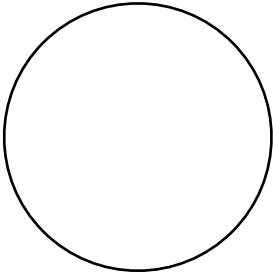
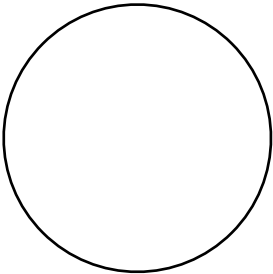


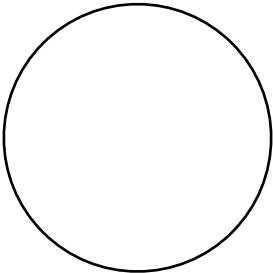
Gradient-based propagation



The most profitable application of the main rule of all time!



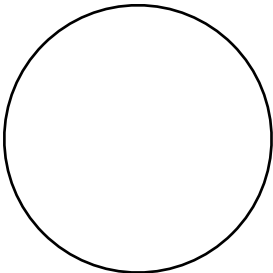








$$w_1(1)$$



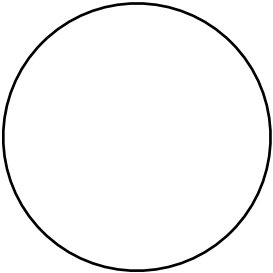
Q

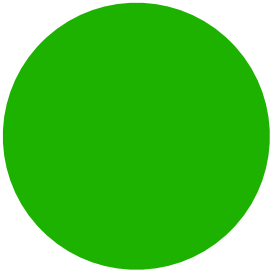
(

1

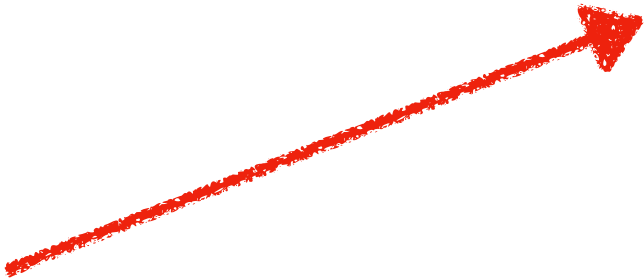
)

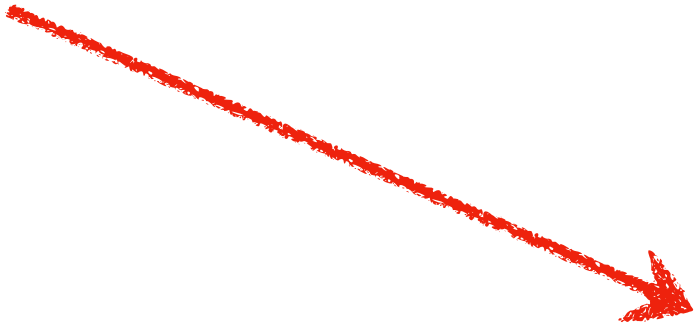
20

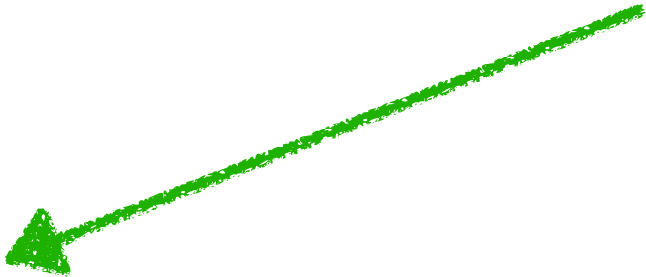
















2

(1)

$w_2(1)$



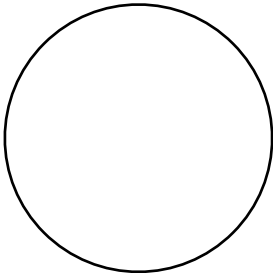






w(2)

