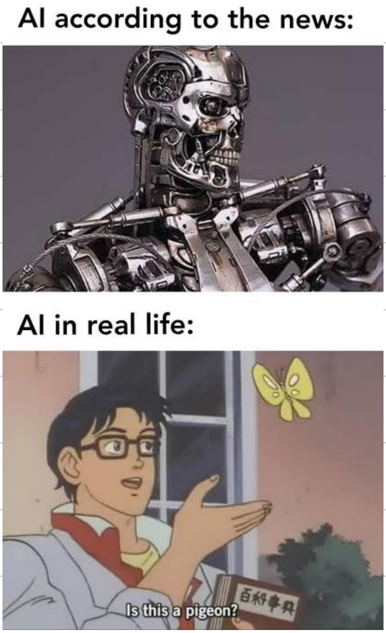
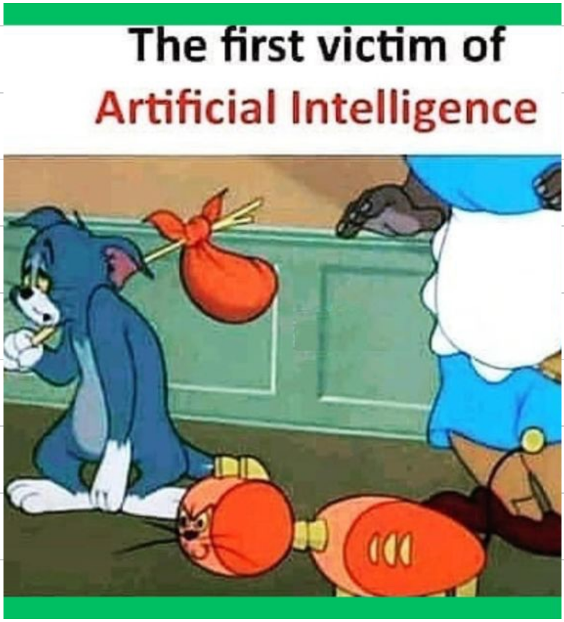
Session-3 # N-Layer Neural Network-1 Feb 09, 2024





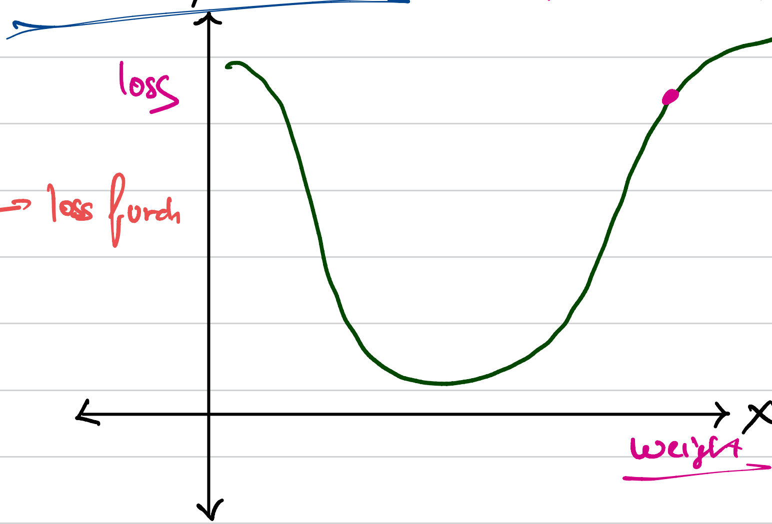
Agenda

Train neural network on Excel Sheet 😈

Lets Talk About Backpropagation

In ML models we have weights → optimise

x_1 — w_1 → High loss ??



