

Sigmoid Function

Definition

In the context of Machine Learning, there is a function that appears very often. We are talking about the **sigmoid function** σ . This is the real function defined as follows:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}, \quad (1)$$

where $x \in \mathbb{R}$. In these notes, we present the most important results related to σ .

Graph

We begin by taking a look at the graph of the sigmoid function:

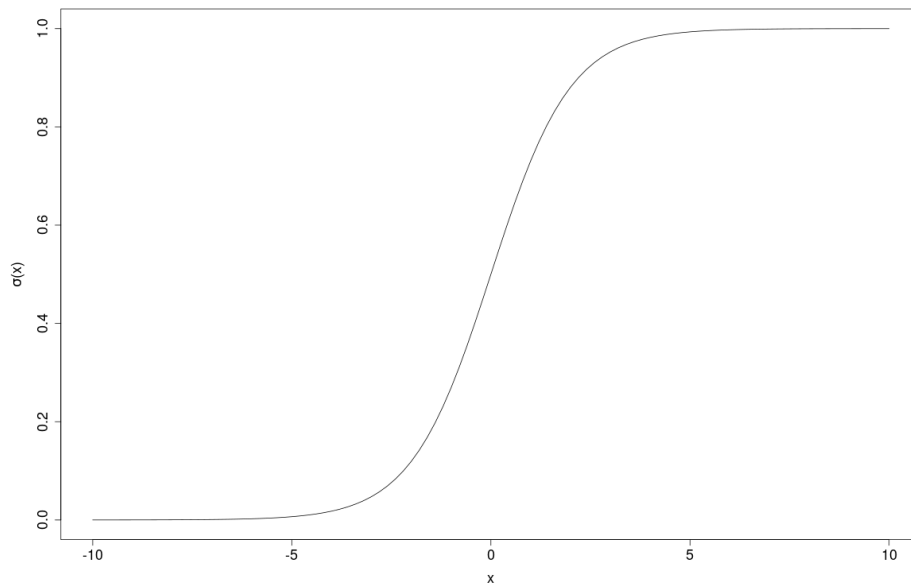


Figure 1: Graph of the sigmoid function.

So essentially σ is a function whose curve has a characteristic S-shape.

Boundedness

Fig. 1 illustrates an important fact: the sigmoid function is bounded. Specifically, the values returned by σ are between 0 and 1. Next, we prove this fact.

First, consider the lower bound of the sigmoid function. It is well-known that the exponential function is strictly positive: $\exp(x) > 0, \forall x \in \mathbb{R}$. This allows us to conclude that the denominator on the RHS of Eq. (1) is also strictly positive, i.e., $1 + \exp(-x) > 0, \forall x \in \mathbb{R}$. Hence, $\sigma(x)$ is the inverse of a strictly positive real number. Therefore, we have

$$\sigma(x) > 0, \quad \forall x \in \mathbb{R}. \quad (2)$$

We continue by considering the upper bound of $\sigma(x)$. Once again, we use the fact that the exponential function is strictly positive. This allows us to establish the following inequality: $1 + \exp(-x) > 1, \forall x \in \mathbb{R}$. Then the expression on the RHS of Eq. (1) is the inverse of a real number greater than 1. Hence, the sigmoid function satisfies

$$\sigma(x) < 1, \quad \forall x \in \mathbb{R}. \quad (3)$$

To finish, we combine (2) and (3) into a single inequality, and write

$$0 < \sigma(x) < 1, \quad \forall x \in \mathbb{R}. \quad (4)$$

Value at the Origin and Special Limits

By using Eq. (1), it is simple to evaluate the sigmoid function at $x = 0$. Since $\exp(0) = 1$, we have

$$\sigma(0) = \frac{1}{1 + \exp(0)} = \frac{1}{2}. \quad (5)$$

We can also compute the limit of $\sigma(x)$ as $x \rightarrow +\infty$. We do so by using the following well-known limit:

$$\lim_{x \rightarrow +\infty} \exp(-x) = 0.$$

This result allows us to prove that, as x increases arbitrarily, the sigmoid function approaches 1:

$$\lim_{x \rightarrow +\infty} \sigma(x) = \frac{1}{1 + 0} = 1. \quad (6)$$

Next, we calculate the limit of $\sigma(x)$ as $x \rightarrow -\infty$. In this case, we use the following limit of the exponential function:

$$\lim_{x \rightarrow -\infty} \exp(-x) = \infty.$$

Hence, as x decreases arbitrarily, the sigmoid function approaches 0:

$$\lim_{x \rightarrow -\infty} \sigma(x) = 0. \quad (7)$$

Relation to the Hyperbolic Tangent

We proceed by deriving an alternative expression for the sigmoid function. More precisely, we show how $\sigma(x)$ can be rewritten in terms of the hyperbolic tangent. Recall that this function is given by

$$\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}.$$

Hence:

$$\tanh\left(\frac{x}{2}\right) = \frac{\exp\left(\frac{x}{2}\right) - \exp\left(-\frac{x}{2}\right)}{\exp\left(\frac{x}{2}\right) + \exp\left(-\frac{x}{2}\right)}.$$

