**Logistic Regression in Machine Learning**

Logistic regression is a supervised machine learning algorithm mainly used for classification tasks where the goal is to predict the probability that an instance of belonging to a given class or not. It is a kind of statistical algorithm, which analyse the relationship between a set of independent variables and the dependent binary variables. It is a powerful tool for decision-making. For example, email spam or not. The difference between linear regression and logistic regression is that linear regression output is the continuous value that can be anything while logistic regression predicts the probability that an instance belongs to a given class or not.



* It is used for predicting the categorical dependent variable using a given set of independent variables.
* 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.

| **Sr.No** | **Linear Regression** | **Logistic Regression** |
| --- | --- | --- |
| **1** | Linear regression is used to predict the continuous dependent variable using a given set of independent variables. | Logistic regression is used to predict the categorical dependent variable using a given set of independent variables. |
| **2** | Linear regression is used for solving Regression problem. | It is used for solving classification problems. |
| **3** | In this we predict the value of continuous variables | In this we predict values of categorical variables |
| **4** | In this we find best fit line. | In this we find S-Curve. |
| **5** | Least square estimation method is used for estimation of accuracy. | Maximum likelihood estimation method is used for estimation of Accuracy |

**Terminologies involved in Logistic Regression:**

* Independent variables: The input characteristics or predictor factors applied to the dependent variable’s predictions.
* Dependent variable: The target variable in a logistic regression model, which we are trying to predict.
* Logistic function: The formula used to represent how the independent and dependent variables relate to one another. The logistic function transforms the input variables into a probability value between 0 and 1, which represents the likelihood of the dependent variable being 1 or 0.
* Odds: It is the ratio of something occurring to something not occurring. it is different from probability as the probability is the ratio of something occurring to everything that could possibly occur.
* Log-odds: The log-odds, also known as the logit function, is the natural logarithm of the odds. In logistic regression, the log odds of the dependent variable are modelled as a linear combination of the independent variables and the intercept.
* Coefficient: The logistic regression model’s estimated parameters, show how the independent and dependent variables relate to one another.
* Intercept: A constant term in the logistic regression model, which represents the log odds when all independent variables are equal to zero.
* Maximum likelihood estimation: The method used to estimate the coefficients of the logistic regression model, which maximizes the likelihood of observing the data given the model.