L → loss funcn

1 Parameter

$\frac{\partial L}{\partial w}$ ↳ ??

how loss increase w.r.t. w

$\frac{\partial L}{\partial w}$

Formula → Gradient descent

$$w = \underbrace{w - \frac{\partial L}{\partial w}}$$

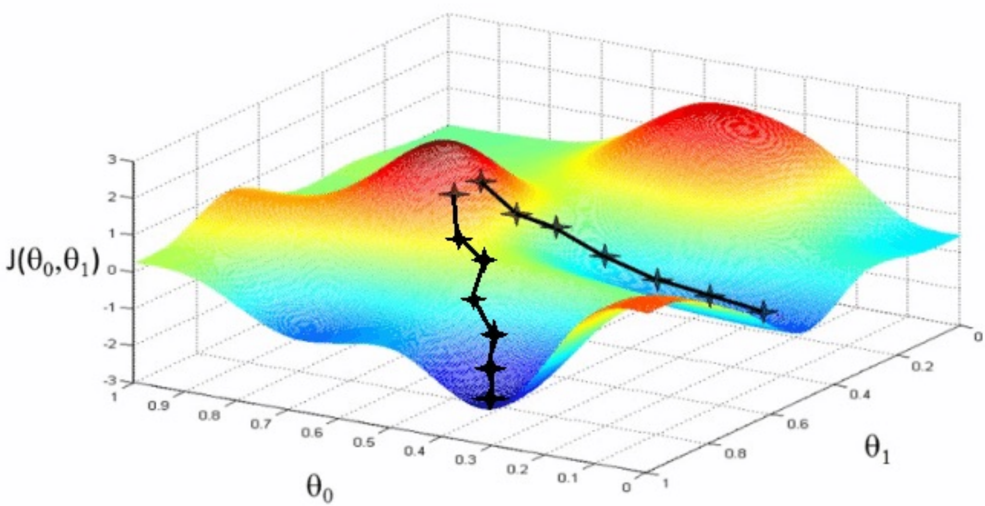
η → Learning rate

0.01

0.000001

$$w = w - \eta \times \frac{\partial L}{\partial w}$$

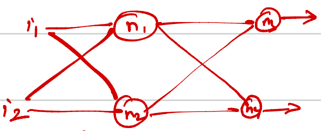
My NN has 2 wt Paramtl



GPT - 3.5 → 135 Billion

GPT - 4 → 1.6 trillion

1 trillion = 1000 Billion



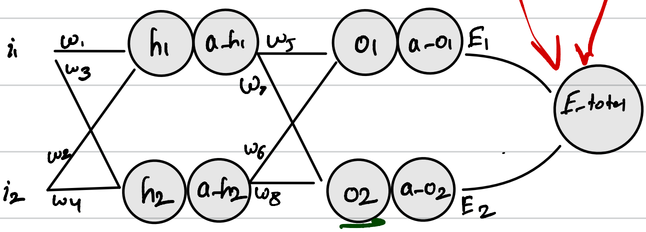
$10^6 \times 10^3 \times 10^3$

total los

not a neuron

10^{12}

LETS BEGIN



starting weights

$w_1 = 0.15$	$w_5 = 0.4$
$w_2 = 0.2$	$w_6 = 0.45$
$w_3 = 0.25$	$w_7 = 0.5$
$w_4 = 0.3$	$w_8 = 0.55$

$E_1 = \frac{1}{2}(t_1 - ao_1)^2$ $\quad E_2 = \frac{1}{2}(t_2 - ao_2$

Assuming only 1 row of target data

$h_1 = w_1 \times i_1 + w_2 \times i_2$

$h_2 = w_3 \times i_1 + w_4 \times i_2$

$ah_1 = \sigma(h_1) = \frac{1}{1+e^{-h_1}}$

$ah_2 = \sigma(h_2)$

$o_1 = w_5 \times a\text{-}h_1 + w_6 \times a\text{-}h_2$

$o_2 = w_7 \times a\text{-}h_1 + w_8 \times a\text{-}h_2$

$a\text{-}o_1 = \sigma(o_1)$

$a\text{-}o_2 = \sigma(o_2)$

$E_1 = \frac{1}{2}(t_1 - a\_o_1)^2$ / $E_2 = \frac{1}{2}(t_2 - a\_o_2)^2$

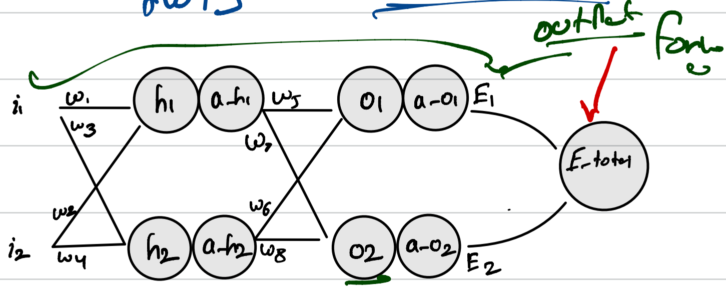
we'w 8 weight $w_1 \rightarrow w_8$

$$w_1 = w_1 - \eta \left[\times \frac{\partial L}{\partial w_1} \right] \rightarrow w_1 \text{ to } w_8$$

