Derivative of the Sigmoid Function

1. What is the Sigmoid Function?

The sigmoid function is:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

It takes any real number and “squashes” it into the range **(0, 1)**.
This is why it's so popular in logistic regression — the output can be interpreted as a probability.

2. Why Do We Need the Derivative?

In optimization (e.g., gradient descent), we need the **slope** of the function to update weights.
The derivative tells us **how much a small change in input changes the output**.
For sigmoid, the derivative is **beautifully simple** and that’s why it's so convenient for machine learning.

3. Step-by-Step Derivation

1. Start with:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

2. Differentiate using the chain rule:
   - Denominator: $(1 + e^{-x})$
   - Use the quotient rule or rewrite first:

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

   - Derivative:

$$\sigma'(x) = -1 \cdot (1 + e^{-x})^{-2} \cdot (-e^{-x})$$

3. Simplify:

$$\sigma'(x) = \frac{e^{-x}}{(1 + e^{-x})^2}$$

4. Now, **the magic trick**:
   - We know $\sigma(x) = \frac{1}{1+e^{-x}}$
   - And $1 - \sigma(x) = \frac{e^{-x}}{1+e^{-x}}$

   Multiply them:

$$\sigma(x) \cdot (1 - \sigma(x)) = \frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{1+e^{-x}} = \frac{e^{-x}}{(1+e^{-x})^2}$$

   That matches our derivative expression.

---

4. Result

The derivative of the sigmoid function is:

$$\sigma'(x) = \sigma(x) \cdot (1 - \sigma(x))$$

---

5. Why This is Useful

- No exponentials are needed after computing σ(x)\sigma(x) the derivative reuses the sigmoid value.
- For logistic regression and neural networks, this saves computation and avoids messy math.
- It also shows that:
  1. When σ(x) is near **0** or **1**, the derivative is small → gradient updates are tiny (this is why sigmoid can cause the vanishing gradient problem).
  2. When σ(x) = 0.5, the derivative is largest (0.25) → learning is fastest here.
