-Logistic Regression-

Further simplified explanation **@S S Roy**

📘 Logistic Regression — Complete Tutorial with Full Worked Example

---

🎯 Objective

To understand how logistic regression models binary outcomes (e.g., success/failure) using a regression-style approach that ensures predicted probabilities lie between 0 and 1. This tutorial covers:

- Theory of logistic regression
- Model formulation
- Likelihood estimation
- A full example with every calculation

📌 1. What is Logistic Regression?

Logistic regression is a **classification algorithm** used when the **dependent variable is binary** (e.g., 0 or 1). It predicts the **probability** of an event occurring, not the value itself.

❌ Why Not Linear Regression?

Linear regression:

$$y = \beta_0 + \beta_1 x$$

...can produce values < 0 or > 1, which are invalid probabilities.

✅ Logistic regression fixes this by modeling the **log-odds** using a sigmoid function.

📐 2. The Sigmoid Function

The **sigmoid** or **logistic function** is:

$$f(z) = \frac{1}{1 + e^{-z}}, \quad \text{where } z = \beta_0 + \beta_1 x$$

This ensures that:

- $f(z) \in (0, 1)$ — valid probability
- Smooth S-shaped curve

🔢 3. Model Equation

The logistic regression model is:

$$\log\left(\frac{p}{1 - p}\right) = \beta_0 + \beta_1 x \quad \text{(logit model)}$$

Or,

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

🧮 4. Estimating Parameters

We estimate β_0 and β_1 using **Maximum Likelihood Estimation (MLE)**.

Likelihood Function:

Given $y_i \in \{0, 1\}$, and $p_i = P(y_i = 1 \mid x_i)$, the likelihood is:

$$L(\beta) = \prod_{i=1}^{n} p_i^{y_i} (1 - p_i)^{1 - y_i}$$

Log-Likelihood:

$$\ell(\beta_0, \beta_1) = \sum_{i=1} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

We maximize ℓ to find optimal β_0, β_1.

📊 5. Fully Worked Example with All Calculations

Let's now use an example dataset:

🎓 Sample Data

Student	Hours Studied (x)	Passed Exam (y)
A	1	0
B	2	0
C	3	1
D	4	1

We want to fit:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$$

🧠 Step 1: Choose Trial Values

Let's try $\beta_0 = -4, \beta_1 = 1.5$ to demonstrate all calculations.

🔢 Step 2: Compute Linear Combination $z = \beta_0 + \beta_1 x$

x	y	$z = -4 + 1.5x$
1	0	$-4 + 1.5(1) = -2.5$
2	0	$-4 + 1.5(2) = -1.0$
3	1	$-4 + 1.5(3) = 0.5$

x	y	$z = -4 + 1.5x$
4	1	$-4 + 1.5(4) = 2.0$

📈 Step 3: Apply Sigmoid to get Probabilities $p = \frac{1}{1+e^{-z}}$

z	e^{-z}	$p = \frac{1}{1+e^{-z}}$
-2.5	$e^{2.5} = 12.182$	$\frac{1}{1+12.182} = 0.0759$
-1.0	$e^{1.0} = 2.718$	$\frac{1}{1+2.718} = 0.2689$
0.5	$e^{-0.5} = 0.6065$	$\frac{1}{1+0.6065} = 0.6225$
2.0	$e^{-2.0} = 0.1353$	$\frac{1}{1+0.1353} = 0.8808$

🧮 Step 4: Compute Log-Likelihood

$$\ell = \sum [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

x	y	p_i	$\log(p_i)$	$\log(1 - p_i)$	Contribution
1	0	0.0759	-2.58	-0.0786	$\log(1 - p) = -0.0786$
2	0	0.2689	-1.313	-0.313	$\log(1 - p) = -0.313$
3	1	0.6225	-0.474	-0.974	$\log(p) = -0.474$
4	1	0.8808	-0.126	-2.136	$\log(p) = -0.126$

Total Log-Likelihood:

$$\ell = -0.0786 + (-0.313) + (-0.474) + (-0.126) = -0.9916$$

🔁 Step 5: Try Other Values

Try different β_0, β_1 pairs to **maximize** ℓ. Tools like Excel Solver, Python's `scipy.optimize`, or R's `glm()` automate this.

🔍 Step 6: Predict New Outcomes

Say a student studied 2.5 hours:

$$z = -4 + 1.5(2.5) = -0.25 \quad \Rightarrow \quad p = \frac{1}{1 + e^{0.25}} = 0.4378$$

Since $p < 0.5$, **predict FAIL (0).**

📊 Step 7: Interpret Coefficients

- $\beta_1 = 1.5$: every extra hour increases log-odds by 1.5
- $e^{1.5} = 4.48$: each hour studied multiplies **odds of passing by 4.48**

✅ Summary Table

Step	Description
1	Formulate model using log-odds
2	Use sigmoid to constrain output to [0,1]
3	Define log-likelihood function
4	Try parameters and compute log-likelihood
5	Choose best parameters (maximize likelihood)
6	Use model to predict and classify
7	Interpret coefficients using odds ratio

🧠 Final Thoughts

Logistic regression is powerful due to:

- **Interpretability** (log-odds and probabilities)
- **Statistical grounding** (likelihood-based)
- **Flexibility** (can be extended to multiple variables, regularized, or made nonlinear)

It remains a **first step in classification modeling** across statistics and machine learning.