

Activation Functions

Trong mô hình Logistic Regression, thông thường hàm kích hoạt được sử dụng là hàm sigmoid. Tuy nhiên trong Neural Network còn nhiều hàm kích hoạt khác với các tính chất khác nhau.

ReLU - Rectified Linear Unit

Một activation function phổ biến là **ReLU**:

- Công thức:

$$g(z) = \max(0, z)$$

- Hành vi:
 - Nếu $z < 0 \rightarrow g(z) = 0$.
 - Nếu $z \geq 0 \rightarrow g(z) = z$ (tuyến tính).

- Ưu điểm:

Cho phép **giá trị dương lớn**, giúp mạng học được các đặc trưng phức tạp hơn.

So sánh 3 Activation Function chính

Tên	Công thức	Đặc điểm
Sigmoid	$g(z) = \frac{1}{1+e^{-z}}$	Giá trị trong (0, 1). Tốt cho đầu ra nhị phân, nhưng có thể gây vanishing gradient .
ReLU	$g(z) = \max(0, z)$	Dễ tính toán, phổ biến nhất hiện nay, không giới hạn phía trên.
Linear	$g(z) = z$	Đôi khi gọi là "không dùng activation". Dùng trong các bài toán hồi quy .

Chọn Activation Function Cho Output Layer

Việc chọn hàm kích hoạt phụ thuộc vào **kiểu giá trị đầu ra (label y)** mà bạn muốn dự đoán:

- **Sigmoid**: Tốt cho bài toán phân loại nhị phân vì trả về giá trị giữa 0 và 1 \rightarrow xác suất.

- **Linear**: Dùng cho hồi quy vì không giới hạn đầu ra.
- **ReLU**: Dùng khi giá trị dự đoán không âm.

Chọn Activation Function Cho Hidden Layers

- Mặc định dùng **ReLU** cho tất cả các lớp ẩn.
- sigmoid **ít dùng** trong lớp ẩn vì:
 - **Sigmoid**: Gradient gần như bằng 0 ở cả hai đầu → làm **học chậm**.
 - **ReLU**: Gradient rõ ràng ở phía dương → giúp học **nhANH hơn**.

Ưu điểm của ReLU:

- Dễ tính toán (chỉ cần `max(0, z)`).
- Hiệu quả hơn trong thực tế và dễ hội tụ hơn khi huấn luyện với gradient descent.

Ngoài ra vẫn còn nhiều hàm kích hoạt khác như: tanh, Leaky Relu, swish, ...

Nếu chỉ sử dụng hàm tuyến tính ở tất cả các tầng của mạng neural, thì:

- Mạng neural sẽ chỉ là một **hàm tuyến tính** của đầu vào.
- Dù bạn có **nhiều tầng**, mạng vẫn không học được gì **phức tạp hơn hồi quy tuyến tính (linear regression)**.

⇒ Chèn hàm phi tuyến vào giữa các tầng mới tạo ra một mạng có khả năng học các hàm phức tạp → phù hợp với các bài toán thực tế