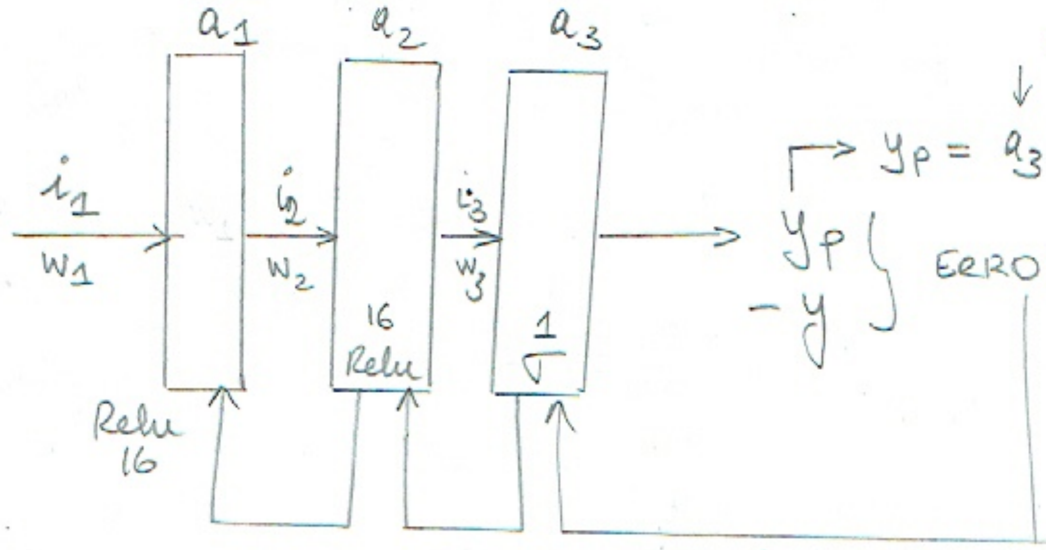
Backpropagation

→ Relu (dot(w_3 i_3) + b)



operator!

↳ $y_p = a_3(i_3 \cdot w_3)$

$\left.\begin{array}{l} y_p \\ -y \end{array}\right\}$ ERROR $= (y_p - y)^2 = C$

feedback signal

Goal of NN → train on training data

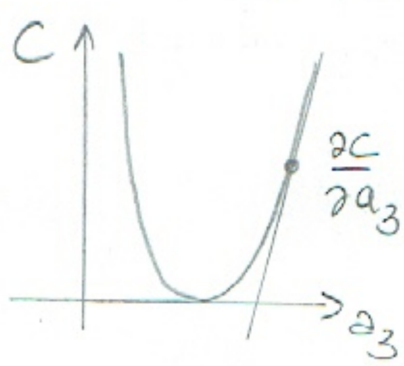
→ as a_i and i_1 and y are fixed, the only way to reduce error is by changing the weights

ERROR $= C = (y_p - y)^2$ and $y_p = a_3(i_3 \cdot w_3) = \sigma(i_3 w_3)$

↳ function of a_3 ↳ $[0,1]$

and a_3 is function of w_3 (1)

→ Nested function.



a_3 and $a_3 = a_3(W_3)$

→ Gradient descent.

(1) $i_3 W_3 = \text{Relu}(\underbrace{i_2 \cdot W_2}_{})$

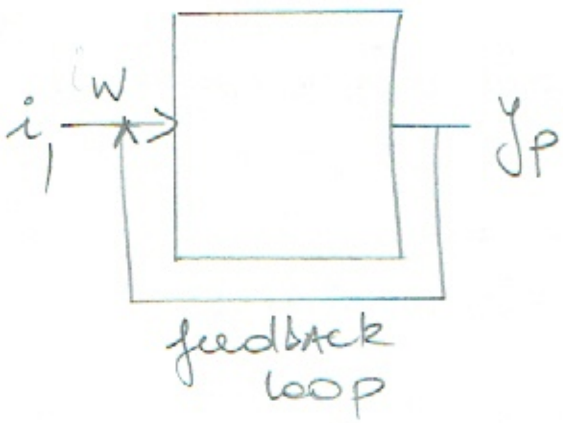
↳ Relu $(i_1 W_1)$

→ $\boxed{y_p = \sigma(i_3 w_3) = \sigma(\text{Relu}(\text{Relu}(i_1 W_1)))}$

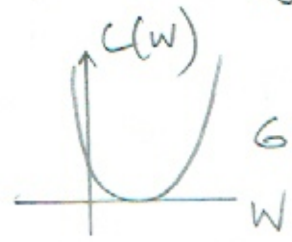
derivative of nested function → chain Rule

Given: $f(g(x))$ → $\frac{\partial f}{\partial x} = \frac{\partial f}{\partial g} \frac{\partial g}{\partial x}$