Next, we add 1 to both sides of the above equation:

$$\begin{aligned} 1 + \tanh\left(\frac{x}{2}\right) &= \frac{\exp\left(\frac{x}{2}\right) + \exp\left(-\frac{x}{2}\right)}{\exp\left(\frac{x}{2}\right) + \exp\left(-\frac{x}{2}\right)} + \frac{\exp\left(\frac{x}{2}\right) - \exp\left(-\frac{x}{2}\right)}{\exp\left(\frac{x}{2}\right) + \exp\left(-\frac{x}{2}\right)} \\ &= \frac{2 \exp\left(\frac{x}{2}\right)}{\exp\left(\frac{x}{2}\right) + \exp\left(-\frac{x}{2}\right)} \\ &= \frac{2}{1 + \exp(-x)}. \end{aligned}$$

From the last equality, we obtain

$$\sigma(x) = \frac{1 + \tanh\left(\frac{x}{2}\right)}{2}. \quad (8)$$

We can use Eq. (8) to establish an interesting property of the sigmoid function. First, consider $\sigma(-x)$. Since the hyperbolic tangent is an odd function, it is simple to prove that

$$\sigma(-x) = \frac{1 - \tanh\left(\frac{x}{2}\right)}{2}. \quad (9)$$

Next, we compute the difference $1 - \sigma(x)$. We do so with the aid of Eq. (8):

$$1 - \sigma(x) = 1 - \frac{1 + \tanh\left(\frac{x}{2}\right)}{2} = \frac{1 - \tanh\left(\frac{x}{2}\right)}{2}. \quad (10)$$

Eqs. (9) and (10) allow us to write the following identity:

$$\sigma(-x) = 1 - \sigma(x). \quad (11)$$

We will utilize the above equation to obtain a simple expression for the derivative of the sigmoid function.

Derivatives

First Derivative

We begin by stating (without proof) the following: by using simple facts about analytic functions, it is possible to show that $\sigma(x)$ is analytic. This means that the sigmoid function is continuous, and so are all of its derivatives. Then we can proceed by computing these derivatives.

Let us start by evaluating the first derivative of $\sigma(x)$. To do so, we can use the definition of the sigmoid function, Eq. (1). By differentiating both sides of this equation with respect to x , we obtain

$$\begin{aligned}\sigma'(x) &= \frac{d}{dx} \left\{ [1 + \exp(-x)]^{-1} \right\} \\ &= -[1 + \exp(-x)]^{-2} [-\exp(-x)] \\ &= \frac{\exp(-x)}{[1 + \exp(-x)]^2}.\end{aligned}\tag{12}$$

The last result is correct. However, the derivative of the sigmoid function can be written in a much simpler form in terms of $\sigma(x)$. Next, we derive this simpler expression. To do so, we compute the product $\sigma(x)[1 - \sigma(x)]$. First, notice that Eq. (1) allows us to write the following:

$$1 - \sigma(x) = \frac{\exp(-x)}{1 + \exp(-x)}.$$

Hence:

$$\sigma(x)[1 - \sigma(x)] = \frac{1}{1 + \exp(-x)} \frac{\exp(-x)}{1 + \exp(-x)} = \frac{\exp(-x)}{[1 + \exp(-x)]^2}.\tag{13}$$

The last expression is exactly what we have on the RHS of Eq. (12). Therefore, we just proved that the derivative of the sigmoid function can be written as

$$\sigma'(x) = \sigma(x)[1 - \sigma(x)].\tag{14}$$

By using Eq. (11), we can also express the derivative of $\sigma(x)$ as

$$\sigma'(x) = \sigma(x)\sigma(-x).\tag{15}$$

This formula makes it very clear that the derivative of the sigmoid function is even:

$$\sigma'(-x) = \sigma'(x).$$

We proceed by discussing two important facts about Eq. (14). First, we remark that $\sigma(x)$ is also known as the **logistic function**. The function $\sigma(x)$ is given this name because the differential equation

$$\frac{dy(x)}{dx} = y(x)[1 - y(x)]$$

is called the logistic equation. $\sigma(x)$ is just one of its particular solutions. Specifically, it is the particular solution that satisfies $y(0) = \frac{1}{2}$.

Moreover, Eq. (14) teaches us something important about the behavior of the sigmoid function: $\sigma(x)$ is monotonically increasing. It is simple to prove this fact with the aid of (4). According to this inequality, both factors on the RHS of Eq. (14) are strictly positive. Hence:

$$\sigma'(x) > 0, \quad \forall x \in \mathbb{R}.\tag{16}$$

It is also possible to reach this conclusion by analyzing Eq. (12). Another confirmation that the inequality (16) is correct comes from the graph of the derivative of $\sigma(x)$ (see Fig. 2).

