearly separable data, the cost function becomes non-convex and optimization algorithm like gradient descent may get stuck in local minima instead of global minima.
* The goal of the training process is to find the set of weights that minimize the cost function. This is done by using optimization algorithm like gradient descent, which iteratively updates the weights in the direction of the negative gradient of the cost function with respect to the weights. In other words, the algorithm repeatedly takes small steps in the direction that reduces the cost the most until the cost function reaches a minimum.
* Overall, the cost function in logistic regression is a measure of how well the model fits the data and it is minimized during the training process to find the best set of weights for the model.