Shallow Learning

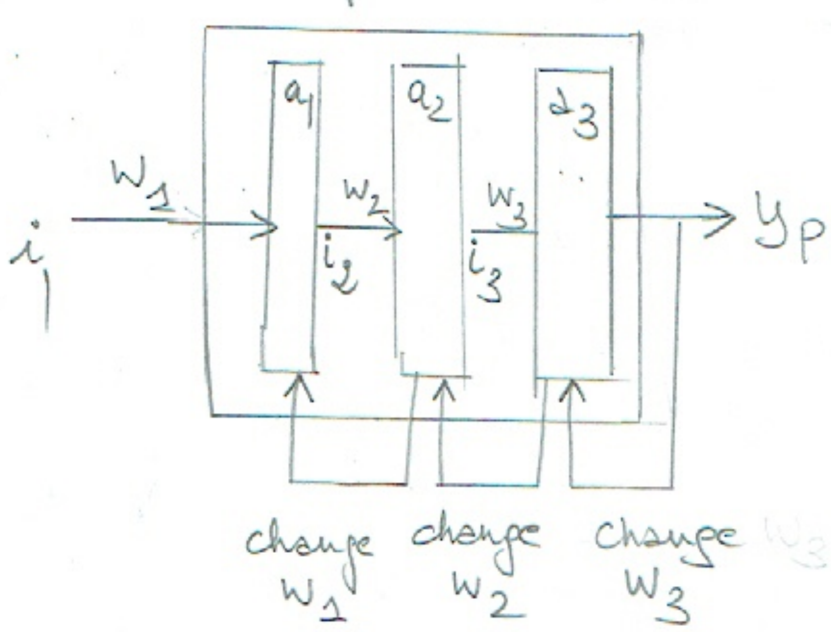


$\rightarrow C = (y_p - y)^2$



Gradient Descent

Deep Learning



function of w_3

↑

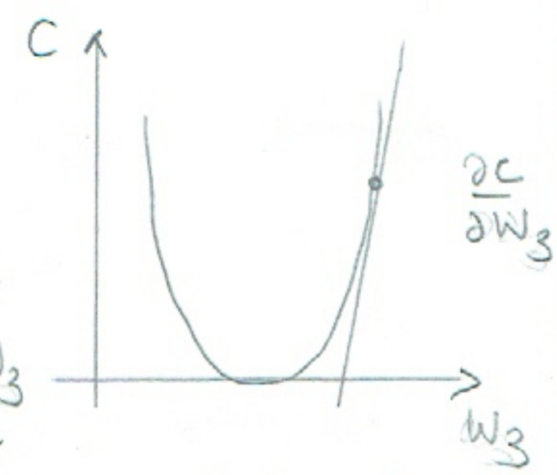
function of a_3

↑

$\rightarrow C = (y_p - y)^2$

↓

Gradient Descent



Update rule $\quad w_3^{t+1} \leftarrow w_3^{t} - \alpha \frac{\partial C}{\partial w_3}$

↳ w_3 not directly accessible

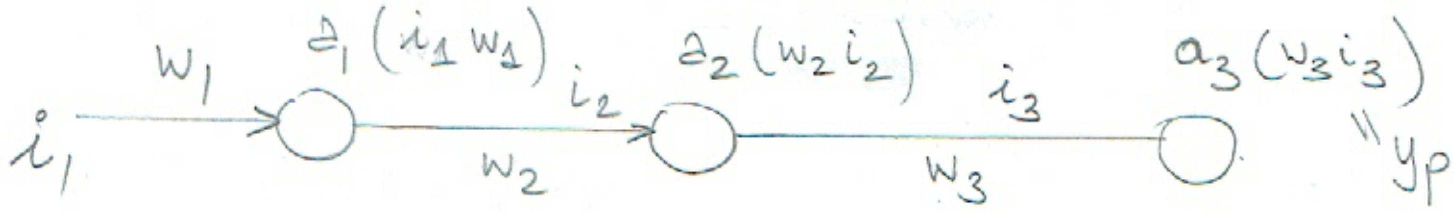
⇓ chain rule

$$\frac{\partial C}{\partial a_3} \frac{\partial a_3}{\partial w_3}$$

what about w_2 and w_1?

$$y_p = a_3(\underbrace{i_3 w_3}_{a_2(i_2 w_2)}) = a_3(a_2(\overbrace{i_2 w_2}^{a_1(i_1 w_1)}))$$

$$a_2(i_2 w_2) \rightarrow a_2 = a_2(w_2)$$



$$-\alpha \frac{\partial C}{\partial a_3} \frac{\partial a_3}{\partial w_3}$$

$$-\alpha \frac{\partial a_2}{\partial w_2} \frac{\partial a_3}{\partial a_2} \frac{\partial C}{\partial a_3}$$

$$-\alpha \frac{\partial a_1}{\partial w_1} \frac{\partial a_2}{\partial a_1} \frac{\partial a_3}{\partial a_2} \frac{\partial C}{\partial a_3}$$

↳ links change in C to changes in w_1,

$$\Rightarrow \boxed{w_3^{t+1} := w_3^{t} - \alpha \left(\frac{\partial a_1}{\partial w_1} \frac{\partial a_2}{\partial a_1} \frac{\partial a_3}{\partial a_2} \frac{\partial C}{\partial a_3} \right)}$$