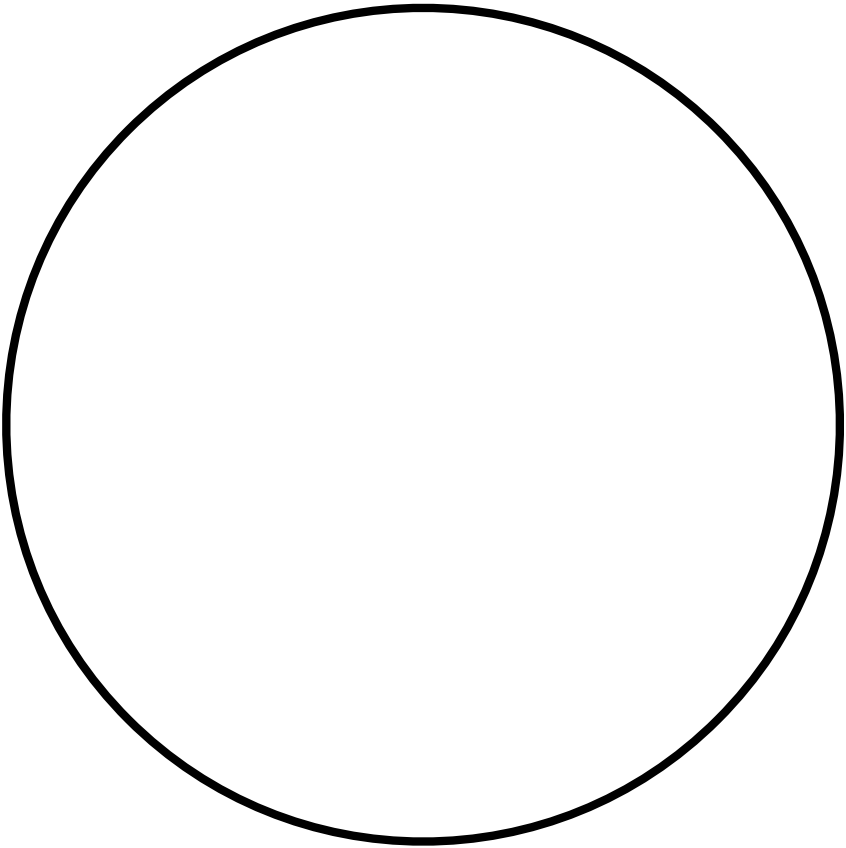
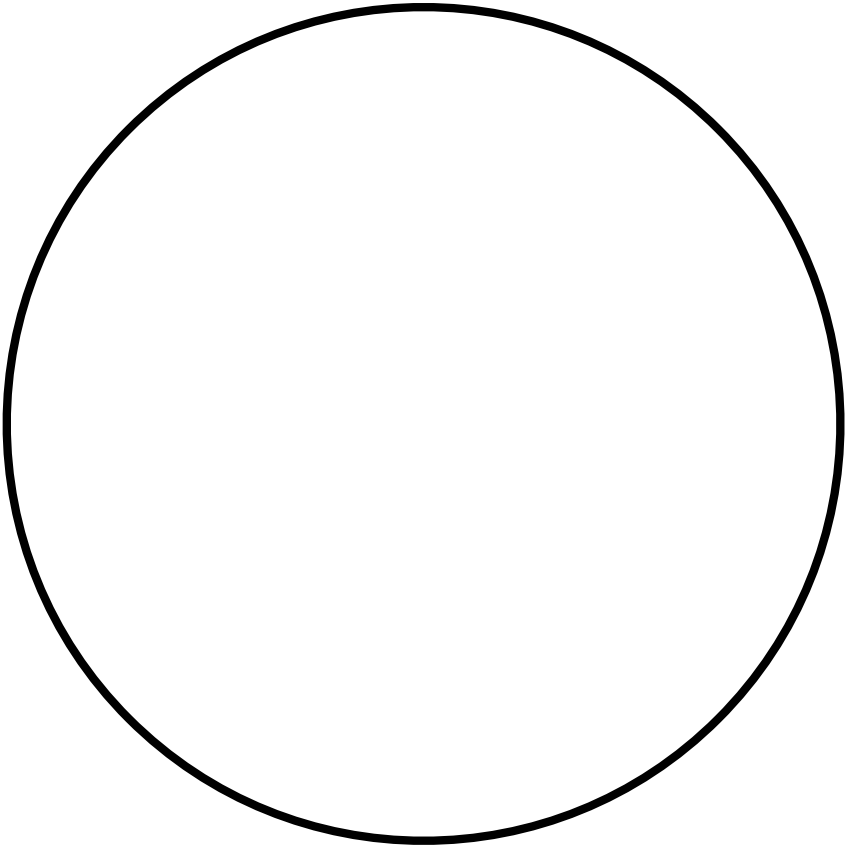
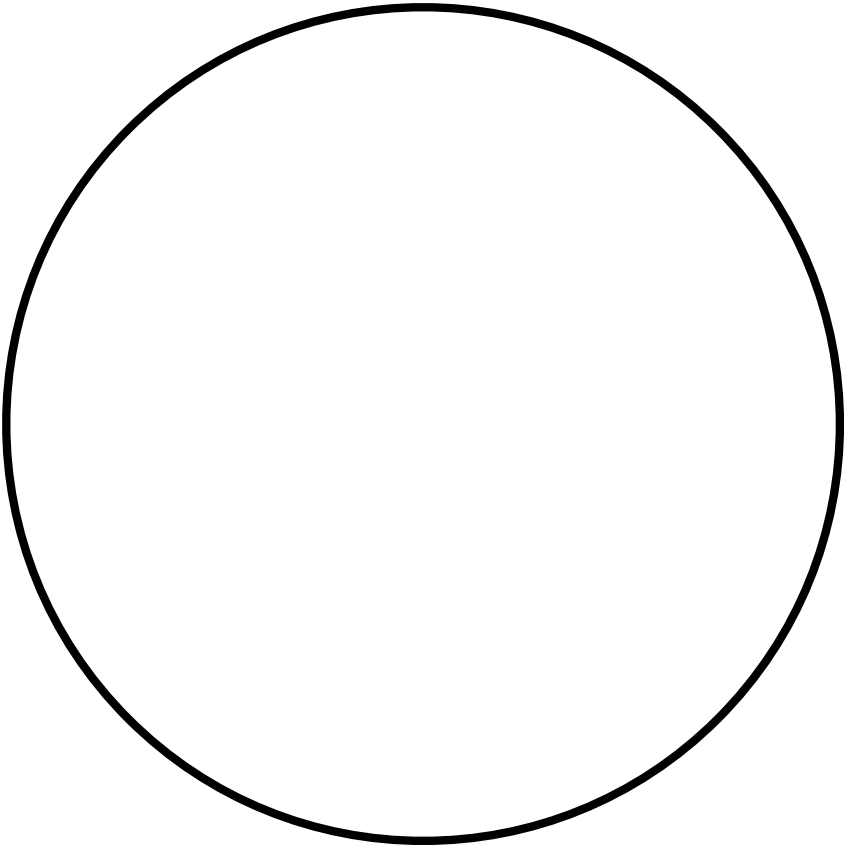
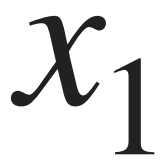
Gradient backpropagation