I

$$\frac{\partial E\_total}{\partial w_5}$$

$\Downarrow$



Chain rule

$$\frac{\partial (E_1 + E_2)}{\partial w_5} \Rightarrow \frac{\partial E_1}{\partial w_5} = \underbrace{\frac{\partial E_1}{\partial a\_o_1}}_{1} \times \underbrace{\frac{\partial a\_o_1}{\partial o_1}}_{2} \times \underbrace{\frac{\partial o_1}{\partial w_5}}_{3}$$

① $\frac{\partial E_1}{\partial a\_o_1}$; $E_1 = \frac{1}{2}(\overbrace{t_1 - a\_o_1}^{x})^2$

$$\frac{dx^2}{x} = 2x \cdot dx$$

$$= 2 \cdot \left(\frac{1}{2} (t_1 - a\_o_1) \right)(0 - 1)$$

$$= -(t_1 - a\_o_1) \Rightarrow a\_o_1 - t_1$$

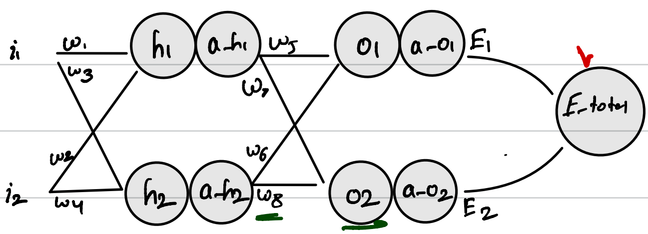
② $\frac{\partial a\_o_1}{\partial o_1}$; $\frac{d\sigma(x)}{dx} = \sigma(x)(1 - \sigma(x))$

$$\frac{\partial a\_o_1}{\partial o_1} = \sigma(o_1) \times (1 - \sigma(o_1))$$

$$\Rightarrow a\_o_1 \times (1 - a\_o_1)$$

(3) $$\frac{\partial o_1}{\partial w_5} = a\_h_1$$

$$\frac{\partial E_1}{\partial w_5} = (a\_o_1 - t_1) \times (a\_o_1 \times (1 - a\_o_1)) \times a\_h_1$$



$$\frac{\partial E_1}{\partial w_6} = (a\_o_1 - t_1) \times (a\_o_1 \times (1 - a\_o_1)) \times a\_h_2$$

$$\frac{\partial E_2}{\partial w_7} = (a\_o_2 - t_2) \times (a\_o_2 \times (1 - a\_o_2)) \times a\_h_1$$

$$\frac{\partial E_2}{\partial w_8} = (a\_o_2 - t_2) \times (a\_o_2 \times (1 - a\_o_2)) \times a\_h_2$$

II

$$\frac{\partial E\_total}{\partial a\_h_1} = \boxed{\frac{\partial E_1}{\partial a\_h_1} + \frac{\partial E_2}{\partial a\_h_1}}$$

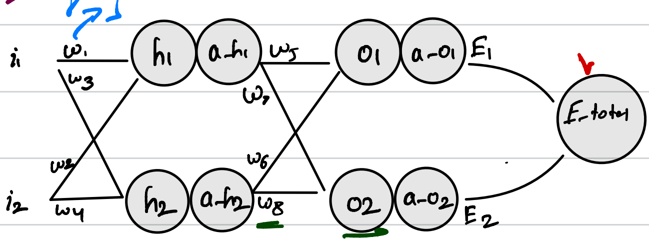
$$\frac{\partial E\_total}{\partial w_1} = \frac{\partial E\_total}{\partial a\_h_1} \times \underbrace{\frac{\partial a\_h_1}{\partial h_1}}_{a\_h_1(1-a\_h_1)} \times \frac{\partial h_1}{\partial w_1} \rightarrow i_1$$

