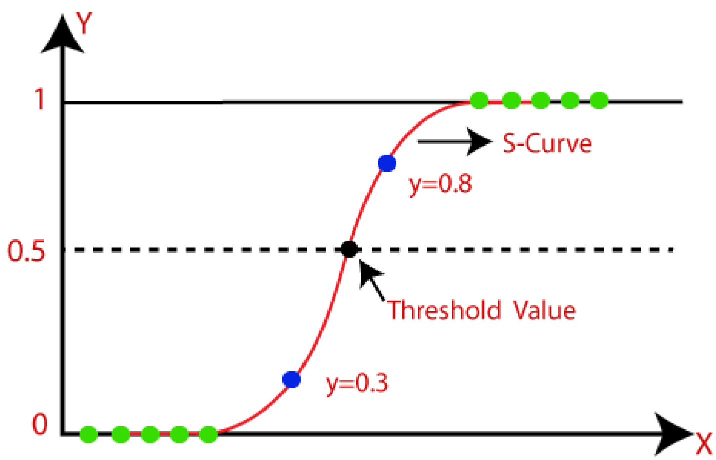
Logistic Regression

- **Logistic Regression** is a type of **supervised machine learning algorithm** used to model the probability of a certain class or event. It's often used for classification and predictive analytics.
- 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).



Logistic Function – Sigmoid Function

- The sigmoid function is a mathematical function used to map the predicted values to probabilities.
- It maps any real value into another value within a range of 0 and 1. The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit, so it forms a curve like the "S" form.

- The S-form curve is called the Sigmoid function or the logistic function.
- In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.

Types of Logistic Regression

On the basis of the categories, Logistic Regression can be classified into three types:

1. **Binomial:** In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.
2. **Multinomial:** In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable, such as "cat", "dogs", or "sheep"
3. **Ordinal:** In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as "low", "Medium", or "High".

Assumptions of Logistic Regression

We will explore the assumptions of logistic regression as understanding these assumptions is important to ensure that we are using appropriate application of the model. The assumption include:

1. Independent observations: Each observation is independent of the other. meaning there is no correlation between any input variables.
2. Binary dependent variables: It takes the assumption that the dependent variable must be binary or dichotomous, meaning it can take only two values. For more than two categories SoftMax functions are used.
3. Linearity relationship between independent variables and log odds: The relationship between the independent variables and the log odds of the dependent variable should be linear.
4. No outliers: There should be no outliers in the dataset.
5. Large sample size: The sample size is sufficiently large

Working of Logistic Regression

Here's a brief overview of how logistic regression works:

1. **Sigmoid Function**: Logistic regression uses the sigmoid function to map the output of the linear equation to a value between 0 and 1. The sigmoid function is defined as:

   $$\sigma(z)=1 / 1+e^{\wedge}-z$$

   Where $z=w0+w1x1+w2x2+...+wnxnz=w0+w1x1+w2x2+...+wnxn$, $wiwi$ are the weights, and $xixi$ are the input features.

2. **Hypothesis Function**: The hypothesis function for logistic regression is the sigmoid function applied to the linear equation:

   $$h\theta(x) = \sigma(\theta Tx) = 1 / 1+e^{\wedge}-\theta Tx$$

3. **Cost Function**: The cost function for logistic regression is the log loss (cross-entropy) function, which measures the difference between the predicted probability and the actual label:

   $$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))]$$

4. **Gradient Descent**: The goal is to minimize the cost function by adjusting the weights $\theta\theta$. This is typically done using gradient descent, where the weights are updated iteratively:

   $$\theta j := \theta j - \alpha\ (\ \partial j(\theta) / \partial(\theta) j$$

   Where $\alpha\alpha$ is the learning rate.

5. **Prediction**: To make a prediction, the probability output by the hypothesis function is compared to a threshold (usually 0.5). If the probability is above the threshold, the instance is classified as belonging to the positive class; otherwise, it is classified as belonging to the negative class.

Logistic Regression Equation

In logistic regression, a logit transformation is applied on the odds—that is, the probability of success divided by the probability of failure. This logistic function is represented by the following formulas:

Logit(pi)=1/(1+exp(−pi))

ln(pi/(1−pi))=Beta0+Beta1*X1+…+Bk*Kk

where `logit (pi)` is the dependent or response variable and `x` is the independent variable