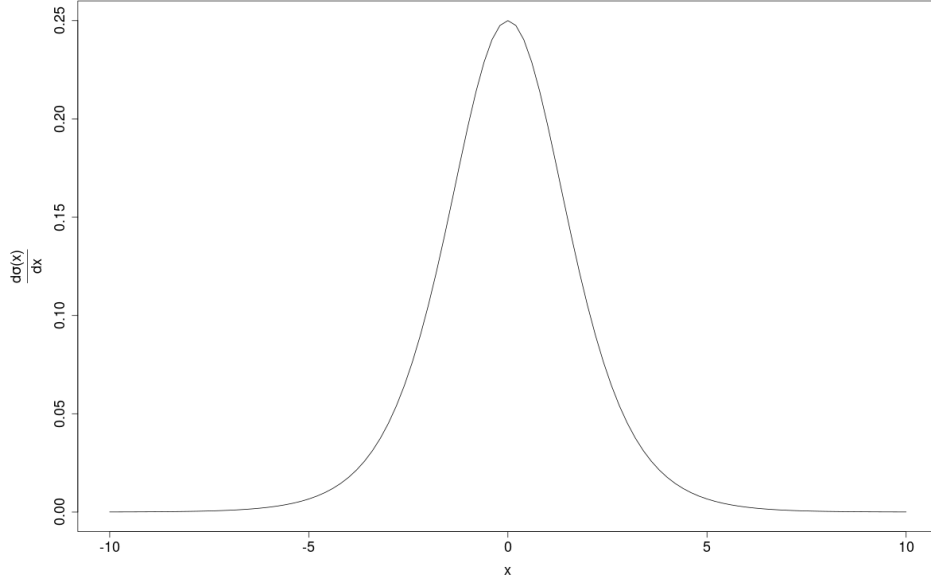


Figure 2: Graph of the first derivative of the sigmoid function.

To finish this discussion about the first derivative of the sigmoid function, we derive expressions for $\sigma'(x)$ in terms of hyperbolic functions. We proceed by using Eq. (15). Since we already have formulas for both factors on the RHS of this equation, we can write

$$\begin{aligned}
 \sigma'(x) &= \sigma(x) \sigma(-x) \\
 &= \frac{1 + \tanh\left(\frac{x}{2}\right)}{2} \frac{1 - \tanh\left(\frac{x}{2}\right)}{2} \\
 &= \frac{1 - \tanh^2\left(\frac{x}{2}\right)}{4} \\
 &= \frac{\operatorname{sech}^2\left(\frac{x}{2}\right)}{4}.
 \end{aligned} \tag{17}$$

Second Derivative

To better understand the sigmoid function, we will also compute the second derivative of $\sigma(x)$. To do so, we can differentiate both sides of Eq. (14) with respect to x :

$$\begin{aligned}
 \sigma''(x) &= \frac{d}{dx} \{ \sigma(x) [1 - \sigma(x)] \} \\
 &= \sigma'(x) - 2\sigma(x) \sigma'(x) \\
 &= \sigma(x) [1 - \sigma(x)] [1 - 2\sigma(x)] \\
 &= 2\sigma^3(x) - 3\sigma^2(x) + \sigma(x).
 \end{aligned} \tag{18}$$

We can use the above results to show that the sigmoid function has an inflection point at $x = 0$. To establish this fact, we consider the third-degree polynomial p defined by

$$p(x) = x(1-x)(1-2x) = 2x^3 - 3x^2 + x. \tag{19}$$

Clearly, one of the roots of this polynomial is $x = \frac{1}{2}$. By comparing Eqs. (18) and (19), it is possible to conclude that the second derivative $\sigma''(x)$ vanishes when the value of $\sigma(x)$ is $\frac{1}{2}$. According to Eq. (5), this happens for $x = 0$. Then we just proved that

$$\sigma''(0) = 0. \tag{20}$$

This is the only situation in which the function $\sigma''(x)$ vanishes. To understand this statement, notice that the other two roots of the polynomial $p(x)$ are $x = 0$ and $x = 1$. Then the second derivative would also vanish if the sigmoid function could yield the values 0 and 1. However, according to Eq. (4), $\sigma(x)$ returns real numbers *limited* by these values. Therefore, $x = 0$ is the only root of $\sigma''(x)$.

