

# Một số loại Activation Function

## Sigmoid (bổ sung)

Lý do chính tại sao chúng ta sử dụng hàm sigmoid là vì nó có phạm vi giá trị liên tục từ 0 đến 1. Do đó, nó đặc biệt được sử dụng cho các mô hình mà phải dự đoán xác suất. Vì vậy xác suất của bất cứ điều gì chỉ tồn tại giữa phạm vi 0 và 1.

Đạo hàm của hàm Sigmoid :

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

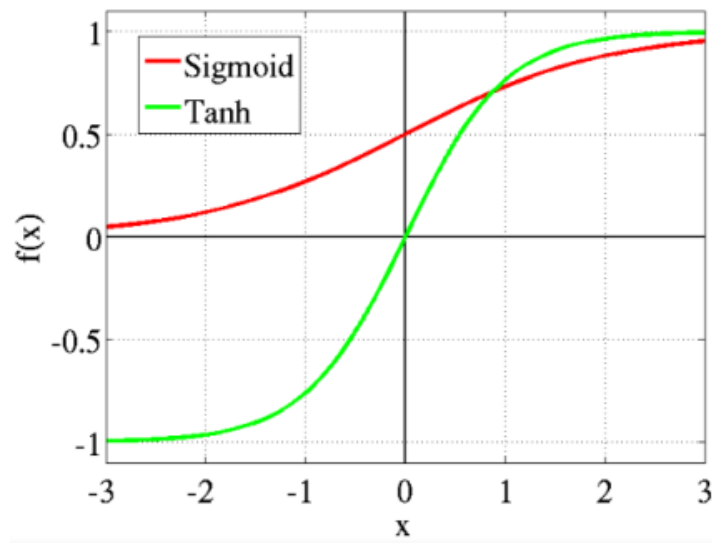
$$\frac{df(x)}{dx} = -\frac{1}{(1 + e^{-x})^2} (-e^{-x}) = \frac{e^{-x}}{(1 + e^{-x})^2} = \frac{1}{1 + e^{-x}} \left(1 - \frac{1}{1 + e^{-x}}\right) = f(x)(1 - f(x))$$

Hàm sigmoid có thể làm cho mạng thần kinh gặp khó khăn trong thời gian training.

## Tanh (Hyperbolic tangent function)

Hàm Tanh khá giống với hàm Sigmoid, có đồ thị hình chữ S.

Phạm vi:  $(-1, 1)$



Hình : So sánh hàm Tanh và Sigmoid

Đạo hàm của hàm Tanh :

$$f(x) = \tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$\frac{d \tanh}{dx} = \frac{\cosh(x) \cdot \sinh(x)' - \sinh(x) \cdot \cosh(x)'}{\cosh(x)^2}$$

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

$$\cosh(x)' = \frac{1}{2}(e^x - e^{-x}) = \sinh(x)$$

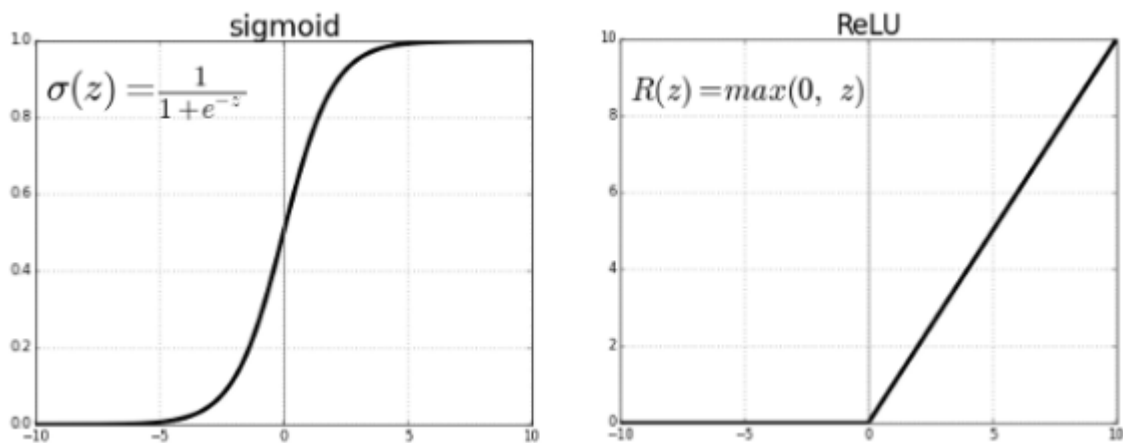
$$\frac{d \tanh}{dx} = \frac{\cosh(x)^2 - \sinh(x)^2}{\cosh(x)^2} = 1 - \left(\frac{\sinh(x)}{\cosh(x)}\right)^2 = 1 - \tanh(x)^2$$

Hàm tanh chủ yếu được sử dụng để phân loại giữa hai lớp.

