Logistic Regression

Logistic Regression

Logistic regression is a method used to predict the **probability** that something belongs to one of two categories. It's like a special type of regression, but instead of predicting a number, it predicts a **probability** that a data point falls into one category or another.

Simple Explanation

Imagine you're trying to predict whether a student passes or fails a test based on their study hours. You have data that shows how many hours each student studied and whether they passed or failed.

- **Study hours**: 1, 2, 3, 4, 5
- **Pass/Fail**: Fail, Fail, Pass, Pass, Pass

Logistic regression helps us create a rule based on the study hours to predict whether the next student will pass or fail. The outcome is always one of two things:

- **Pass** (1)
- **Fail** (0)

How does it work?

1. **Input**: Logistic regression takes the **input data**, like study hours.
2. **Formula**: It uses a formula to calculate a number (called a **logit**) based on the input. This logit will be a number that can be between $-\infty$ and $+\infty$.
3. **Sigmoid function**: Then, the formula passes this number through a function called the **sigmoid function**. The sigmoid function squashes this number into a **probability** between 0 and 1.
    - If the result is **closer to 1**, it means the student is more likely to **pass**.

- If the result is **closer to 0**, it means the student is more likely to **fail**.

4. **Decision**: Based on the probability, logistic regression will decide the outcome:
   - If the probability is greater than 50% (0.5), it predicts **Pass**.
   - If the probability is less than 50% (0.5), it predicts **Fail**.

Example

- A student studied 3 hours.
- The model calculates the probability that the student will pass and gets 0.75.
- Since 0.75 is greater than 0.5, the model predicts the student will **pass**.

Why is it called "Logistic"?

It's called **logistic regression** because of the **logistic (sigmoid)** function used to convert the result into a probability.

So, in simple terms, logistic regression helps us predict if something will belong to one group or another, based on input features (like hours of study) and a probability.