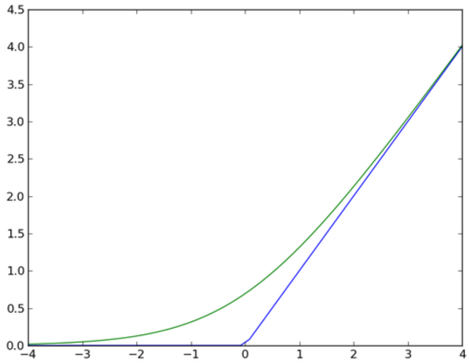
Rectifier (neural networks)



Plot of the rectifier (blue) and softplus (green) functions near x = 0.

In the context of artificial neural networks, the **rectifier** is an activation function defined as

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

where *x* is the input to a neuron. This activation function has been argued to be more biologically plausible (*cortical neurons are rarely in their maximum saturation regime*) [12] than the widely used logistic sigmoid (which is inspired by probability theory; see logistic regression) and its more practical[13] counterpart, the hyperbolic tangent.

A unit employing the rectifier is also called a **rectified linear unit** (**ReLU**).[14] A smooth approximate to the rectifier is the analytic function

$$f(x) = \log(1 + e^x)$$

which is called the **softplus** function.[15] The derivative of softplus is $f'(x) = e^x/(e^x + 1) = 1/(1 + e^{-x})$, i.e. the logistic function.

Rectified linear units find applications in computer vision using deep neural nets.[12]

1 Variants

1.1 Noisy ReLUs

Rectified linear units can be extended to include Gaussian noise, making them noisy ReLUs, giving[14]

$$f(x) = \max(0, x + \mathcal{N}(0, \sigma(x)))$$

Noisy ReLUs have been used with some success in restricted Boltzmann machines for computer vision tasks.[14]

1.2 Leaky ReLUs

Leaky ReLUs allow a small, non-zero gradient when the unit is not active.[16]

$$f(x) = \begin{cases} z & \text{if } z > 0 \\ 0.01z & \text{otherwise} \end{cases}$$

2 Advantages

- **Biological plausibility**: One-sided, compared to the antisymmetry of tanh.
- **Sparse activation**: For example, in a randomly initialized networks, only about 50% of hidden units is activated (having a non-zero output).
- **Efficient gradient propagation**: No vanishing gradient problem or exploding effect.
- **Efficient computation**: Only comparison, addition and multiplication.

Rectified linear units, compared to sigmoid function or similar activation functions, allow for faster and effective training of deep neural architectures on large and complex datasets. The common trait is that they implement local competition between small groups of units within a layer ($max(x, 0)$ can be interpreted as competition with a fixed value of 0), so that only part of the network is activated for any given input pattern.[17]

3 Potential problems

- **Ill-conditioning of parametrization**: There are infinitely many ways of setting values to parameters of a rectifier network to express an overall network function.[12]:319
- **Non-differentiable at zero**: however it is differentiable at any point arbitrarily close to 0.

4 References

[1] Boyce, Charles (1990). *Shakespeare A to Z*. New York: Roundtable Press.

[2] Brockett, Oscar G.; Hildy, Franklin J. (2007). *History of Theatre* (Foundation ed.). Boston: Allyn and Bacon.

[3] Bryson, Bill (2007). *Shakespeare: the world as stage*. New York: HarperCollins.

[4] Dunton-Downer, Leslie; Riding, Alan. (2004). *Essential Shakespeare Handbook*. New York: DK.

[5] Foakes, R.A. ed. (2002). *King Lear. The Arden Shakespeare*. London: Thompson Learning.

[6] Greenbalt, Stephen (1997). *Comedies. The Norton Shakespeare based on the Oxford ed*. New York: W.W. Norton &Co.

[7] Hoy, Cyrus, ed (1992). *Hamlet. Norton ciritical ed*. New York: W.W. Norton & Co.

[8] *The Illustrated Library Shakespeare*. Bath: Robert Frederick, Ltd. 2004.

[9] Jamieson, Lee. "Shakespeare Tragedies". *About.com*. Retrieved 2014-10-04.

[10] McEachern, Claire, ed (2013). *The Cambridge Companion to Shakespearean Tragedy*. Cambridge: Cambridge UP.

[11] Mowat, Barbara A.; Werstine, Paul eds (2013). *The Tragedy of Julius Caesar. Folger Shakespeare Library*. New York: Washington P.

[12] Xavier Glorot, Antoine Bordes and Yoshua Bengio (2011). "Deep sparse rectifier neural networks". AISTATS.

[13] Yann LeCun, Leon Bottou, Genevieve B. Orr and Klaus-Robert Müller (1998). "Efficient BackProp". In G. Orr and K. Müller. *Neural Networks: Tricks of the Trade*. Springer.

[14] Vinod Nair and Geoffrey Hinton (2010). "Rectified linear units improve restricted Boltzmann machines". ICML.

[15] F. Bélisle, Y. Bengio, C. Dugas, R. Garcia, C. Nadeau (2001). Second-Order Functional Knowledge for Better Option Pricing

[16] Andrew L. Maas, Awni Y. Hannun, Andrew Y. Ng. Rectifier Nonlinearities Improve Neural Network Acoustic Models. 2014

[17] RK Srivastava, J Masci,F Gomez, and J Schmidhuber (2014). "Understanding Locally Competitive Networks". arXiv.

5 Text and image sources, contributors, and licenses

5.1 Text

- **Rectifier (neural networks)** *Source:* http://en.wikipedia.org/wiki/Rectifier_(neural_networks)?oldid=629861804 *Contributors:* Glenn, Tobias Bergemann, Qwertyus, Kri, Bhny, Johndburger, Spymark, Rakeshchalasani, Ngocminh.oss, Khazar2, Ginsuloft, Enerjiparki, Velvel2 and Anonymous: 3

5.2 Images

- **File:LampFlowchart.svg** *Source:* http://upload.wikimedia.org/wikipedia/commons/9/91/LampFlowchart.svg *License:* CC-BY-SA-3.0 *Contributors:* vector version of Image:LampFlowchart.png *Original artist:* svg by Booyabazooka

- **File:Rectifier_and_softplus_functions.png** *Source:* http://upload.wikimedia.org/wikipedia/commons/f/f7/Rectifier_and_softplus_functions.png *License:* CC0 *Contributors:* Own work *Original artist:* Qwertyus

5.3 Content license