Artificial neural networks and backpropagation

E. Decencière

MINES ParisTech
PSL Research University
Center for Mathematical Morphology

Contents

Contents

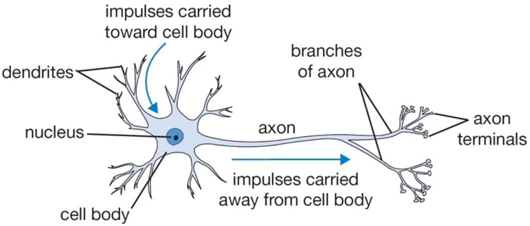
Artificial neural networks and deep learning history

For a very complete state of the art on deep learning, see the overview by Schmidhuber [Schmidhuber, 2015].

- 1958: Rosenblatt's perceptron [Rosenblatt, 1958]
- 1980's: the backpropagation algorithm (see, for example, the work of Le Cun [LeCun, 1985])
- 2006-: CNN implementations using Graphical Processing Units (GPU): up to a 50 speed-up factor.
- 2011-: super-human performances [Cireșan et al., 2011]
- 2012: Imagenet image classification won by a CNN [Krizhevsky et al., 2012].

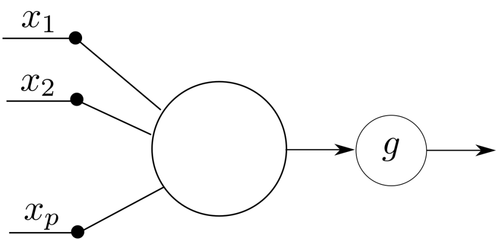
Contents

Neuron



- The human brain contains 100 billion (10^{11}) neurons
- A human neuron can have several thousand dendrites

Artificial neuron

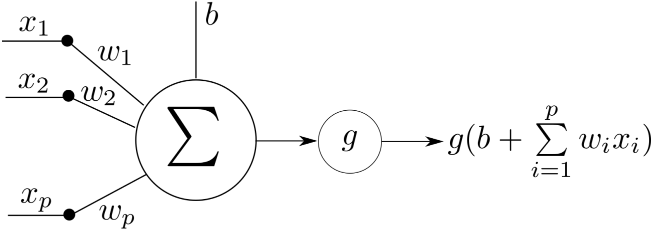


General principle

An artificial neuron takes p inputs $\{x_i\}_{1 \leq i \leq p}$, combines them to obtain a single value, and applies an activation function to the result.

- The first artificial neuron model was proposed by [McCulloch and Pitts, 1943]
- Input and output signals were binary
- Input dendrites could be inhibitory or excitatory

Artificial neuron

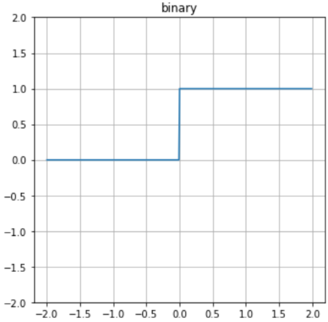


Classical neuron

Bias **b** is a variable linked to the neuron. The activation function g somehow decides, depending on the input, if a signal (the neuron's activation) is produced.

Activation: binary

$$g(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

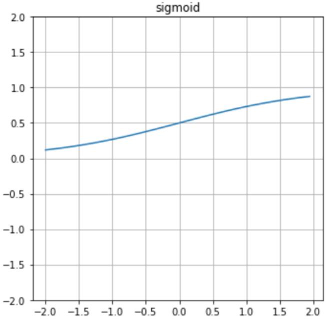


Remarks

- Biologically inspired
+ Simple to compute
+ High abstraction
- Gradient nil except on one point

Activation: sigmoid

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

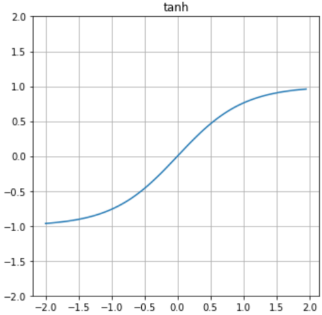


Remarks

+ Similar to binary activation, but with usable gradient
- However, gradient tends to zero when input is far from zero
- More computationally intensive

