Understanding Logistic Regression

A Supervised Learning Algorithm for Binary Classification

What is Logistic Regression?

Supervised learning algorithm used for binary classification

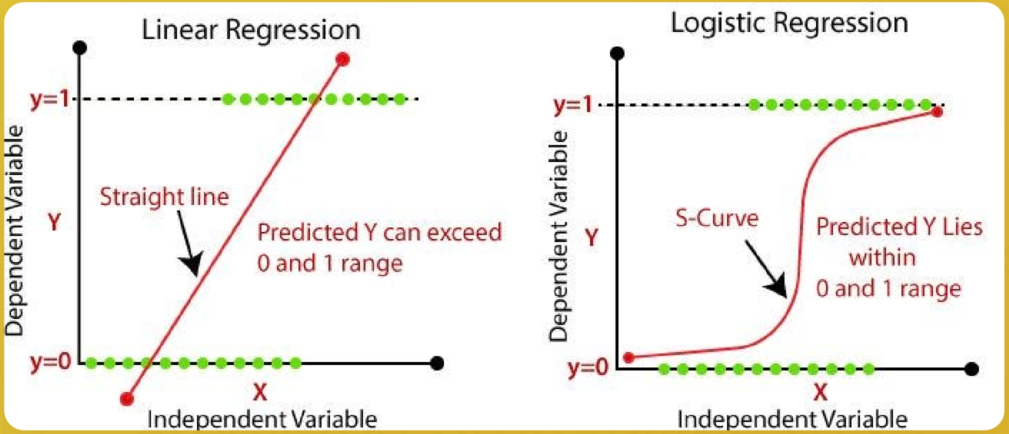
Output values (Y) are either:

- 0 (No)
- 1 (Yes)

Example:

Cat in image?
Yes → 1

Cat in image?
No → 0

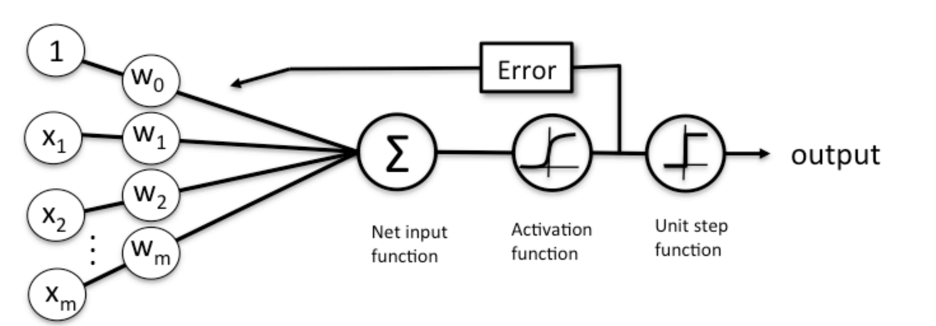


What Are We Trying to Do?

- We have an **input X** (e.g., a digital image)
- We want to **calculate the probability** that this image is of a cat
- This probability is denoted by **Ŷ (Y hat)**, our prediction of Y

$$P(Y = 1 \mid X)$$

We want to know the probability that **Y = 1** given **X**



Schematic of a logistic regression classifier.

Components of the Model

X: Feature Vector

The input data, such as pixel values of an image

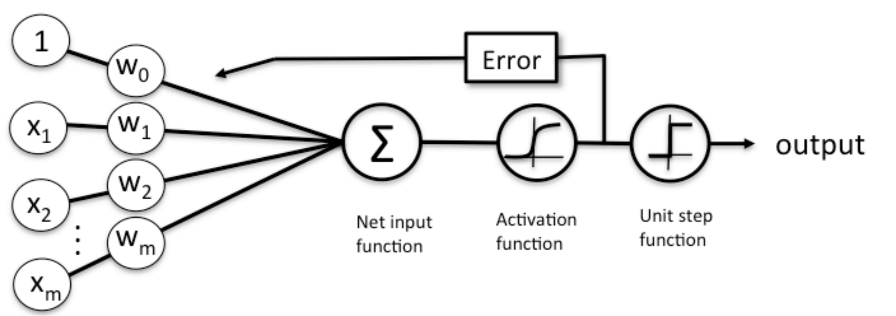
W: Weight Vector

Has the same dimensions as X

b: Bias

A single real number

Goal: Use **X, W, and b** to produce **Ŷ** (the predicted output)



Schematic of a logistic regression classifier.