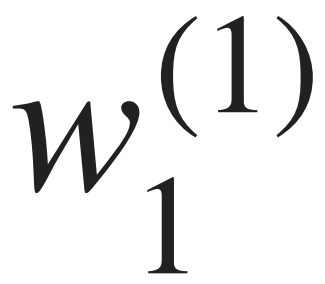
The most profitable application of the chain rule of all time!

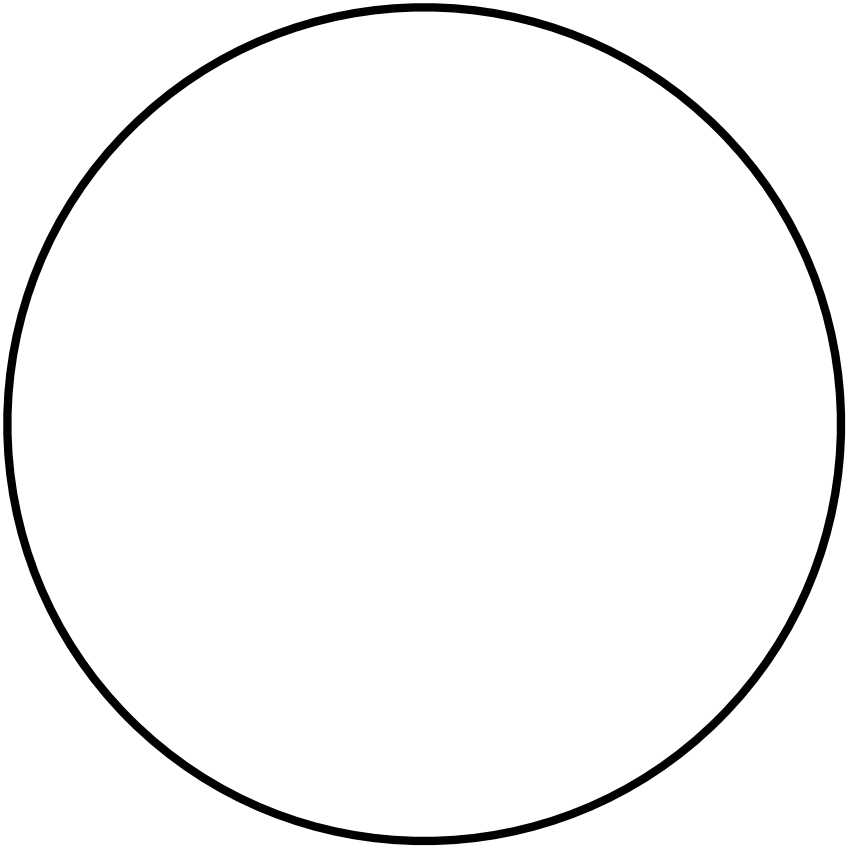


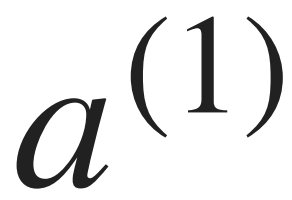


