Back propagation

do feedforward and get the error
go in the opposite direction.
increase the weights that come from better
classification and decrease ones that are worse.

more formally:

$$W_{ij}^{(k+1)} \leftarrow W_{ij}^{(k)} - \alpha \frac{\partial E}{\partial W_{ij}^{(k)}}$$

Some reminders: chain rule

$$A = f(x)$$
$$B = g \circ f(x)$$
$$\frac{\partial B}{\partial x} = \frac{\partial B}{\partial A} \frac{\partial A}{\partial x}$$

$$g'(x) = 6(x)(1 - g(x))$$

we now have all the ingredients to train
neural networks!