RV



o

ccivvovv



C

E

∂CE

$\partial \sigma_{softmax}$

$$\frac{\partial \sigma_{softmax}}{\partial z(2)}$$

$\partial z^{(2)}$  $\partial w^{(2)}$

softmax



$\partial a^{(1)}$

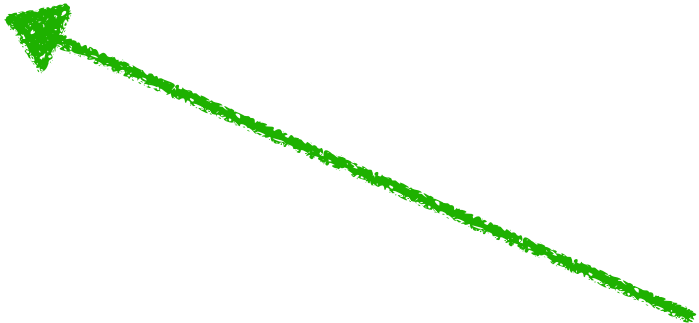
 $\partial w_2^{(1)}$

$\partial a^{(1)}$  $\partial w_1^{(1)}$

$\partial \sigma_{\text{activation}}$

$\partial a^{(1)}$

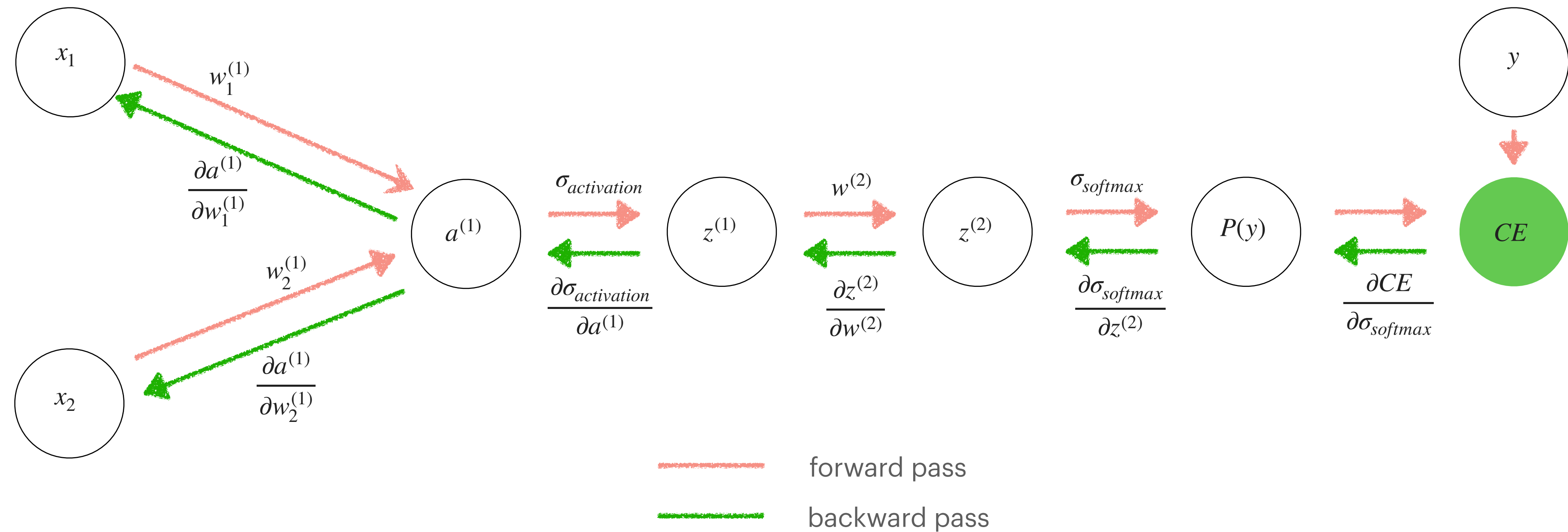




forward pass

backwards

Gradient backpropagation



The most profitable application of the chain rule of all time!

Neural network visualization