$$\frac{\partial E_1}{\partial a\_h_1} = \frac{\partial E_1}{\partial a\_o_1} \times \frac{\partial a\_o_1}{\partial o_1} \times \frac{\partial o_1}{\partial a\_h_1}$$

$$(a\_o_1 - t_1) \times (a\_o_1 \times (1 - a\_o_1)) \times w_5$$

$$\Rightarrow \frac{\partial E_1}{\partial a\_h_1} = (a\_o_1 - t_1) \times a\_o_1 \times (1 - a\_o_1) \times w_5$$

$$\Rightarrow \frac{\partial E_2}{\partial a\_h_1} = \frac{\partial E_2}{\partial a\_o_2} \times \frac{\partial a\_o_2}{\partial o_2} \times \frac{\partial o_2}{\partial a\_h_1}$$



$$(a\_o_2 - t_2) \times a\_o_2 \times (1 - a\_o_2) \times w_7$$

$$\frac{\partial E\_total}{\partial a\_h_1} = (a\_o_1 - t_1) \times a\_o_1 \times (1 - a\_o_1) \times w_5 + (a\_o_2 - t_2) \times a\_o_2 \times (1 - a\_o_2) \times w_7$$

$$\frac{\partial E\_total}{\partial w_1} = \left\{ (a\_o_1 - t_1) \times a\_o_1 \times (1 - a\_o_1) \times w_5 + (a\_o_2 - t_2) \times a\_o_2 \times (1 - a\_o_2) \times w_7 \right\} \times a\_h_1 (1 - a\_h_1) \times i_1$$

$$\frac{\partial E\_total}{\partial w_2} = \left\{ (a\_o_1 - t_1) \times a\_o_1 \times (1 - a\_o_1) \times w_5 + (a\_o_2 - t_2) \times a\_o_2 \times (1 - a\_o_2) \times w_7 \right\} \times a\_h_1 (1 - a\_h_1) \times i_2$$

$$\frac{\partial E\_total}{\partial w_3} = \left\{ \begin{array}{c} (a\_o_1 - t_1) \times a\_o_1 \times (1 - a\_o_1) \times w_6 \\ + \\ (a\_o_2 - t_2) \times a\_o_2 \times (1 - a\_o_2) \times w_8 \end{array} \right\}$$

$$\times \; a\_h_2 (1 - a\_h_2) \times i_1$$

$$\frac{\partial E\_total}{\partial w_4} = \left\{ \begin{array}{c} (a\_o_1 - t_1) \times a\_o_1 \times (1 - a\_o_1) \times w_6 \\ + \\ (a\_o_2 - t_2) \times a\_o_2 \times (1 - a\_o_2) \times w_8 \end{array} \right\}$$

$$\times \; a\_h_2 (1 - a\_h_2) \times i_2$$

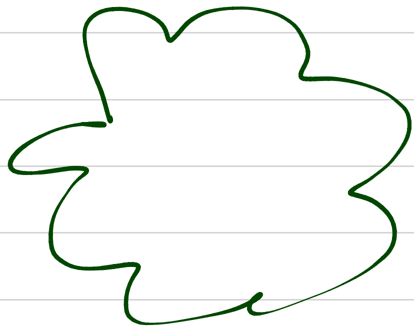
$$\vec{w} = \begin{bmatrix} w_1 & w_2 & w_3 \\ w_4 & w_5 & w_6 \\ w_7 & w_8 & w_9 \end{bmatrix}$$

