Given

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

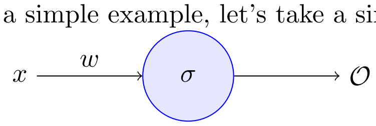
recall that the derivative

$$\frac{\partial \sigma}{\partial x} = \sigma(x)(1-\sigma(x))$$

Also recall that the error function is given by

$$E = \frac{1}{2}\sum_{t\in\text{Data}} (\mathcal{O}-t)^2$$

As a simple example, let's take a single neuron with only one incoming weight and no bias.



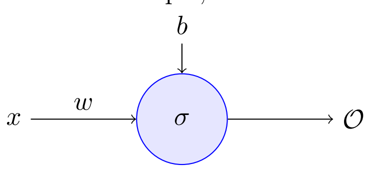
The equation which defines this very tiny network is

$$\mathcal{O} = \sigma(wx)$$

So, let's compute the derivative of the error E wrt the weight w:

$$\frac{\partial E}{\partial w} = (\mathcal{O}-t)\mathcal{O}(1-\mathcal{O})x = (\sigma(wx)-t)\sigma(wx)(1-\sigma(wx))x$$

As our next example, let's take a single neuron with only one incoming weight and a bias.



The equation which defines this very tiny network is

$$\mathcal{O} = \sigma(wx+b)$$

The derivative of the error works out exactly the same as before with respect to the weight, so let's do it for the bias b this time:

$$\frac{\partial E}{\partial b} = (\mathcal{O}-t)\mathcal{O}(1-\mathcal{O})\frac{\partial(wx+b)}{\partial b} = (\sigma(wx)-t)\sigma(wx)(1-\sigma(wx))$$