

# Logistic Regression

Linear Regression không phải lựa chọn hợp lý cho bài toán phân loại vì một số lý do như sau:

- Linear Regression có thể tạo ra giá trị **< 0 hoặc > 1**
- Khi thêm một điểm dữ liệu **xa về phía bên phải**, đường thẳng bị kéo lệch  
→ **Decision boundary (ranh giới quyết định)** cũng lệch theo  
→ Dự đoán sai cho phần lớn dữ liệu!

## Logistic Regression

**Đặc điểm:**

- Luôn cho giá trị đầu ra  $\hat{y} \in (0, 1)$
- Dùng để giải quyết **bài toán phân loại nhị phân**
- Mặc dù tên gọi là "regression", **Logistic Regression dùng cho classification**

**Dạng tổng quát:**

- Tính  $z = w \cdot x + b$
- Sau đó, đưa  $z$  qua **Hàm Sigmoid (Logistic Function)**:

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

→ Kết quả là một số **luôn nằm trong khoảng (0, 1)**

## Hàm Sigmoid – Logistic Function

Giá trị $z$	Đầu ra sigmoid $\sigma(z)$
$z \rightarrow -\infty$	$\rightarrow 0$
$z = 0$	0.5
$z \rightarrow +\infty$	$\rightarrow 1$

## Tổng kết

$$f(x) = \hat{y} = \sigma(z) = \sigma(w \cdot x + b) = \frac{1}{1 + e^{-(w \cdot x + b)}}$$

## Decision Boundary

- Là đường (hoặc đường cong) mà tại đó:

$$z = w \cdot x + b = 0$$

- Đây là “ranh giới” giữa dự đoán 0 và 1:
  - $z \geq 0 \rightarrow \hat{y} = 1$
  - $z < 0 \rightarrow \hat{y} = 0$