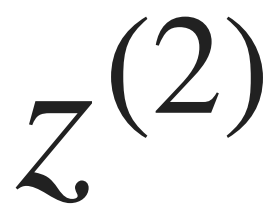


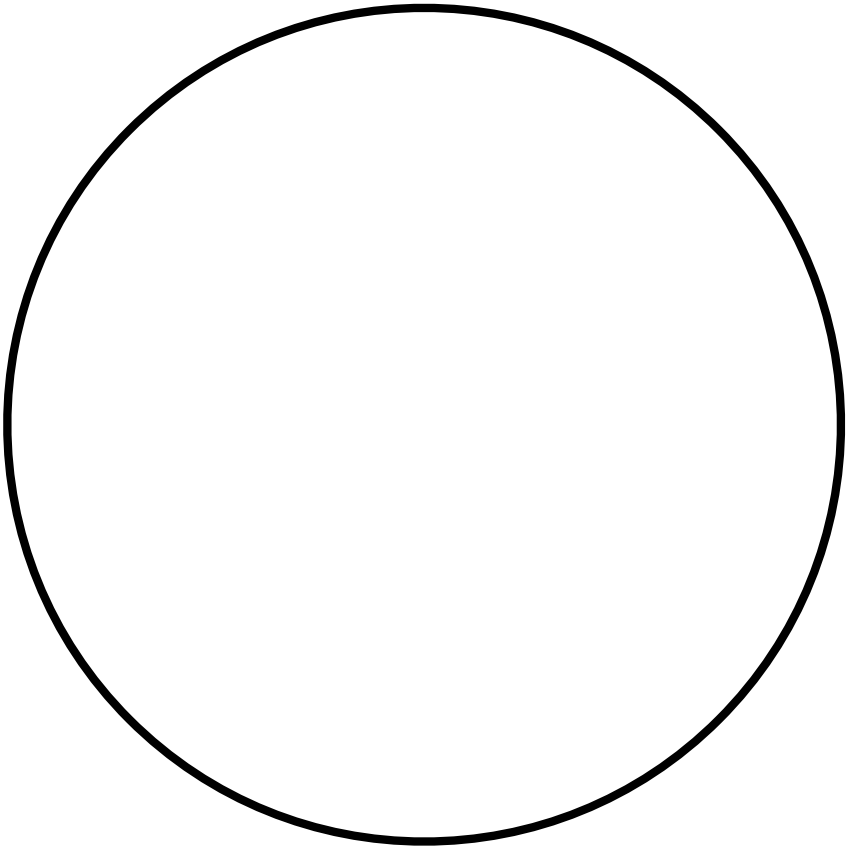


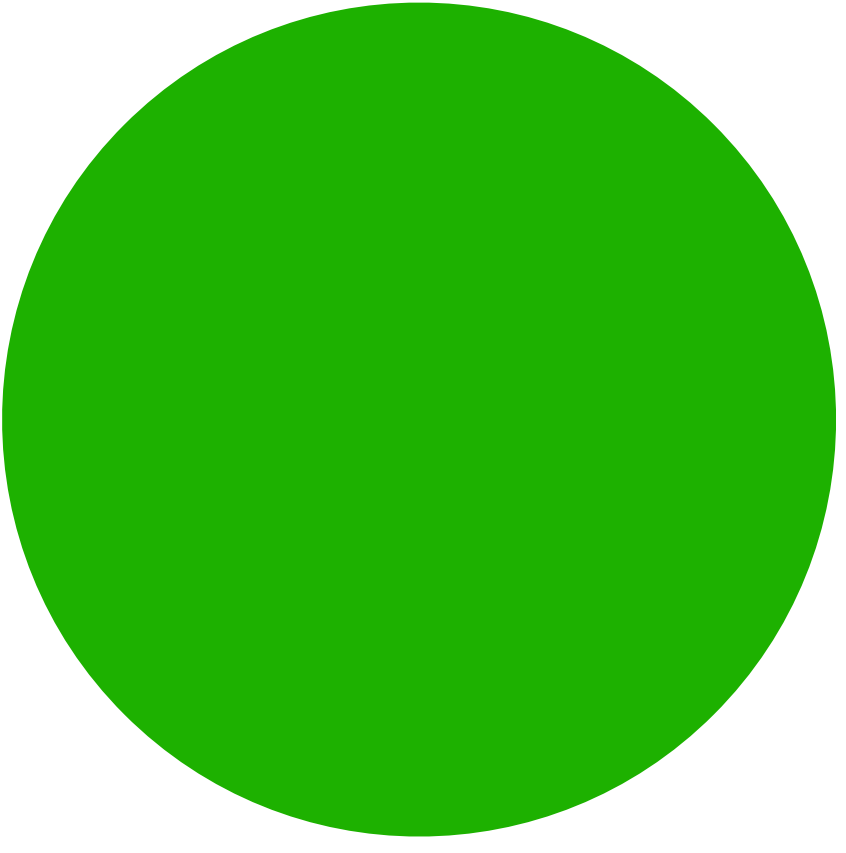


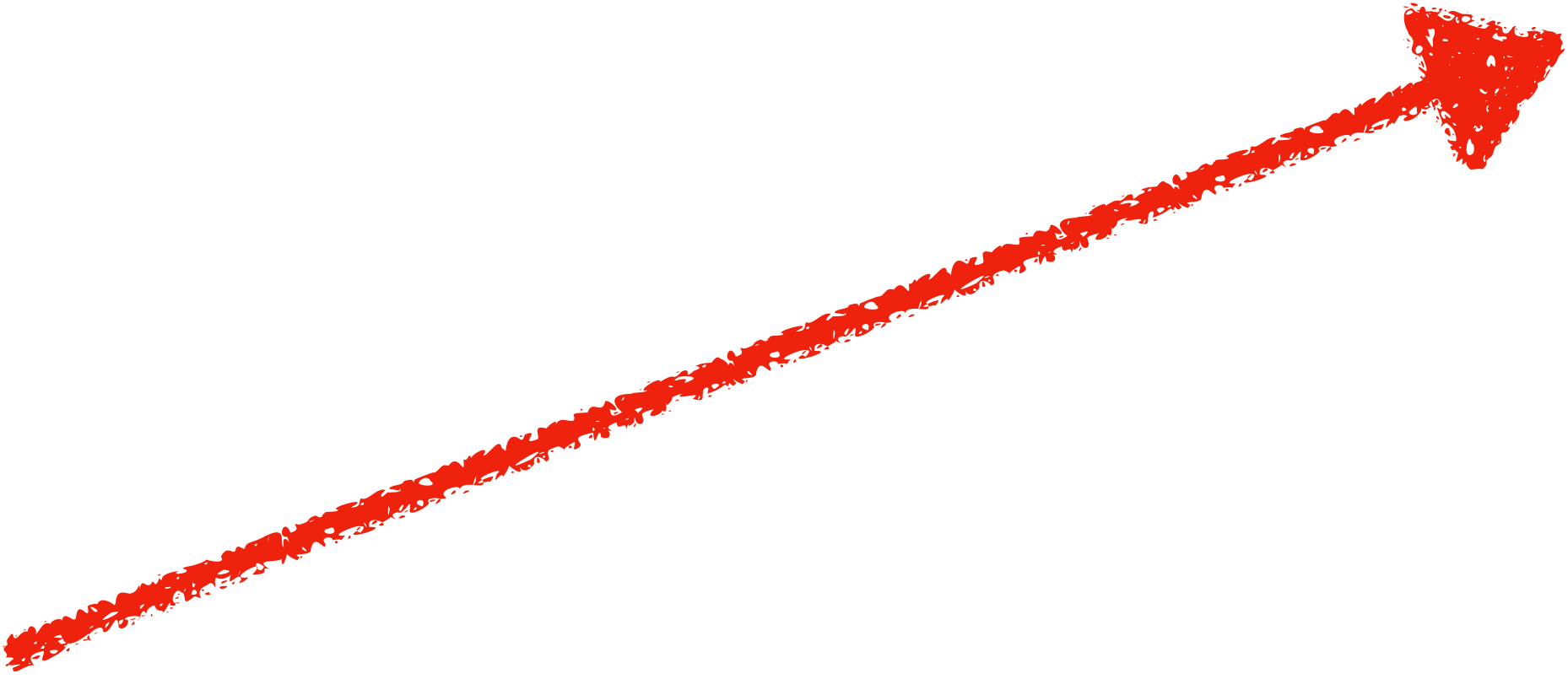