How Do We Calculate the Prediction?

- First compute $Z = W^TX + b$
- Problem: Z might be **greater than 1** or **negative**
- This doesn't make sense for a **probability** (must be between 0 and 1)
- Solution: Pass Z through the **Sigmoid Function**

Linear Combination

W^T	×	X	+	b	=	Z
Weights		Features		Bias		Output

$$Z = W^TX + b$$

But Z can be any real number, not a valid probability

Σ The Sigmoid Function

Takes **any real number (Z)** and converts it to a value **between 0 and 1**

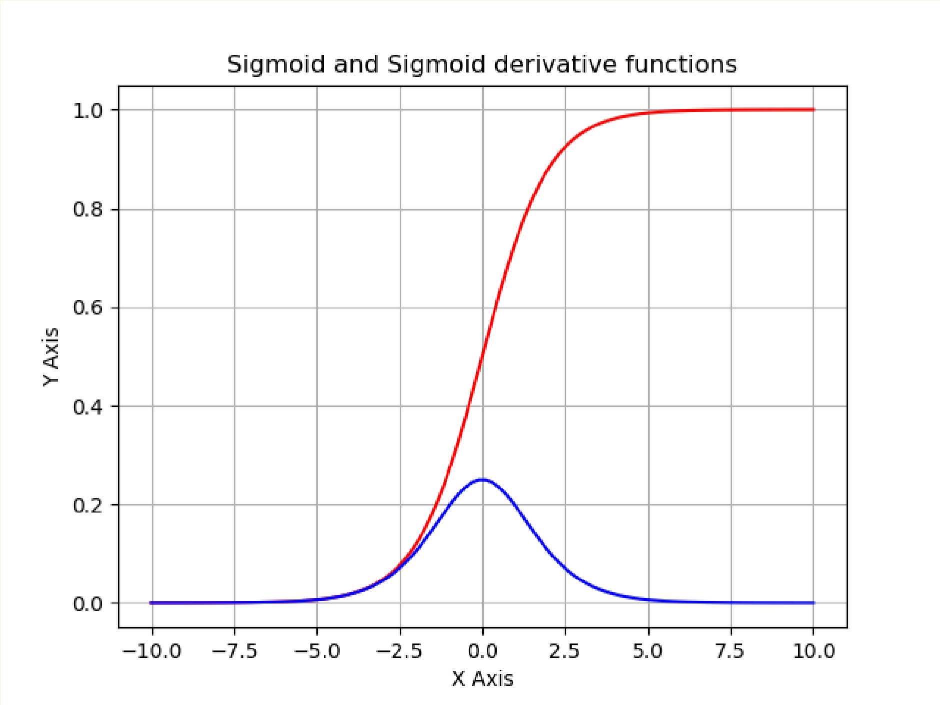
Formula

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

When **Z** is very large → output ≈ **1**

When **Z** is very small (negative) → output ≈ **0**

Transforms **linear values** into **meaningful probabilities**



The Complete Relationship in the Model

Complete Formula

$$\hat{Y} = \sigma(W^T X + b)$$

1. Compute $Z = W^T X + b$
2. Pass Z through the **sigmoid function**
3. Get $\hat{Y}$ between 0 and 1 → probability that $Y = 1$

Logistic Regression Model

$$W^T \times X + b = Z$$

$$Z \rightarrow \sigma(Z) \rightarrow \hat{Y}$$

$\hat{Y}$ represents the probability that $Y = 1$ given X

What Are We Trying to Learn?

- Learn the values of **W** (weights) and **b** (bias)
- Make predicted value $\hat{Y}$ as close as possible to actual value Y
- Minimize the difference between predictions and true labels
- Build a model that accurately predicts whether an image contains a **cat** or not

Learning Process

$\boxed{W, b} \rightarrow \boxed{\hat{Y} = \sigma(W^T X + b)}$

$\boxed{\hat{Y}} \approx \boxed{Y}$

Through optimization, we find the best W and b values

Summary

Key Points About Logistic Regression

Binary classification algorithm for 0/1 outputs

Uses **sigmoid function** to convert linear output to probability

Learns **weights (W)** and **bias (b)** to make accurate predictions

Final output: **Probability** that Y = 1 given X

Complete Logistic Regression Formula

$$\hat{Y} = \sigma(W^T X + b)$$

Where σ is the sigmoid function that transforms any real number into a probability between 0 and 1