## ReLU (Rectified Linear Unit)

ReLU là activation function được sử dụng nhiều nhất trên thế giới hiện nay, nó được sử dụng trong Convolutional Neural Networks (CNN) or Deep Learning.

$$f(x) = \max(0, x)$$



Hình : So sánh hàm ReLU và Sigmoid

Phạm vi:  $[0, \infty)$

Vấn đề: với bất kỳ input có giá trị âm đưa vào cho hàm ReLU sẽ biến giá trị thành 0 ngay lập tức trong biểu đồ, do đó ảnh hưởng đến biểu đồ kết quả bằng cách không ánh xạ các giá trị âm một cách thích hợp.

## Leaky ReLUs và Parametric ReLUs

Leaky ReLUs và Parametric ReLUs là biến thể từ ReLU để giải quyết vấn đề của ReLU.

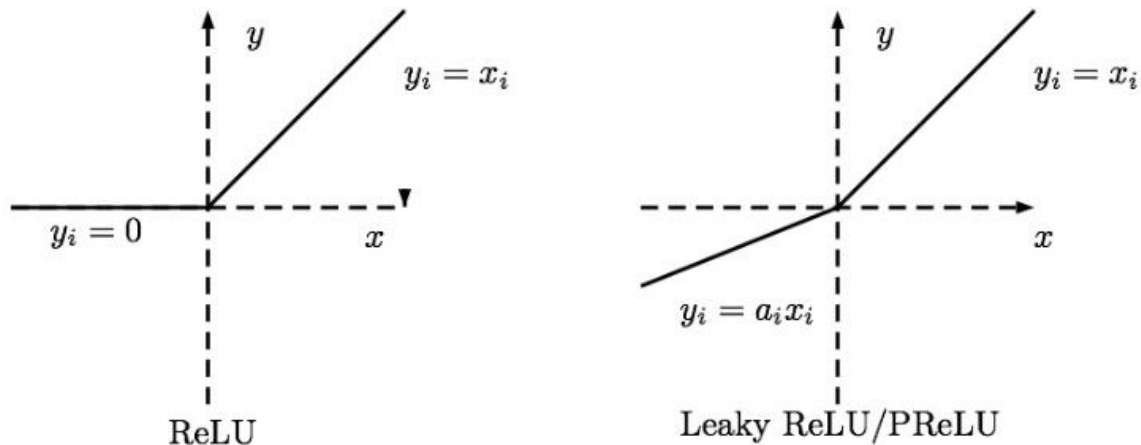
Leaky ReLUs:

$$f(x) = \begin{cases} x, & x > 0 \\ 0.01x, & otherwise \end{cases}$$

Parametric ReLUs:

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha x, & \text{otherwise} \end{cases} ; \text{ if } \alpha \leq 1, f(x) = \max(x, \alpha x)$$

Phạm vi:  $[-\infty, \infty)$



Hình: So sánh giữa ReLU và LeakyReLU/PreLU

## Softmax function

Hàm sigmoid chỉ có thể xử lý hai lớp. Các hàm ReLUs sẽ không làm mất hiệu lực trong quá trình đào tạo. Tuy nhiên, phân loại lại là vấn đề.

Hàm softmax có output nằm trong khoảng từ 0 đến 1, giống như hàm sigmoid. Nhưng nó chia mỗi output sao cho tổng số output bằng 1.

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \text{ for } j = 1, 2, \dots, K$$

Với  $\mathbf{z}$ :  $K$ -dimensional vector

Links nguồn:

[https://en.wikipedia.org/wiki/Activation\\_function](https://en.wikipedia.org/wiki/Activation_function)

<https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>

<https://www.analyticsindiamag.com/most-common-activation-functions-in-neural-networks-and-rationale-behind-it/>

<https://github.com/Kulbear/deep-learning-nano-foundation/wiki/ReLU-and-Softmax-Activation-Functions>