| | |
|---|---|
| | |
| | |
| | |

## Zero to hero

~~$X = 96 \times 3$~~

$X.shape = 32 \times 3$

examples

block size or char indices

$C[x] = 32 \times 3 \times 3$ ✓

prepare input for NN

Input Layer

$9 \times 100$

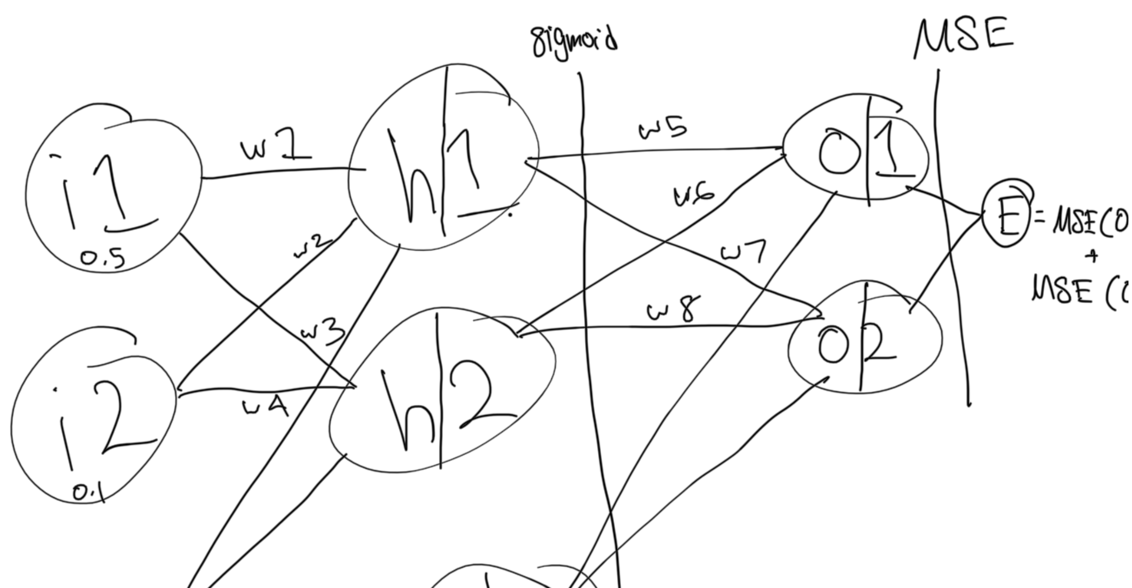
a

OK apparently we need to start a new section, to avoid lag buildup. lol, wtf apple, this is the exact type of product bloat that Jobs tried to fight. But orgs w/ bad incentives push features for promotions instead of love & stability.

OK anyways, back to backprop :)

Repeat process for $w_5 \ldots w_8$

- Let's find out how ~~to~~ changing $w_1$ affects our total error



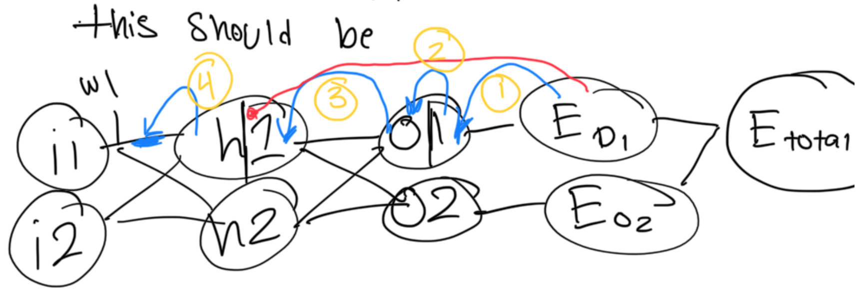
| | |
|---|---|
| + | |
| | |
| | |
| | |

Ok idk what's happening, this notes app is ok for light use, but shitty if you need anything more powerful.

$$\frac{\partial E_{total}}{\partial out_{h1}} = \frac{\partial E_{out_1}}{\partial out_{h_1}} + \frac{\partial E_{out_2}}{\partial out_{h_1}}$$

ok now let's calculate what this should be

$$\boxed{\frac{\partial E_{o_1}}{\partial out_{h_1}}}$$




$out_{h_1}$ $E_{o_1}$

$$= \frac{\partial net_{h_1}}{\partial w_1} \times \frac{\partial net_{o_1}}{\partial out_{h_1}} \times \frac{\partial out_{o_1}}{\partial net_{o_1}} \times \frac{\partial E_{o_1}}{\partial out_{o_1}}$$

| | |
|---|---|
| | |

$$\frac{\partial E_2}{\partial out_{h1}}$$ so


=

$$= \frac{\partial net_{o2}}{\partial out_{h1}} \times \frac{\partial out_{o2}}{\partial net_{o2}} \times \frac{\partial E_2}{\partial out_{o2}}$$

so now putting it all together

$$\frac{\partial E_{total}}{\partial out_{h1}} = \frac{\partial E_1}{\partial out_{h1}} + \frac{\partial E_2}{\partial out_{h1}}$$

$$= \frac{\partial net_{o1}}{\partial out_{h1}} \times \frac{\partial out_{o1}}{\partial net_{o1}} \times \frac{\partial E_1}{\partial out_{o1}}$$

