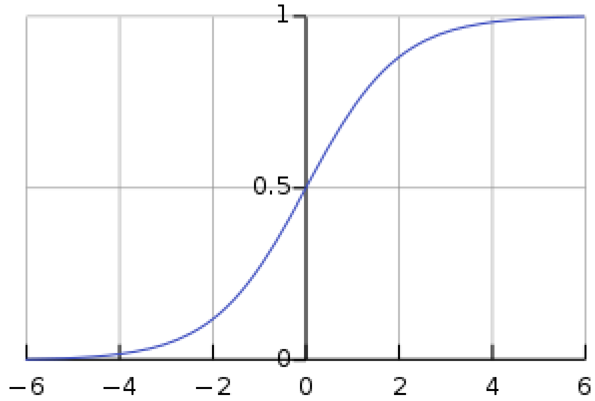
WIKIPEDIA

Sigmoid function

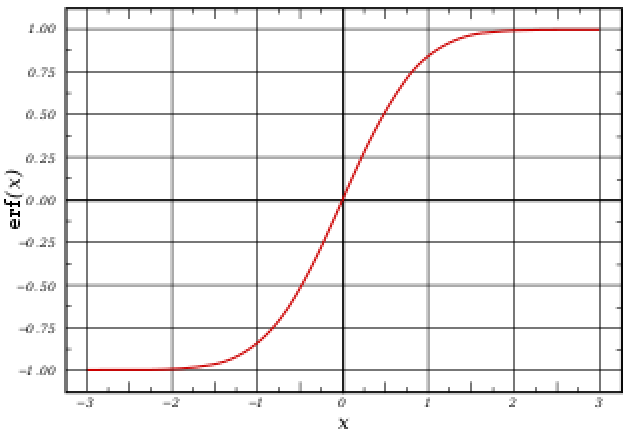
A **sigmoid function** is a mathematical function having a characteristic "S"-shaped curve or **sigmoid curve**. A standard choice for a *sigmoid function* is the logistic function shown in the first figure and defined by the formula

$$S(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1}.$$ [1] Other standard sigmoid functions are given in the Examples section.



The logistic curve

Special cases of the sigmoid function include the Gompertz curve (used in modeling systems that saturate at large values of x) and the ogee curve (used in the spillway of some dams). Sigmoid functions have domain of all real numbers, with return value monotonically increasing most often from 0 to 1 or alternatively from −1 to 1, depending on convention.



Plot of the error function

A wide variety of sigmoid functions including the logistic and hyperbolic tangent functions have been used as the activation function of artificial neurons. Sigmoid curves are also common in statistics as cumulative distribution functions (which go from 0 to 1), such as the integrals of the logistic distribution, the normal distribution, and Student's *t* probability density functions.

Contents

Definition

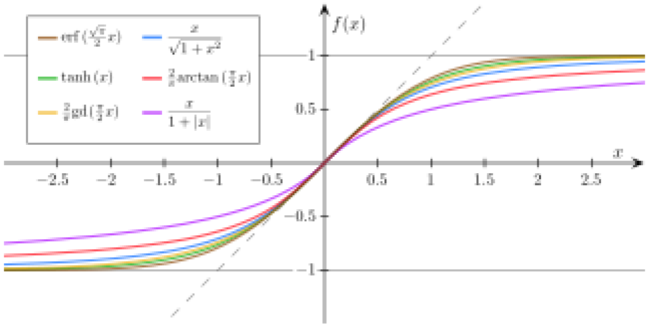
A sigmoid function is a bounded, differentiable, real function that is defined for all real input values and has a non-negative derivative at each point.[1] A sigmoid "function" and a sigmoid "curve" refer to the same object.

Properties

In general, a sigmoid function is monotonic, and has a first derivative which is bell shaped. A sigmoid function is constrained by a pair of horizontal asymptotes as $x \to \pm\infty$.

The sigmoid function is convex for values less than 0, and it is concave for values more than 0. Because of this, the sigmoid function and its affine compositions can possess multiple optima.

Examples



Some sigmoid functions compared. In the drawing all functions are normalized in such a way that their slope at the origin is 1.