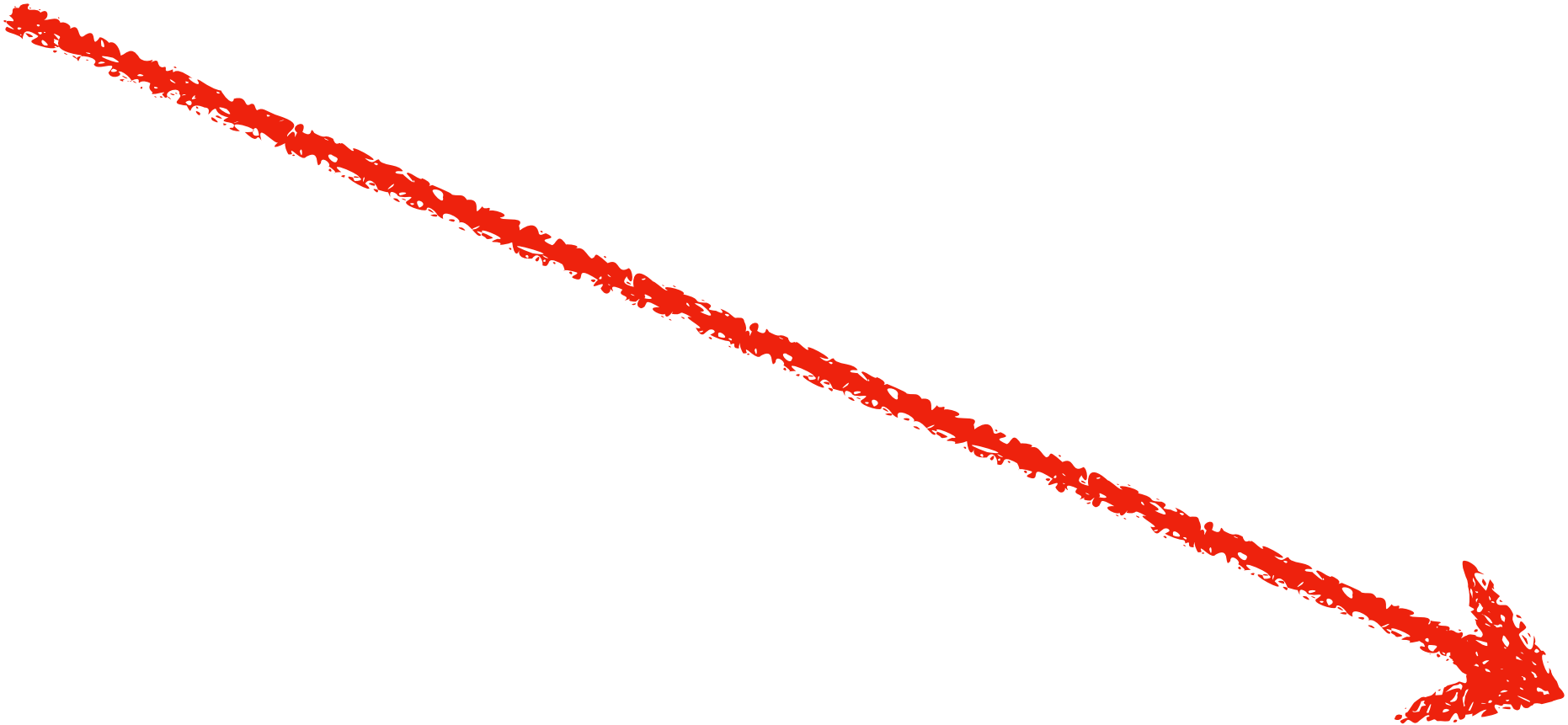


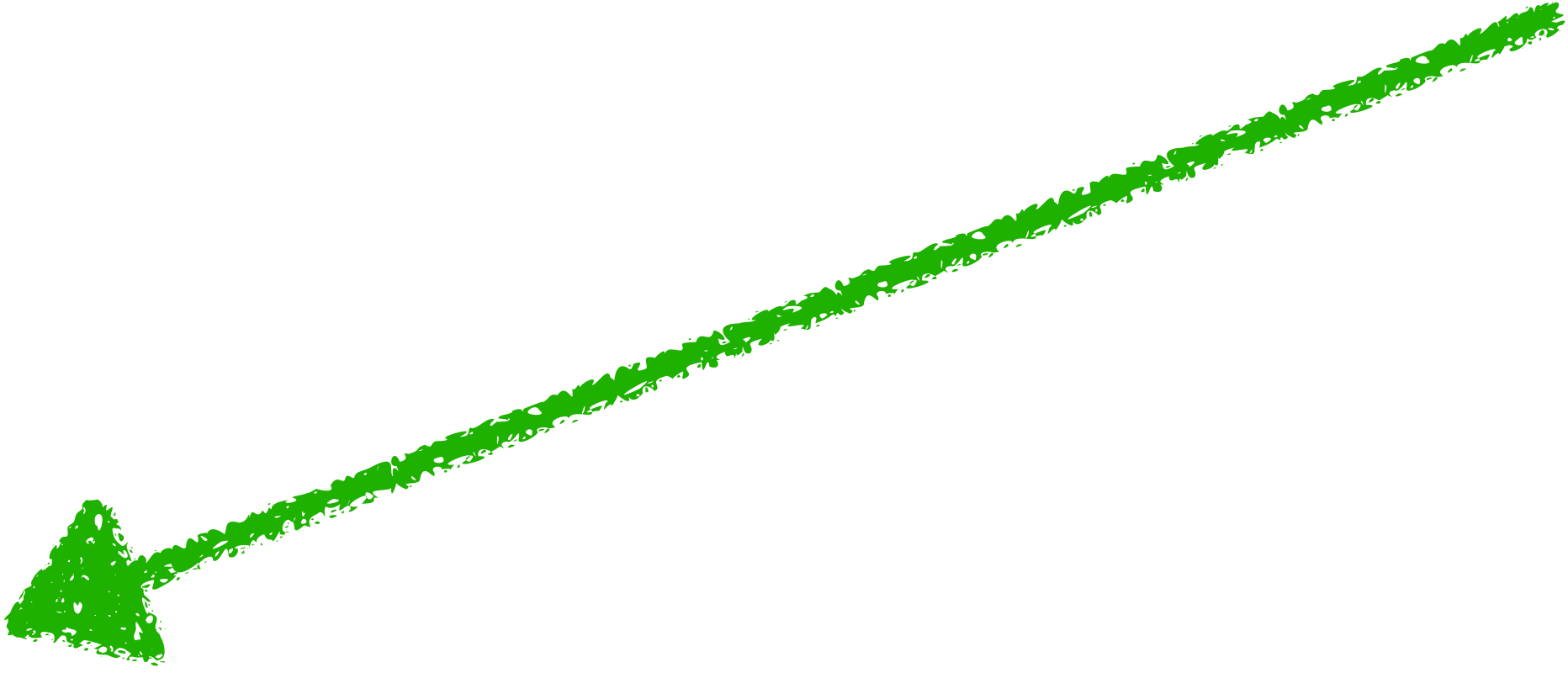


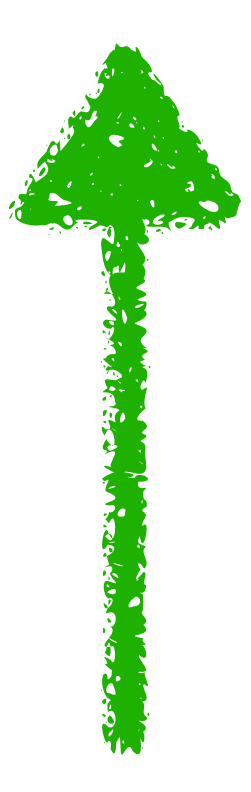


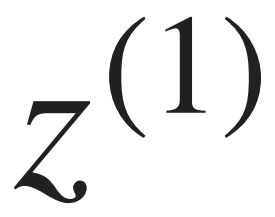


