Activation functions 108-W4-P4-171RMC216

Audars Karlis Klints

RTU Telecomunications faculty

March, 2018

The sigmoid function

The sigmoid function is defined as follows:

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

The sigmoid function derivative $\sigma'(x)$ is defined as follows:

$$\sigma'(x) = \frac{d}{dx}dx = \sigma(x)(1-\sigma(x))$$

Here's a detailed derivation:

$$\frac{d}{dx}dx =$$

$$= \frac{d}{dx}\left[\frac{1}{1+e^{-x}}\right] = \frac{d}{dx}(1+e^{-x})^{-1} =$$

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

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

$$= \frac{1}{1+e^{-x}} \times \left(\frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}\right) = \frac{1}{1+e^{-x}} \times \left(1 - \frac{1}{1+e^{-x}}\right) =$$

$$= \sigma(x)(1-\sigma(x))$$

Function graph

As represented in following graph, it has small output changes in (0 and 1) when the varies in (-∞ and $+\infty$). Mathematically, the function is continuous. A typical sigmoid function is represented in the following graph:

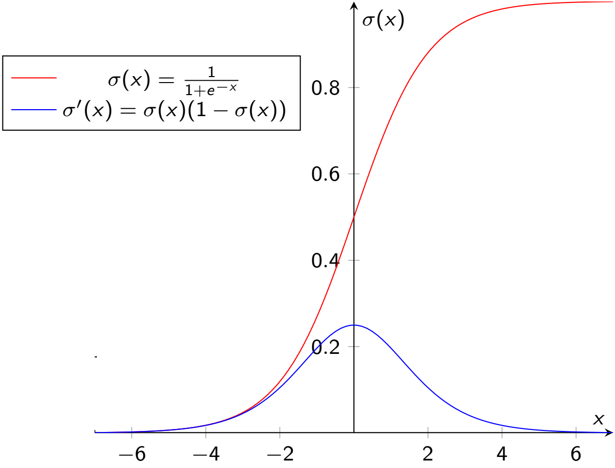

Figure: Sigmoid function

Sigmoid graph usage

Audars Karlis Klints

A neuron can use the sigmoid for computing the nonlinear function $\sigma(z = wx + b)$. Note that, if $z = wx + b$ is very large and positive, then $e^{-z} \to 0$, so $\sigma(z) \to 1$, while $z = wx + b$ if is very large and negative $e^{-z} \to \infty$ so $\sigma(z) \to 0$. In other words, a neuron with sigmoid activation has a behavior similar to the perceptron, but the changes are gradual and output values, such as 0.5539 or 0.123191, are perfectly legitimate. In this sense, a sigmoid neuron can answer *maybe*.

Activation function - ReLU

The sigmoid is not the only kind of smooth activation function used for neural network. Recently, a very simple function called "Rectified linear unit (ReLU)" became very popular because it generates very good experimental results. A ReLU is simply defined as $f(x) = max(0, x)$, and the nonlinear function is represented in the following graph. As you can see in the following graph, the function is zero for negative values, and it grows linearly for positive values:

Activation function (ReLU) graph

Audars Karlis Klints

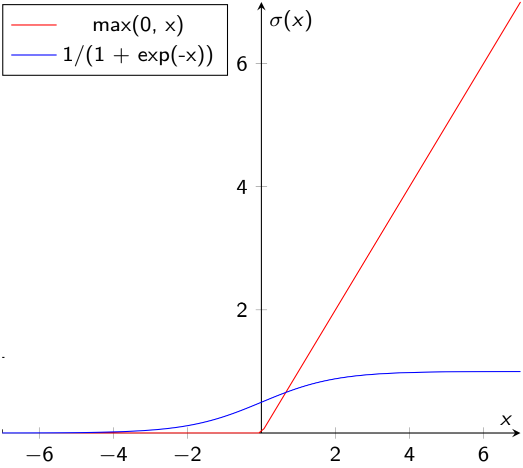

Figure: Activation function (ReLU) graph

Activation functions

Audars Karlis Klints

Sigmoid and ReLU are generally calles *activationfunctions* in neural network jargon. In the *TestingdifferentoptimizersinKeras* section, we will see that those gradual changes, typical of sigmoid and ReLU functions, are the basic building blocks to developing a learning algorithm which adapts little by little, by progressively reducing the mistakes made by our nets. An example of using the activation function σ with the $(x_1, x_2, ..., x_m)$ input vector, $(\omega_1, \omega_2, ..., \omega_m)$ weight vector, b bias, and Σ summation is given in the following diagram:

Activation function diagram

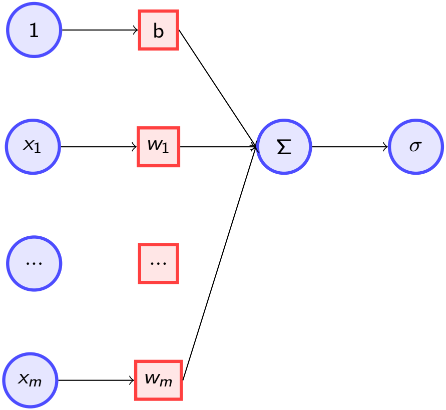

Figure: Activation function diagram