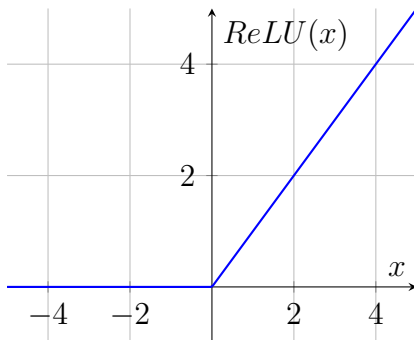


Activation Functions in Neural Networks

Introduction

Activation functions introduce nonlinearity to neural networks, allowing them to learn complex mappings. Below are the most commonly used activation functions with their formulas, shapes, and output ranges.

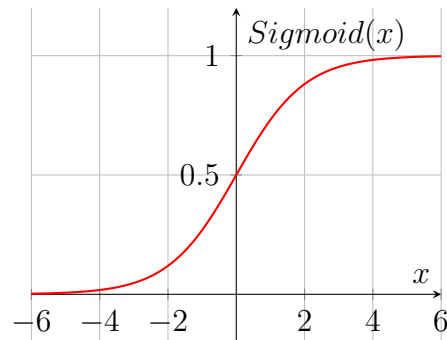
ReLU (Rectified Linear Unit)



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

Range: $[0, \infty)$ **Note:** ReLU outputs zero for negative inputs and is linear for positive values.

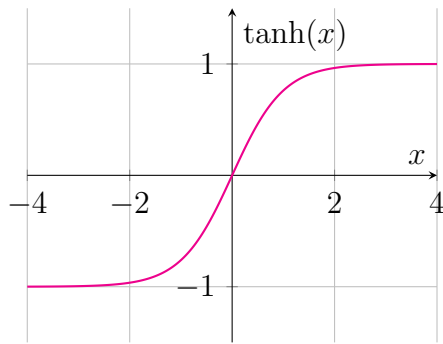
Sigmoid Function



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

Range: $(0, 1)$ **Note:** Compresses inputs into the range $(0, 1)$. Commonly used in binary classification.

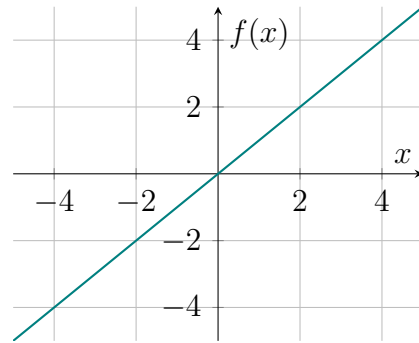
Hyperbolic Tangent (Tanh)



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

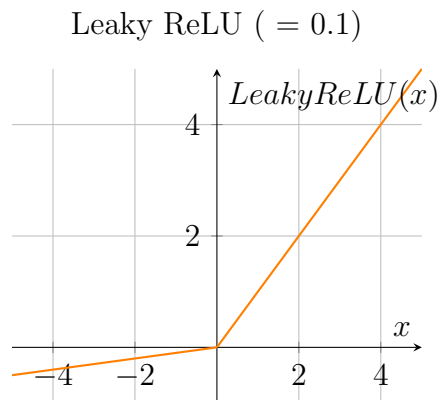
Range: $(-1, 1)$ **Note:** Centered at zero, often preferred over sigmoid for hidden layers.

Linear (Identity) Function



$$f(x) = x$$

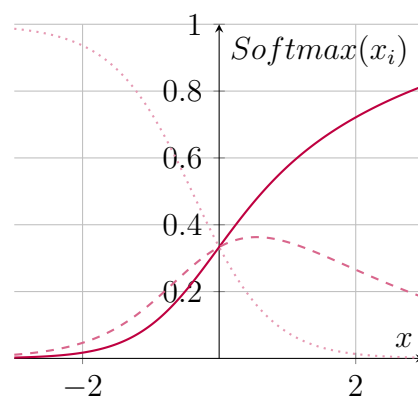
Range: $(-\infty, \infty)$ **Note:** No nonlinearity; used in regression output layers.



$$f(x) = \begin{cases} x, & x > 0 \\ 0.1x, & x \leq 0 \end{cases}$$

Range: $(-\infty, \infty)$ **Note:** Allows a small gradient when $x < 0$, preventing dead neurons.

Softmax Function (Example for 3 Classes)



$$f_i(x) = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

Range: $(0, 1)$ with $\sum_i f_i(x) = 1$ **Note:** Converts raw scores to probabilities, used in multi-class output layers.