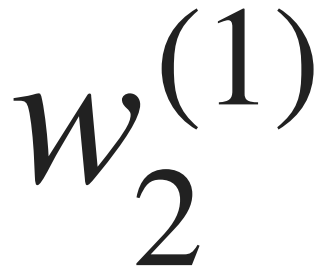


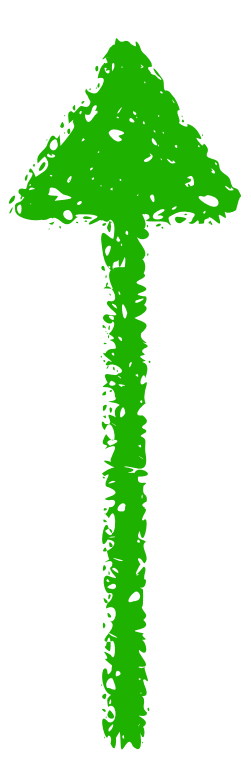


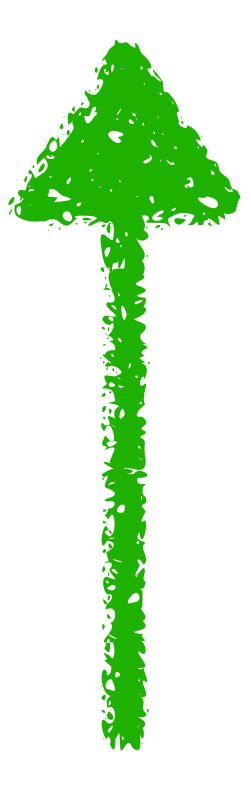


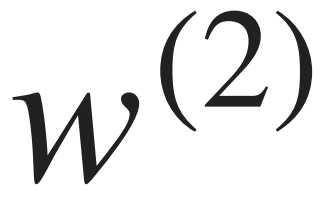