- Logistic function

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

- Hyperbolic tangent (shifted and scaled version of the logistic function, above)

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

- Arctangent function

$$f(x) = \arctan x$$

- Gudermannian function

$$f(x) = \operatorname{gd}(x) = \int_0^x \frac{1}{\cosh t}\, dt = 2\arctan\left(\tanh\left(\frac{x}{2}\right)\right)$$

- Error function

$$f(x) = \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$$

- Generalised logistic function

$$f(x) = (1+e^{-x})^{-\alpha}, \quad \alpha > 0$$

- Smoothstep function

$$f(x) = \begin{cases} \dfrac{\int_0^x \left(1-u^2\right)^N du}{\int_0^1 \left(1-u^2\right)^N du}, & |x| \le 1 \\ \operatorname{sgn}(x) & |x| \ge 1 \end{cases} \qquad N \ge 1$$

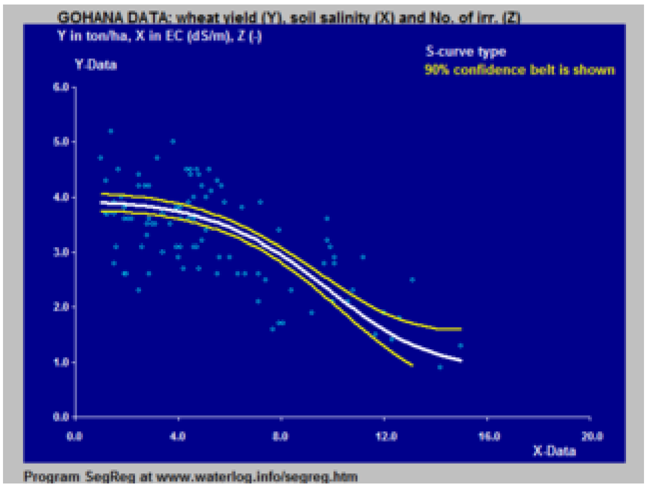
- Some algebraic functions, for example

$$f(x) = \frac{x}{\sqrt{1+x^2}}$$

The integral of any continuous, non-negative, "bump-shaped" function will be sigmoidal, thus the cumulative distribution functions for many common probability distributions are sigmoidal. One such example is the error function, which is related to the cumulative distribution function of a normal distribution.

Applications

Many natural processes, such as those of complex system learning curves, exhibit a progression from small beginnings that accelerates and approaches a climax over time. When a specific mathematical model is lacking, a sigmoid function is often used.[3]



Inverted logistic S-curve to model the relation between wheat yield and soil salinity. [2]

The van Genuchten–Gupta model is based on an inverted S-curve and applied to the response of crop yield to soil salinity.

Examples of the application of the logistic S-curve to the response of crop yield (wheat) to both the soil salinity and depth to water table in the soil are shown in logistic function#In agriculture: modeling crop response.

In artificial neural networks, sometimes non-smooth functions are used instead for efficiency; these are known as hard sigmoids.

In biochemistry and pharmacology, the Hill equation and Hill-Langmuir equation are sigmoid functions.

See also

- Heaviside step function
- Logistic regression
- Logit
- Softplus function
- Softmax function
- Weibull distribution
- Fermi–Dirac statistics

References

1. Han, Jun; Morag, Claudio (1995). "The influence of the sigmoid function parameters on the speed of backpropagation learning". In Mira, José; Sandoval, Francisco (eds.). *From Natural to Artificial Neural Computation*. Lecture Notes in Computer Science. **930**. pp. 195–201. doi:10.1007/3-540-59497-3_175 (https://doi.org/10.1007%2F3-540-59497-3_175). ISBN 978-3-540-59497-0.
2. Software to fit an S-curve to a data set [1] (https://www.waterlog.info/sigmoid.htm)
3. Gibbs, M.N. (Nov 2000). "Variational Gaussian process classifiers". *IEEE Transactions on Neural Networks*. **11** (6): 1458–1464. doi:10.1109/72.883477 (https://doi.org/10.1109%2F72.883477). PMID 18249869 (https://www.ncbi.nlm.nih.gov/pubmed/18249869).

- Mitchell, Tom M. (1997). *Machine Learning*. WCB–McGraw–Hill. ISBN 978-0-07-042807-2.. In particular see "Chapter 4: Artificial Neural Networks" (in particular pp. 96–97) where Mitchell uses the word "logistic function" and the "sigmoid function" synonymously – this function he also calls the "squashing function" – and the sigmoid (aka logistic) function is used to compress the outputs of the "neurons" in multi-layer neural nets.
- Humphrys, Mark. "Continuous output, the sigmoid function" (http://www.computing.dcu.ie/~humphrys/Notes/Neural/sigmoid.html). Properties of the sigmoid, including how it can shift along axes and how its domain may be transformed.