2021/01/19

Exercise 3

Consider the neural network in Figure 1 with input $x \in \mathbb{R}$, 3 hidden layers with one node each and one output $y \in \mathbb{R}$.

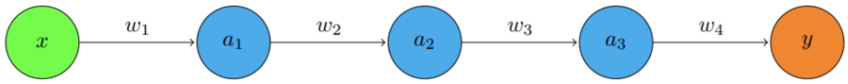
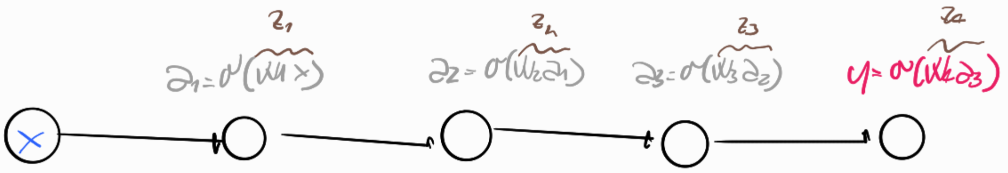


Figure 1: Simple neural network.

In the network each node corresponds to the sigmoid of the previous node multiplied by some weight *i.e.* $a_i = \sigma(w_i a_{i-1})$, $i = 1, \dots, 4$ where $a_0 = x$ and $a_4 = y$.

- By using the chain rule compute $\frac{\partial y}{\partial x}$.
- Compute the maximum of σ' and discuss how this is related to the vanishing gradients problem.



Backpropagation and chain rule

$$\frac{dy}{dx} = \frac{d a_4}{\partial x} = \frac{\partial a_4}{\partial z_4} \cdot \frac{\partial z_4}{\partial a_3} \cdot \frac{\partial a_3}{\partial z_3} \cdot \frac{\partial z_3}{\partial a_2} \cdot \frac{\partial a_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial a_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial x}$$

- $\frac{\partial z_1}{\partial x} = w_1$
- $\frac{\partial a_1}{\partial z_1} = \sigma'(z_1) = \sigma(z_1)(1-\sigma(z_1))$
- $\frac{\partial z_2}{\partial a_1} = w_2$
- $\frac{\partial a_2}{\partial z_2} = \sigma(z_2)(1-\sigma(z_2))$
- $\frac{\partial z_3}{\partial a_2} = w_3$
- $\frac{\partial a_3}{\partial z_3} = \sigma(z_3)(1-\sigma(z_3))$
- $\frac{\partial z_4}{\partial a_3} = w_4$
- $\frac{\partial a_4}{\partial z_4} = \sigma(z_4)(1-\sigma(z_4))$

$$\frac{\partial y}{\partial x} = w_1 w_2 w_3 w_4 \cdot \sigma'(z_1)\,\sigma'(z_2)\,\sigma'(z_3)\,\sigma'(z_4) = \prod_{i=1}^{4} w_i \prod_{i=1}^{4} \sigma'(z_i)$$

Obs: $\max(\sigma'(z)) = 0.25$ when $z = 0.5$

$\hookrightarrow$ gradient $\sim 0.25^L$ when L - depth