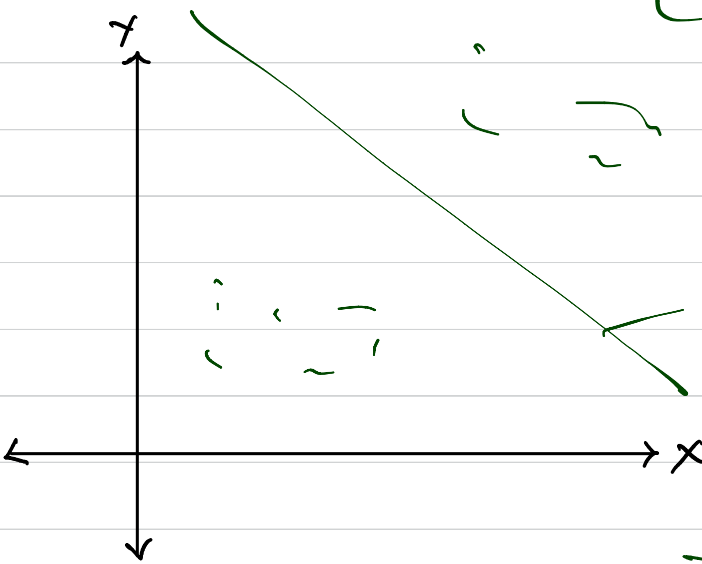
$$\& \quad w_0 = \underline{0.1}$$

$$\vec{w} = \vec{w_6} - \alpha \frac{\partial L}{\partial w \text{ matrix}}$$

10

1000

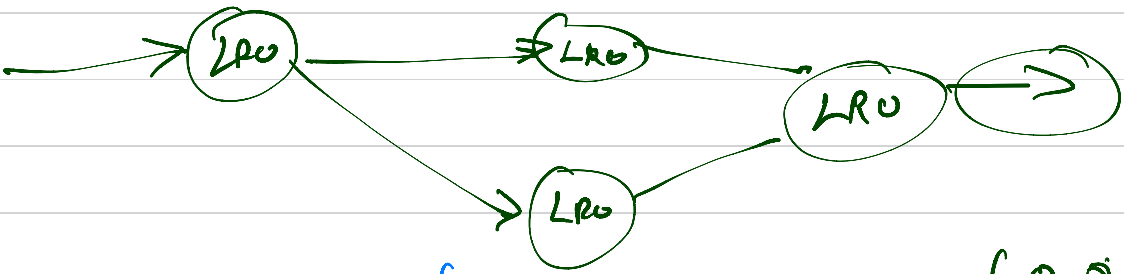


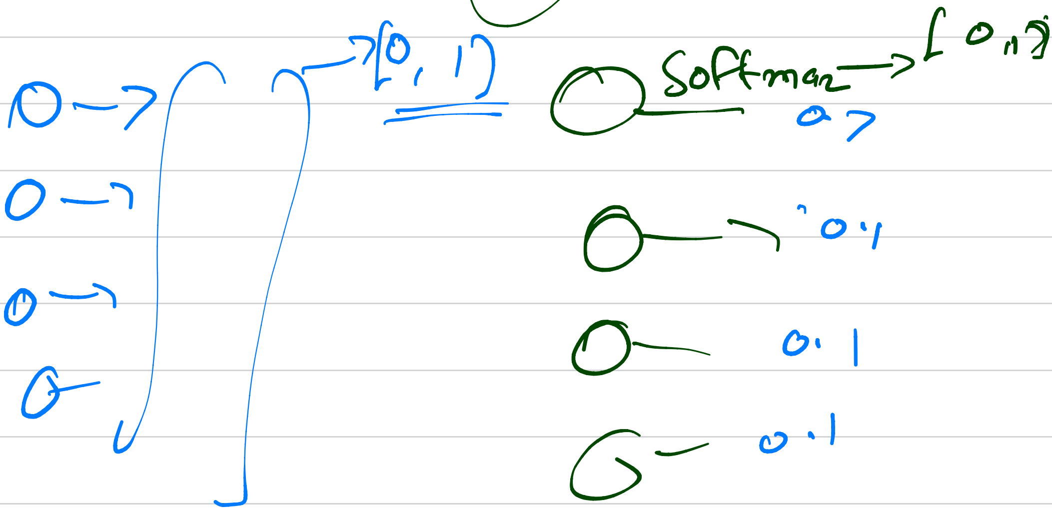


$w_2x_1 + w_1x_2 + w_0$

$w_3x_3 + w_2x_2 + w_1x_1 + w_0 = 0$

$w_1x_1 + w_2x_2 - - - - w_8x_8 + w_0 = 0$





argmax(softmax output)