Activation: hyperbolic tangent

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

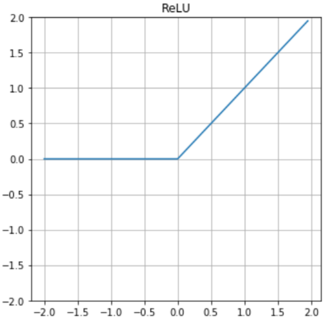


Remarks

- Similar to sigmoid

Activation: rectified linear unit

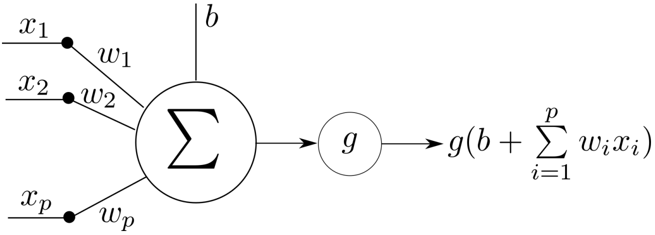
$$g(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$$



Remarks

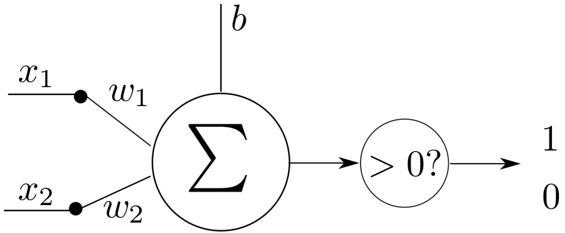
+ Usable gradient when activated
+ Fast to compute
+ High abstraction

What can an artifical neuron compute?

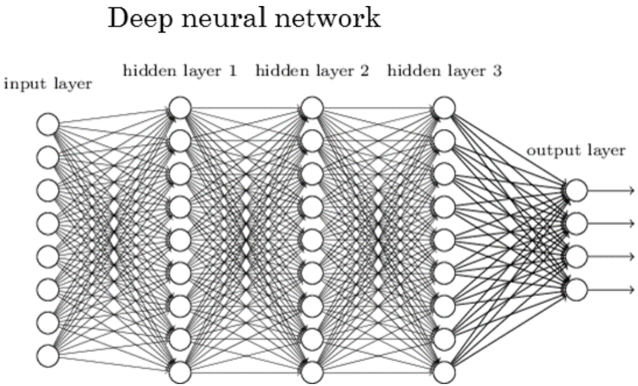


In $\mathbb{R}^p$, $b + \sum_{i=0}^{p} w_i x_i = 0$ gives a hyperplane. For a given point $\mathbf{x} = \{x_0, \ldots, x_p\}$, decisions are made according to the side of the hyperplane it belongs to.

The power of an artificial neuron: illustration

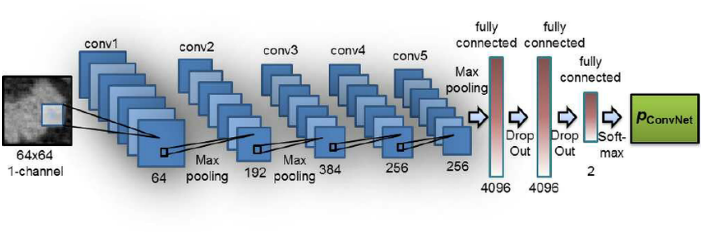


Neural network



(from http://www.jtoy.net)

Convolutional neural networks (ConvNets or CNNs)



(from https://www.researchgate.net)

The triggering factor to the success of neural networks

- Appropriate architectures: graphical processing units (GPUs)
- Optimized software
- Large annotated databases

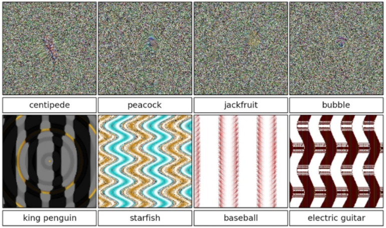
Contents

Practical considerations

For a deep-learning solution to work, you need:

- A lot of annotated data
- A lot of fiddling (different architectures; hyper-parameters)
- GPUs, at least from training

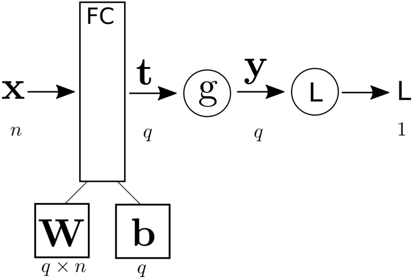
Deep learning can produce astonishing results
[Nguyen et al., 2015]...



