Logistic Regression

Definition - Logistic Regression is **primarily used for binary classification problems** where the output variable can take only two distinct classes (e.g., yes/no, 0/1, spam/not spam). Despite its name, **logistic regression is used for classification, not regression**.

How It Works

1 - **Model Formulation**: Logistic Regression predicts the probability of a binary outcome using the logistic function (also known as the **sigmoid function**). The model is expressed as:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n)}}$$

- P(Y = 1|X) is the probability that the target variable Y is 1 given the features X
- B_0 is the intercept
- B_1, B_2, …, B_n are the coefficients of the features X_1, X_2, …, X_n
- e is the base of the natural logarithm

2 - **Training the Model**: Training involves finding the optimal values for the coefficients (B) that best fit the training data. This is typically done using **Maximum Likelihood Estimation (MLE)** which seeks to maximize the **likelihood function**.

To calculate likelihood function, first the algorithm uses the sigmoid function to calculate the predicted probabilities for each datapoint. Assume, for first data point the model predicts a probability of 0.731 for being class 1 (true), for second data point the model predicts a probability of 0.119 for being class 1 (true).

The likelihood function is the product of these individual likelihoods: 0.731*0.119 = 0.087

So the goal is to maximize this likelihood function by adjusting parameter values that are in the sigmoid function. This process can be done using optimization techniques like Gradient Descent.

3 - **Making Predictions**: To make predictions, the model calculates the probability that an instance belongs to the positive class using the logistic function. This probability is then compared to a threshold (usually 0.5). If the probability exceeds the threshold, the instance is classified as the positive class; otherwise, it is classified as the negative class. This threshold can be adjusted depending on the specific requirements of the classification problem, such as achieving a better balance between precision and recall.