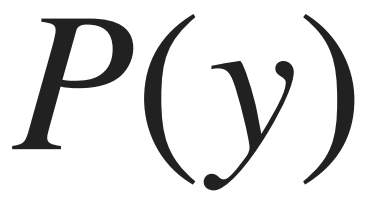


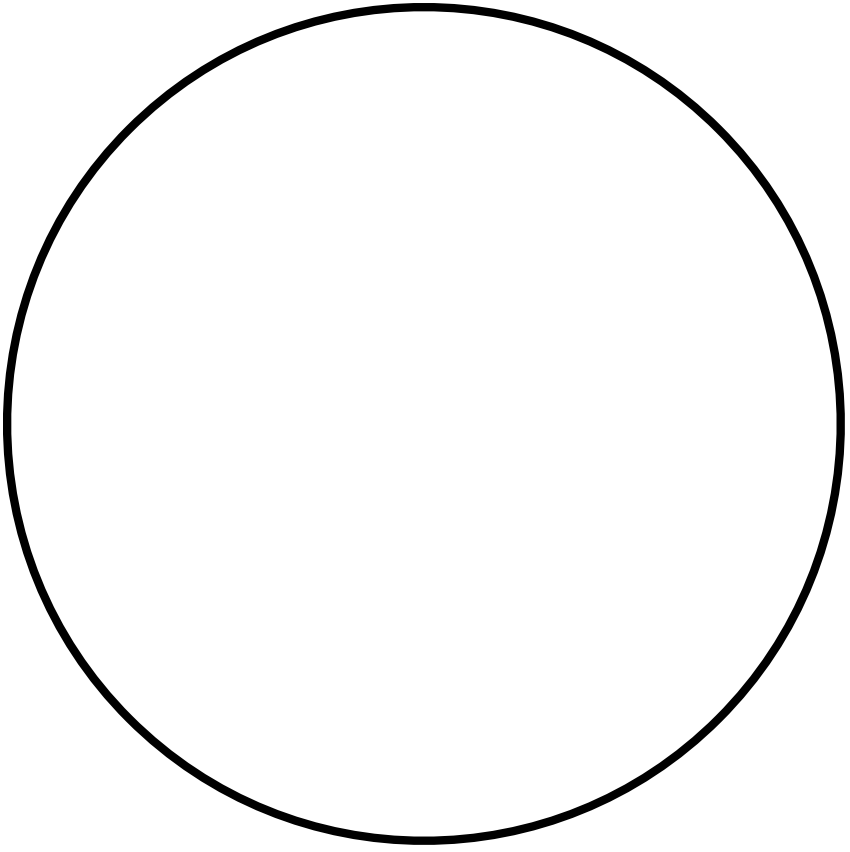

$w_2^{(1)}$





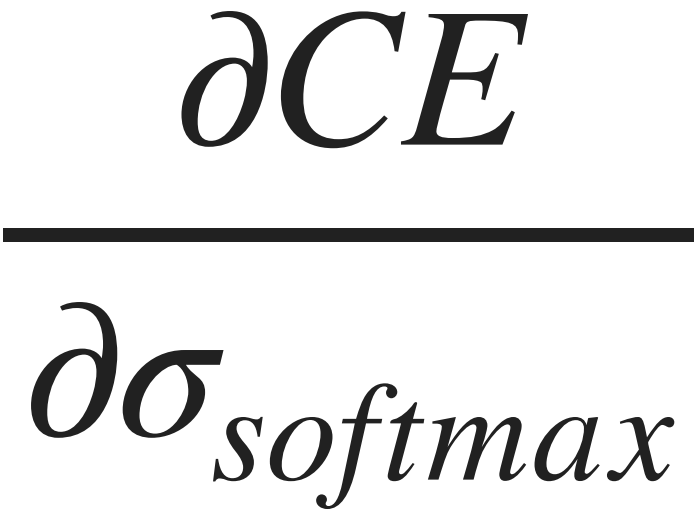


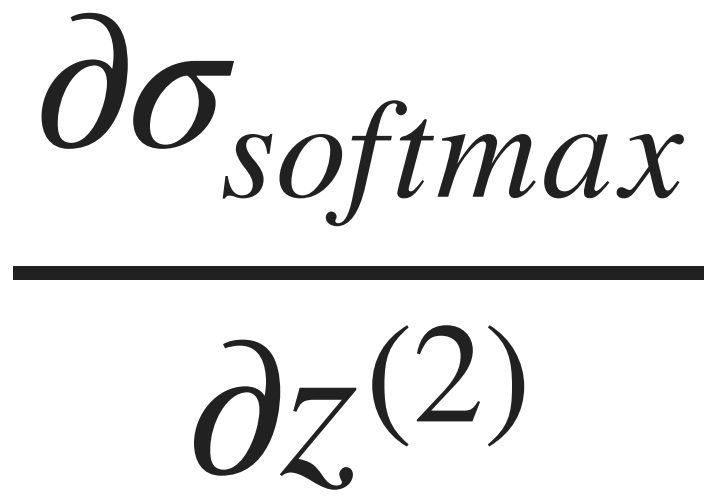

P(y)

