

AI Coffe Club (09/01/2020)

Activation Functions

Last modification: January 8, 2020

Activation functions are used to modify or control the output of layers of a neural network. More precisely, they are applied to each neuron and they determine whether they should be activated or not and, in more complex cases, the magnitude of such activation.

Although activation functions were traditionally used as a mathematical gate, a simple step function, to turn the neuron on and off depending on a threshold to produce binary classifications, nowadays they are used for more complex purposes.

Modern neural networks use non-linear activation functions to create complex mappings between the inputs and the output and generate more complex decision boundaries to tackle difficult classification problems.

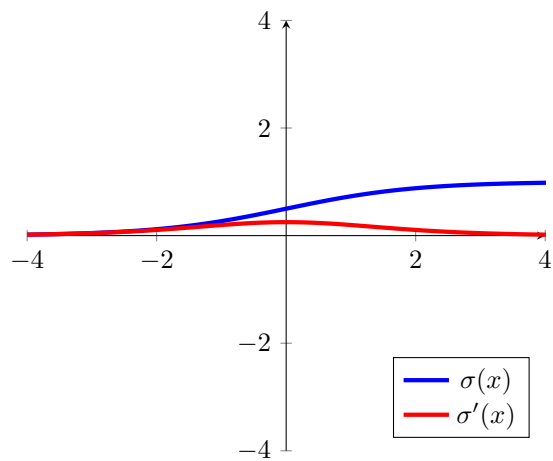
Many activation functions have been proposed to date, most of them trying to strike a balance between making the training easier, increasing representation capacity, avoiding common training stalemates, and computational efficiency. Table 1 shows a summarized overview.

Remark: we need activation functions to be differentiable so as to perform backpropagation optimization.

Name	Function	Derivative	Range
Sigmoid	$\sigma(x) = \frac{1}{1+e^{-x}}$	$\sigma'(x) = \sigma(x)(1 - \sigma(x))$	$(0, 1)$
Hyperbolic Tangent	$\sigma(x) = \frac{e^{2x}-1}{e^{2x}+1}$	$\sigma'(x) = 1 - \sigma(x)^2$	$(-1, 1)$
ReLU	$\sigma(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases}$	$\sigma(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases}$	$[0, \infty)$
Leaky ReLU	$\sigma(x) = \begin{cases} 0.01x & x \leq 0 \\ x & x > 0 \end{cases}$	$\sigma(x) = \begin{cases} 0.01 & x \leq 0 \\ x & x > 0 \end{cases}$	$(-\infty, \infty)$
Parametric ReLU	$\sigma(x) = \begin{cases} \alpha x & x \leq 0 \\ x & x > 0 \end{cases}$	$\sigma(x) = \begin{cases} \alpha & x \leq 0 \\ x & x > 0 \end{cases}$	$(-\infty, \infty)$
ELU	$\sigma(x) = \begin{cases} \alpha(e^x - 1) & x \leq 0 \\ x & x > 0 \end{cases}$	$\sigma(x) = \begin{cases} \alpha + \sigma(x) & x \leq 0 \\ x & x > 0 \end{cases}$	$(-\alpha, \infty)$
Softplus	$\sigma(x) = \ln(1 + e^x)$	$\sigma'(x) = \frac{1}{1+e^{-x}}$	$(0, \infty)$
Swish	$\sigma(x) = \frac{x}{1+e^{-\beta x}}$	$\sigma'(x) = \beta \sigma(x) + \frac{1}{1+e^{-\beta x}}(1 - \beta \sigma(x))$	

Table 1: Common activation functions.

