Measure error

- Prediction : p → Depends on weights: $w_1, \dots, w_n$ & biases: $b_0, \dots, b_m$
- True value: y (target)
- Loss Function $L(p, y)$ → Tells us how wrong our approximation is

Updating weights and biases (make p similar to y)

Usually involves computing $\frac{\partial L}{\partial w_i}$, for instance

$$w_i \mapsto w_i - \alpha \frac{\partial L}{\partial w_i}, \quad b_j \mapsto b_j - \alpha \frac{\partial L}{\partial w_i}$$

→ Hard to compute since p is a NN

The chain rule

For $\frac{dy}{dx}$. Let $y(u(x))$, then: $\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}$.

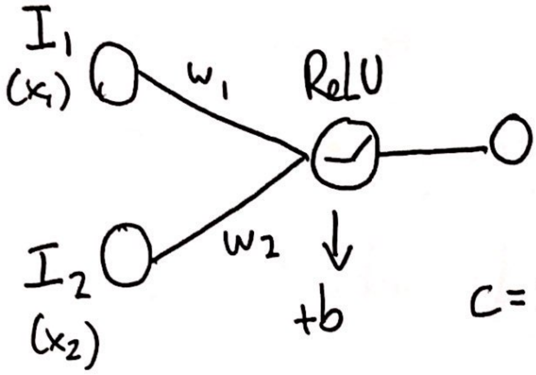
If $y(u_1(x), u_2(x), \dots, u_n(x))$, then:

$$\frac{dy}{dx} = \frac{dy}{du_1} \frac{du_1}{dx} + \frac{dy}{du_2} \cdot \frac{du_2}{dx} + \dots + \frac{dy}{du_n} \cdot \frac{du_n}{dx}$$

Most important formulas

Consider the NN

$$L = \frac{1}{2m} \sum_{i=1}^{N} \underbrace{(p-y)^2}_{C=(p-y)^2}$$



$$c = x_1 w_1 + x^2 w_2 + b = c(w_1, w_2, b)$$

$$p = \mathrm{ReLU}(x_1 w_1 + x_2 w_2 + b) = \mathrm{ReLU}(c)$$

By the chain rule:

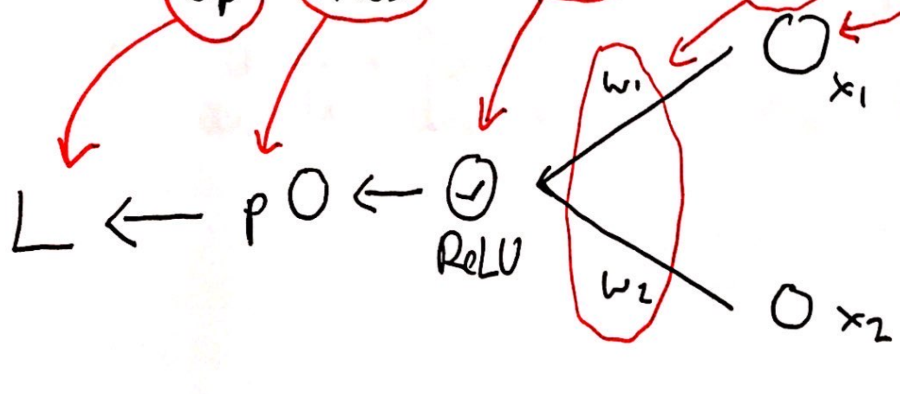
$$\frac{\partial L}{\partial w_1}(p,y) = \frac{\partial L}{\partial p}\frac{\partial p}{\partial w_1} + \frac{\partial L}{\partial y}\underbrace{\frac{\partial y}{\partial w_1}}_{=0}$$

y does not depend on w_1

$$= \frac{\partial L}{\partial p} \cdot \frac{\partial p}{\partial \mathrm{ReLU}} \cdot \frac{\partial \mathrm{ReLU}}{\partial w_1}$$

$$= \frac{\partial L}{\partial p} \cdot \frac{\partial p}{\partial \mathrm{ReLU}} \cdot \frac{\partial \mathrm{ReLU}}{\partial c} \cdot \frac{\partial c}{\partial w_1}$$

$$= \frac{\partial L}{\partial p} \cdot \frac{\partial p}{\partial \mathrm{ReLU}} \cdot \frac{\partial \mathrm{ReLU}}{\partial c} \cdot \frac{\partial c}{\partial w_1} \cdot \left(\frac{\partial w_1}{\partial w_1} + \overbrace{\frac{\partial w_2}{\partial w_1}}^{=0} + \overbrace{\frac{\partial b}{\partial w_1}}^{=0} \right)$$



We are moving backwards!

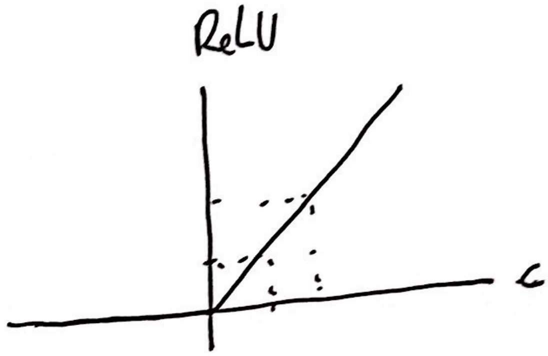
BACKPROPAGATION

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial p} \cdot \frac{\partial p}{\partial ReLU} \cdot \frac{\partial ReLU}{\partial c} \cdot \frac{\partial c}{\partial w_1}$$

$$\frac{\partial L}{\partial p} = \frac{1}{2m} \sum_{i=1}^{N} 2(p-y) = \frac{1}{m} \sum_{i=1}^{N} (p-y)$$

$$\frac{\partial p}{\partial ReLU} = 1 \quad (p = ReLU)$$

$$\frac{\partial ReLU}{\partial c} = \begin{cases} 1, & \text{if } c \geq 0 \\ 0, & \text{if } c < 0 \end{cases}$$



$$\frac{\partial c}{\partial w_1} = \frac{\partial}{\partial w_1} (x_1 w_1 + x_2 w_2 + b) = x_1$$

$$\Rightarrow \frac{\partial L}{\partial w_1} = \frac{1}{m} \sum_{i=1}^{N} (p-y) \cdot (1 \text{ if } c \geq 0,\ 0) \cdot x_1$$

$$\Rightarrow \frac{\partial L}{\partial w_2} = \frac{1}{m} \sum_{i=1}^{N} (p-y) \cdot (1 \text{ if } c \geq 0,\ 0) \cdot x_2$$

$$\Rightarrow \frac{\partial L}{\partial b} = \frac{1}{m} \sum_{i=1}^{N} (p-y) \cdot (1 \text{ if } c \geq 0,\ 0)$$