Back propagation

do feedforward and get the error

go at the opposit direction.

increase the weights that come from better classification and decrease ones that are worse.

more formally:

$$w_{ij}^{\prime(k)} \leftarrow w_{ij}^{(k)} - \alpha \frac{\partial E}{\partial w_{ij}^{(k)}}$$

Some reminders: chain rule

$A = f(x)$

$B = g \circ f(x)$

$$\frac{\partial B}{\partial x} = \frac{\partial B}{\partial A} \frac{\partial A}{\partial x}$$

$$\sigma'(x) = \sigma(x)(1-\sigma(x))$$

---

we now have all the ingredients to train neural networks!