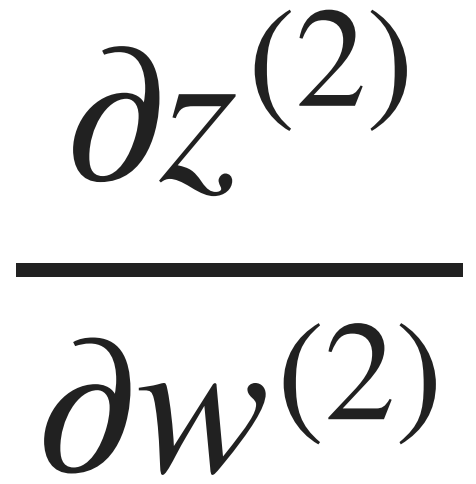


$\sigma_{activation}$

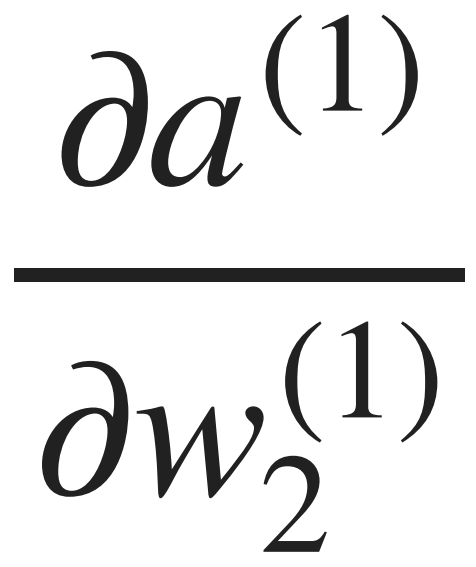
CE

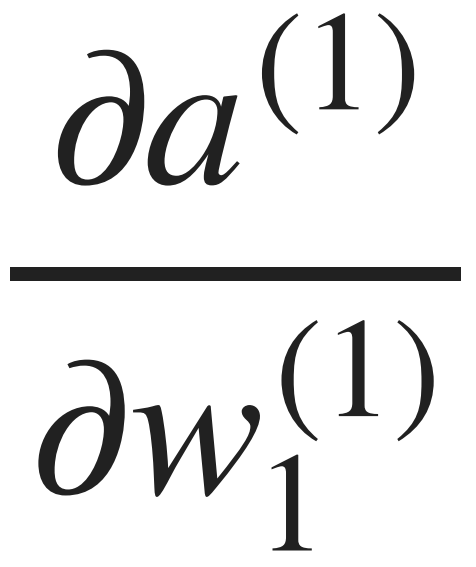

$\frac{\partial CE}{\partial \sigma_{softmax}}$

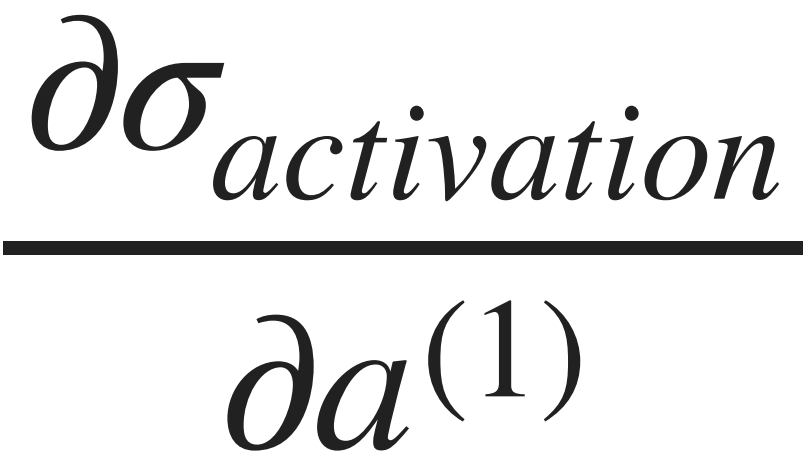


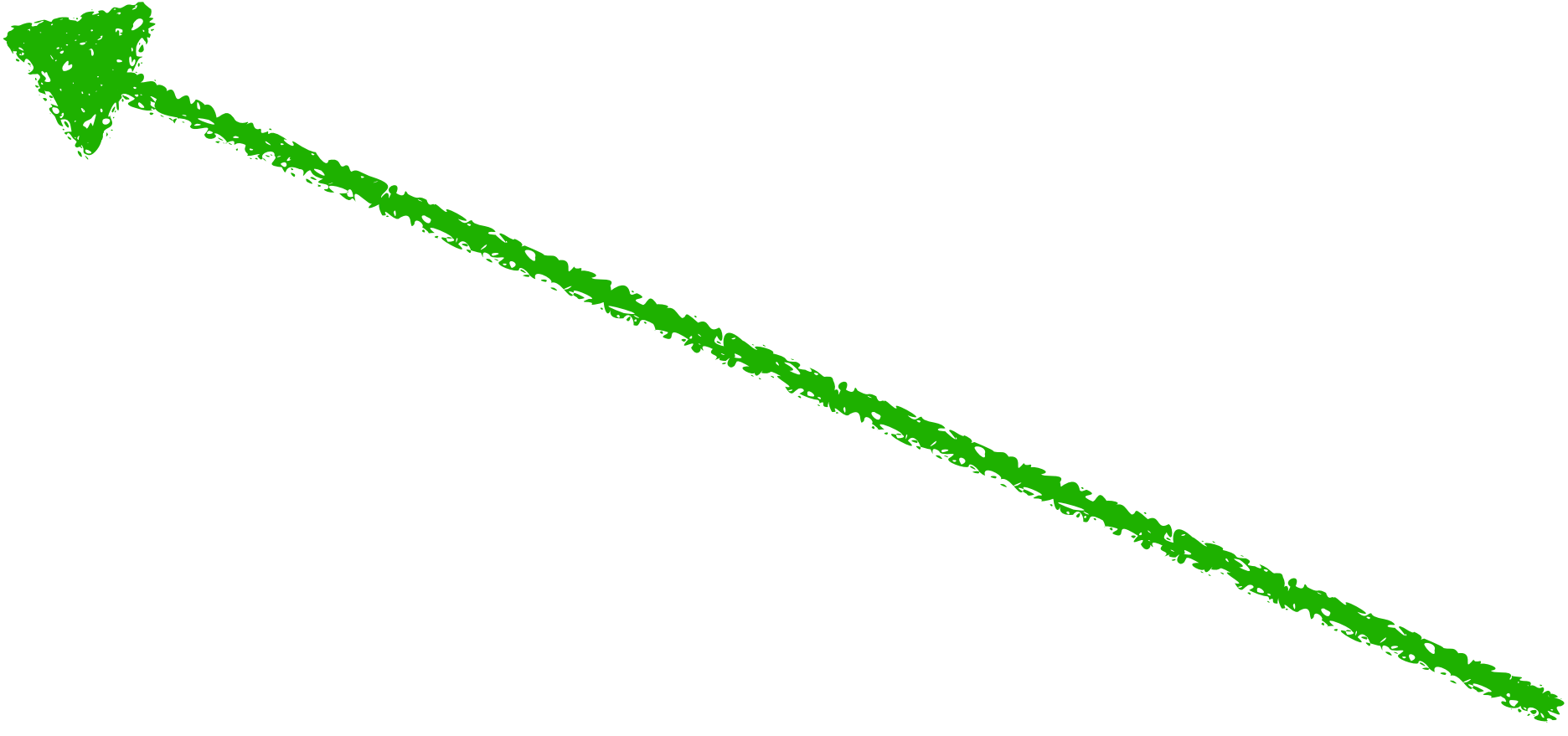