The web giants

- Google, Facebook, Microsoft, Amazon etc. are actively investing in deep-learning
- Competition is intense
- Most of them are sharing their deep learning libraries

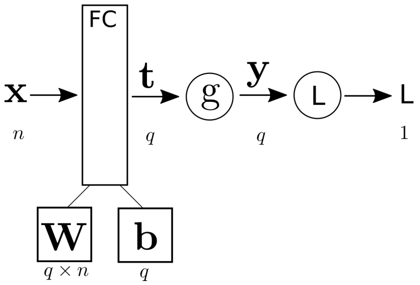
Backpropagation through a fully connected layer



Setup:

$$n, q \in \mathbb{N}^*$$
$$\mathbf{x} \in \mathbb{R}^n$$
$$\mathbf{W} \in \mathbb{R}^q \times \mathbb{R}^n$$
$$\mathbf{b}, \mathbf{t}, \mathbf{y} \in \mathbb{R}^q$$
$$L \in \mathbb{R}$$

Backpropagation through a fully connected layer



Local gradients:

$$\frac{\partial \mathbf{t}}{\partial \mathbf{W}} = \mathbf{x}^t$$

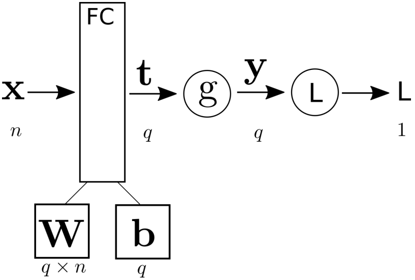
$$\frac{\partial \mathbf{t}}{\partial \mathbf{b}} = 1$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{t}} = \mathrm{g}'$$

Forward pass:

$$\begin{aligned} \mathbf{t} &= \mathbf{W}\mathbf{x} + \mathbf{b} \\ \mathbf{y} &= \mathrm{g}(\mathbf{W}\mathbf{x} + \mathbf{b}) \\ L &= L(\mathbf{y}) \end{aligned}$$

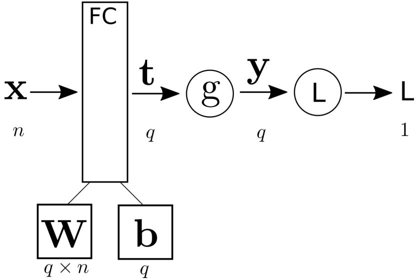
Backpropagation through a fully connected layer



Backpropagation:

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{t}} &= \frac{\partial L}{\partial \mathbf{y}} \cdot \frac{\partial \mathbf{y}}{\partial \mathbf{t}} \\ &= \frac{\partial L}{\partial \mathbf{y}} \odot \mathrm{g}'(\mathbf{t}) \end{aligned}$$

Backpropagation through a fully connected layer



Backpropagation:

$$\begin{aligned}\frac{\partial L}{\partial \mathbf{W}} &= \frac{\partial L}{\partial \mathbf{t}} . \frac{\partial \mathbf{t}}{\partial \mathbf{W}} \\ &= \frac{\partial L}{\partial \mathbf{y}} \odot g'(\mathbf{t}).\mathbf{x}^t\end{aligned}$$

$$\frac{\partial L}{\partial \mathbf{b}} = \frac{\partial L}{\partial \mathbf{y}} \odot g'(\mathbf{t})$$

References I

[Cireşan et al., 2011] Cireşan, D., Meier, U., Masci, J., and Schmidhuber, J. (2011). A committee of neural networks for traffic sign classification. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 1918–1921. IEEE.

[Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc.

[LeCun, 1985] LeCun, Y. (1985). Une procedure d'apprentissage pour reseau a seuil asymmetrique (A learning scheme for asymmetric threshold networks).

[McCulloch and Pitts, 1943] McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.

[Nguyen et al., 2015] Nguyen, A., Yosinski, J., and Clune, J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 427–436.

[Rosenblatt, 1958] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408.

References II

[Schmidhuber, 2015] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117.