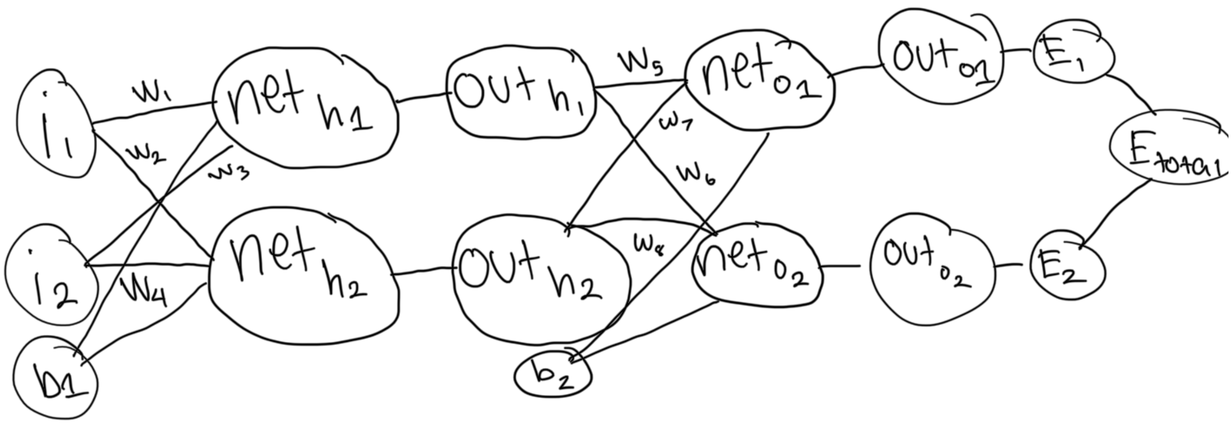
$$+ \frac{\partial net_{o2}}{\partial out_{h1}} \times \frac{\partial out_{o2}}{\partial net_{o2}} \times \frac{\partial E_2}{\partial out_{o2}}$$

Now that we've calculated 3, let's put it all together

$$\frac{\partial E_{total}}{} = \frac{\partial net_{h1}}{\partial w_1} \times \frac{\partial out_{h1}}{\partial net_{h1}} \times \frac{\partial E_{total}}{\partial out_{h1}}$$

| | |
|---|---|
| | |

OK now let's see how $\frac{\partial E_{total}}{\partial w_2}$




$$\frac{\partial E_{total}}{\partial W_2} = \frac{\partial net_{h_2}}{\partial W_2} \times \frac{\partial out_{h_2}}{\partial net_{h_2}} \times \frac{\partial E_{total}}{\partial out_{h_2}}$$

$$= i1 \times (out_{h_2}) \times (1 - out_{h_2}) \times \frac{\partial E_{total}}{\partial out_{h_2}}$$

let's calculate this and see what happens

$$E_{total} = E_1 + E_2$$

$$\frac{\partial E_{total}}{\partial out_{h_2}} = \frac{\partial E_1}{out_{h_2}} + \frac{\partial E_2}{out_{h_2}}$$

$$= \frac{\partial net_{o_1}}{\partial out_{h_2}} \times \frac{\partial out_{o_1}}{\partial net_{o_1}} \times \frac{\partial E_1}{\partial out_{o_1}}$$

$$+$$

$$\frac{\partial net_{o_2}}{\partial out_{h_2}} \times \frac{\partial out_{o_2}}{\partial net_{o_2}} \times \frac{\partial E_2}{\partial out_{o_2}}$$

