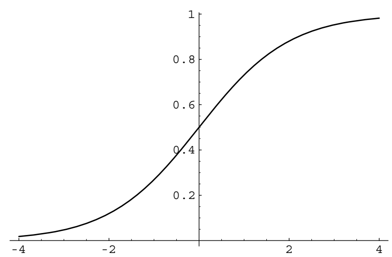
Neural activation functions

In a neural network, each neuron has an *activation function* which specifies the output of a neuron to a given input. Neurons are 'switches' that output a '1' when they are sufficiently activated, and a '0' when not. One of the activation functions commonly used for neurons is the *sigmoid* function:

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = \frac{1}{1+e^{-x}} \tag{1}$$

This function looks like an S (hence its name):



When one uses the neural network to learn, the updating of the contributions depends on the steepness (slope) of the activation function. How and why is explained in the lectures on neural networks (with the *delta rule* and *back-propagation*) and we cannot treat that here. But mathematically, the steepness of the activation function is given by the *derivative*, and we can compute with that. To save computation, the neural network engineer attempts to express this derivative in terms of the function value itself, so to express $f'(x)$ in terms of ($f(x)$).

(1) Verify that the figure indeed represents the sigmoid function: check what happens when $x \to -\infty$, $x \to \infty$, and at $x = 0$.

(2) From the figure, it seems that the function has a certain symmetry around the point $(0, \frac{1}{2})$. How would you express that symmetry mathematically ? Can you prove that it holds? (**Answer**: $f(x) - f(0) = f(0) - f(-x)$, so $f(x) + f(-x) = 2f(0)$.)

(3) Compute the derivative $f'(x)$.

(4) The formula for the derivative $f'(x)$ can be expressed as a function of $f(x)$. Do so. (**Answer**: $f'(x) = f(x)(1 - f(x))$.)

(5) The previous answer is a differential equation; its solution give the complete family of functions with this property. Give the family (so solve the differential equation!). How do the members differ. (**Answer**: $y = \frac{1}{1+ce^{-x}}$, with $c = 0$; and also $y = 0$ and $y = 1$. Apart from those two, the members differ in steepness.

Now a neural network engineer wants to use the final neuron to change the direction of a motor. To do so, she prefers its output to be between -1 and 1. She decides to use a new activation function g derived simply from the old as:

$$g(x) = 2f(x) - 1 \tag{2}$$

Again, she wants to save computation by expressing the derivative in terms of the function value, so to express $g'(x)$ in terms of $g(x)$. But she needs to be careful!

(4) Check that this is indeed a good choice (draw this new function).

(5) From the definition of g, you should expect g' to be just twice f'. Verify this by computing the derivative.

(6) Express $g'(x)$ in terms of $g(x)$. (**Answer**: $g'(x) = \frac{1}{2}(1 - g(x)^2)$.)

(7) Note that the derivative $f'(x)$ in terms of x becomes twice as big, but that the formula for the derivative in terms of $g(x)$ changes more than just by a factor of 2: it is $\frac{1}{2}(1 - g(x)^2)$ rather than $2g(x)(1 - g(x))$!

Moral: be careful with derivatives re-expressed in terms of the function values!