However, we are not finished. We still have to show that the sign of the second derivative changes at $x = 0$. To do so, we begin by rewriting $\sigma''(x)$ as follows:

$$\sigma''(x) = \sigma'(x) [1 - 2\sigma(x)]. \quad (21)$$

We already know that the first factor on the RHS of this equation is always positive. Then we only need to study the sign of the second factor. Define $f(x) = 1 - 2\sigma(x)$. With the aid of the results we obtained so far, it is straightforward to establish that this function has the following properties: (1) $f(x)$ is monotonically decreasing; (2) $f(x)$ vanishes at $x = 0$. These properties imply that $f(x) < 0$ for $x > 0$, and $f(x) > 0$ for $x < 0$. Since $f(x)$ and $\sigma''(x)$ have the same sign, an analogous statement about $\sigma''(x)$ is true. Hence, for $x = 0$, $\sigma''(x)$ equals zero, and this function has different signs for positive and negative values of x . This can be seen clearly in the graph of the second derivative, Fig. 3. Then we have completed our proof that the sigmoid function has an inflection point at $x = 0$.

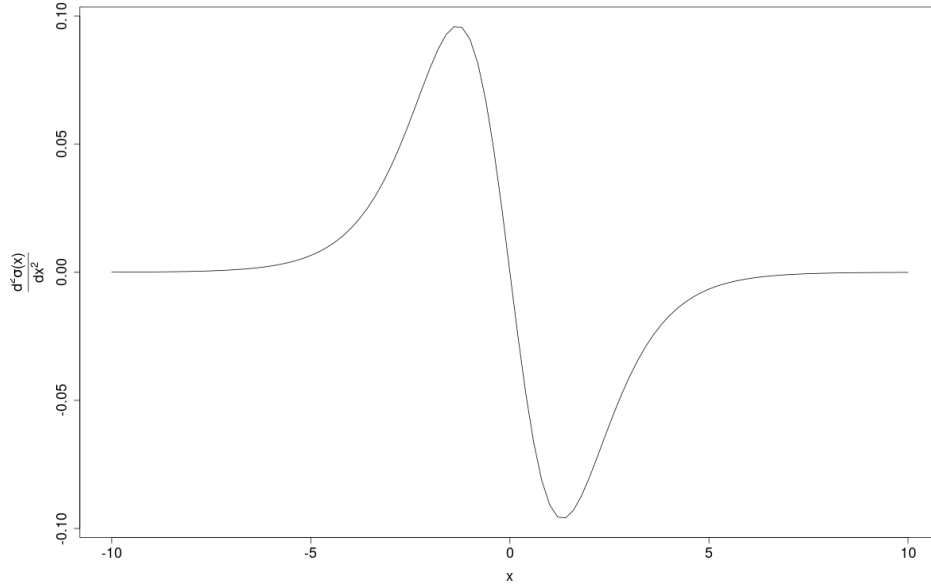


Figure 3: Graph of the second derivative of the sigmoid function.

Next, we derive explicit formulas for the second derivative. First, we obtain an expression in terms of the exponential function. To do so, we can use the following relation: $\sigma''(x) = \sigma'(x) f(x)$. We already have several equations for the first derivative. Then let us consider the function $f(x)$. It can be written as

$$\begin{aligned} f(x) &= 1 - 2\sigma(x) \\ &= 1 - \frac{2}{1 + \exp(-x)} \\ &= \frac{\exp(-x) - 1}{1 + \exp(-x)}. \end{aligned} \quad (22)$$

To continue, we multiply the last result by $\sigma'(x)$:

$$\begin{aligned} \sigma''(x) &= \sigma'(x) f(x) \\ &= \frac{\exp(-x)}{[1 + \exp(-x)]^2} \frac{\exp(-x) - 1}{1 + \exp(-x)} \\ &= \frac{\exp(-x) [\exp(-x) - 1]}{[1 + \exp(-x)]^3}. \end{aligned} \quad (23)$$

We can follow an analogous procedure to express the second derivative in terms of hyperbolic functions. We begin by rewriting the function $f(x)$. With the aid of Eq. (8), it is possible to write

$$\begin{aligned} f(x) &= 1 - 2\sigma(x) \\ &= 1 - 1 - \tanh\left(\frac{x}{2}\right) \\ &= -\tanh\left(\frac{x}{2}\right). \end{aligned} \quad (24)$$

Notice that the last equality shows that $f(x)$ is an odd function. We already established that the first derivative $\sigma'(x)$ is even. These facts allow us to conclude that the second derivative $\sigma''(x)$ is odd. To complete our derivation, we only need to multiply the above expression by our result for $\sigma'(x)$ in terms of the hyperbolic secant:

$$\begin{aligned}
 \sigma''(x) &= \sigma'(x) f(x) \\
 &= \frac{\operatorname{sech}^2\left(\frac{x}{2}\right)}{4} \left[-\tanh\left(\frac{x}{2}\right)\right] \\
 &= -\frac{\tanh\left(\frac{x}{2}\right) \operatorname{sech}^2\left(\frac{x}{2}\right)}{4}.
 \end{aligned} \tag{25}$$