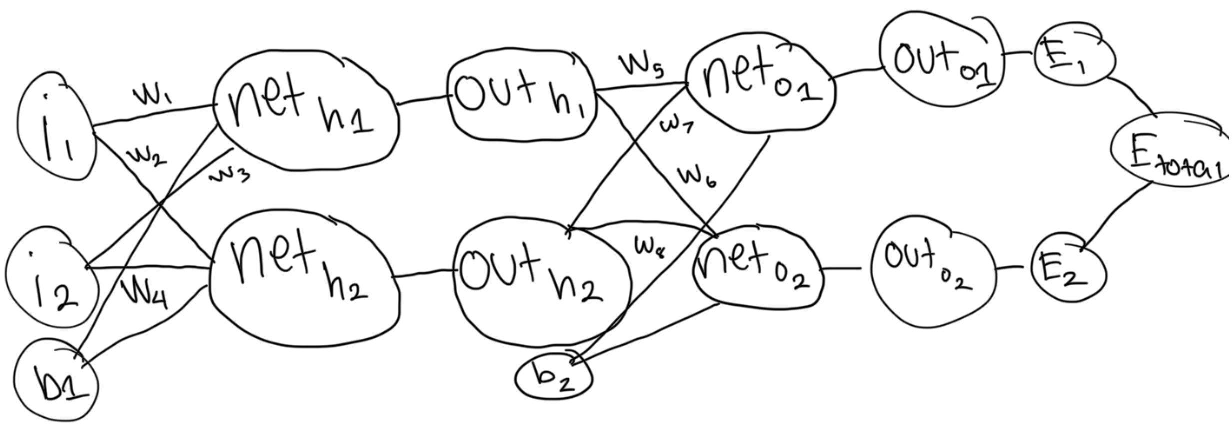
r

| | |
|---|---|
| | |

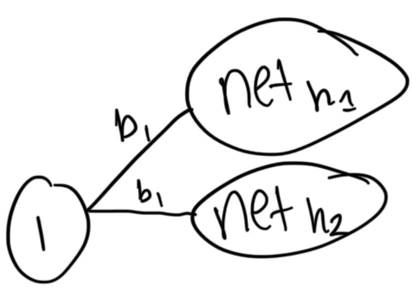
What about our bias parameters $b_1$ and $b_2$?

Let's see how they affect our total error over time.

$$\frac{\partial E_{total}}{\partial b_1}$$




To be more precise...

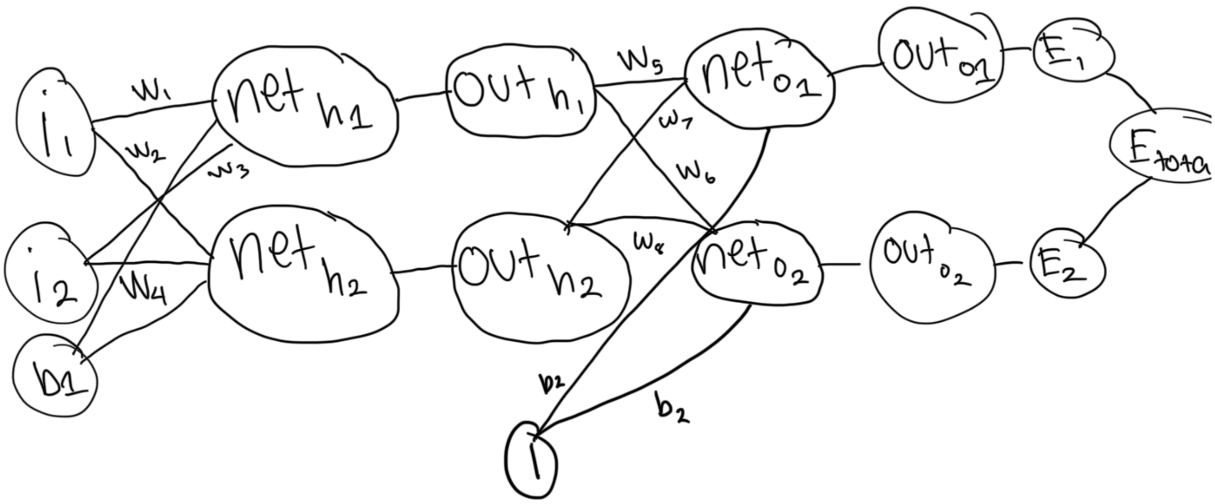
$b_1 =$ 


so $\frac{\partial E_{total}}{\partial b_1}$

If we shake $b_1$, what happens?

r

ok cool now $\frac{\partial E_{total}}{\partial b_2}$




$$\frac{\partial E_{total}}{\partial b_2} = \frac{\partial E_1}{\partial b_2} + \frac{\partial E_2}{\partial b_2}$$

↑
Focus

$$\frac{\partial E_1}{\partial b_2} = \frac{\partial net_{o_1}}{\partial b_2} \times \frac{\partial out_{o_1}}{\partial net_{o_1}} \times \frac{\partial E_1}{\partial out_{o_1}}$$

$$= 1 \times (out_{o_1}) \times (1-out_{o_1}) \times -(target_1 - out_{o_1})$$