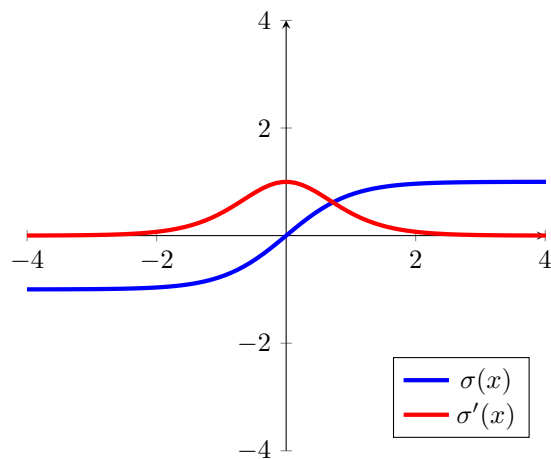
1 Sigmoid



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

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \quad (2)$$

2 Hyperbolic Tangent



$$\sigma(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1} \quad (3)$$

$$\sigma'(x) = 1 - \sigma(x)^2 \quad (4)$$

3 Rectified Linear Unit

$$\sigma(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases} \quad (5)$$

$$\sigma'(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases} \quad (6)$$

4 Leaky Rectified Linear Unit

$$\sigma(x) = \begin{cases} 0.01x & x \leq 0 \\ x & x > 0 \end{cases} \quad (7)$$

$$\sigma'(x) = \begin{cases} 0.01 & x \leq 0 \\ 1 & x > 0 \end{cases} \quad (8)$$

5 Parametric Rectified Linear Unit

$$\sigma(x) = \begin{cases} \alpha x & x \leq 0 \\ x & x > 0 \end{cases} \quad (9)$$

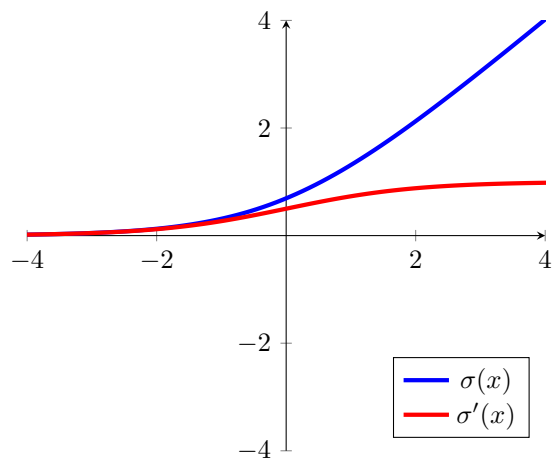
$$\sigma'(x) = \begin{cases} \alpha & x \leq 0 \\ 1 & x > 0 \end{cases} \quad (10)$$

6 Exponential Linear Unit

$$\sigma(x) = \begin{cases} \alpha(e^x - 1) & x \leq 0 \\ x & x > 0 \end{cases} \quad (11)$$

$$\sigma'(x) = \begin{cases} \alpha + \sigma(x) & x \leq 0 \\ 1 & x > 0 \end{cases} \quad (12)$$

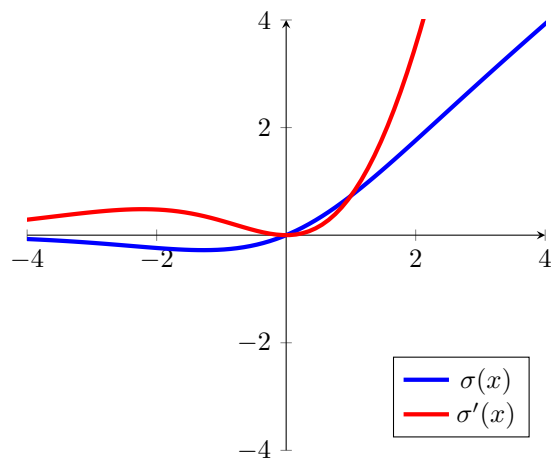
7 Softplus



$$\sigma(x) = \ln(1 + e^x) \quad (13)$$

$$\sigma'(x) = \frac{1}{1 + e^{-x}} \quad (14)$$

8 Swish



$$\sigma(x) = \frac{x}{1 + e^{-x}} \quad (15)$$

$$\sigma'(x) = x\sigma(x) \quad (16)$$