σsoftmax





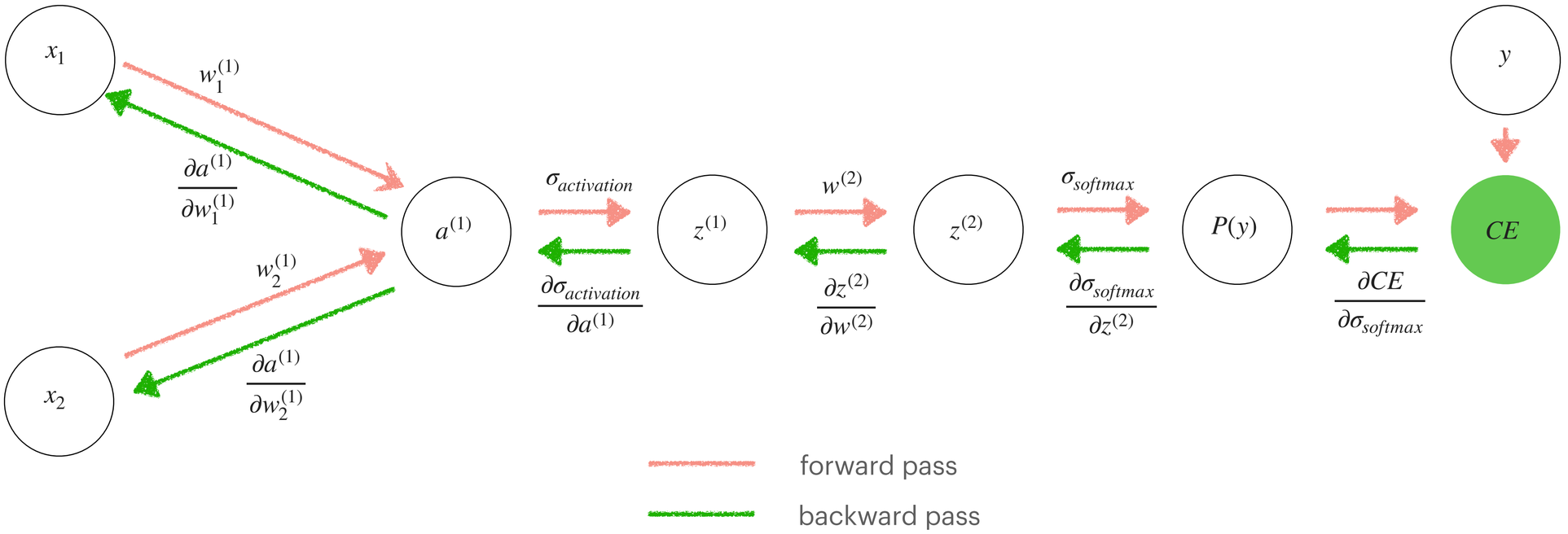

$\frac{\partial \sigma_{activation}}{\partial a^{(1)}}$



forward pass

backward pass

Gradient backpropagation



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

Neural network visualization