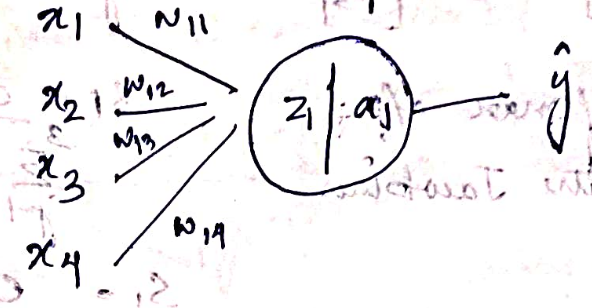
2.C. Saturates in the case of Sigmoid

Lab 2

Layer 1



$$\begin{bmatrix} 0.2 \\ -1.3 \\ 2 \\ 0.01 \end{bmatrix}$$

$$w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + w_{14}x_4$$

$w_{11} = 0.001$
$w_{12} = 0.01$
$w_{13} = -0.005$
$w_{14} = -1.2$

$$z_1 = (0.001 \times 0.2) + (0.01 \times -1.3) + (-0.005 \times 2) + (-1.2 \times 0.01)$$

$$= 0.0002 + (-0.013) + (-0.01) + (-0.012)$$

$$= \cancel{-0.024} \; -0.0348$$

$$a_1 = softmax(z)$$

$$z = -0.024$$

$$\frac{e^{-0.0348}}{e^{-0.0348}} = 1$$

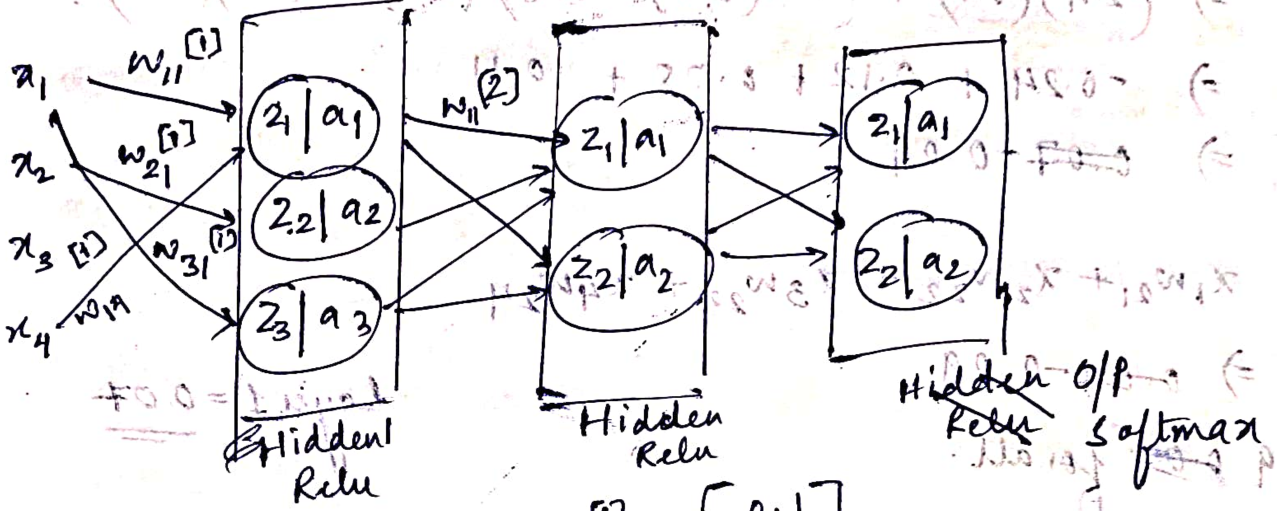
$$\cancel{\frac{e^{0.2}}{e^{-1.3} + e^{0.2} + e^{2} + e^{0.01}}}$$

$$= \frac{1.22}{0.27 + 0.27 + 7.3 + 1.1}$$

$$\begin{bmatrix} 0.2 \\ -1.3 \\ 2 \\ 0.01 \end{bmatrix} \begin{bmatrix} 0.001 \\ 0.01 \\ -0.005 \\ -1.2 \end{bmatrix} = \begin{bmatrix} 0.2 \\ -1.3 \\ 2 \\ 0.01 \end{bmatrix} \begin{bmatrix} 0.001 & 0.01 & -0.005 & -1.2 \end{bmatrix}$$

$$= 0.002 + 0 - 0.013 + 0.01 - 0.012$$

$$= \cancel{-0.024} \; -0.0348$$



$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2.4 \\ 1.2 \\ -0.8 \\ 1.1 \end{bmatrix} \qquad \begin{matrix} w_{11}^{[1]} = \\ w_{21}^{[1]} \end{matrix} \begin{bmatrix} 0.1 \\ 0.1 \\ \vdots \\ 0.1 \end{bmatrix}$$

$$x_1 = \frac{-2.4 + 2.4}{2} = 0$$

$$x_2 = \frac{1.2 + 1.2}{2} = \frac{2.4}{2} = 1.2$$

$$x_3 = \frac{-0.8 + 0.8}{2} = 0$$

$$x_4 = \frac{1.1 + 1.1}{2} = \frac{2.2}{2} = 1.1$$

$$z_1 = w_{11} x_1 + x_1 w_{21} + x_1 w_{31} + x_4 w_{14}$$
$$= (0.1)(-2.4) + (-2.4)(0.1) + (-2.4)(0.1) + (1.1)(0.1)$$

$x_1 w_{11} + x_1 w_{21} + x_1 w_{31} + x_2 w_{12} + x_2 w_{22} + x_3 w_{32}$

$+ x_3 w_{13} + x_3 w_{23} + x_3 w_{33} + x_4 w_{14} + x_4 w_{24}$

$+ x_4 w_{34}$

$\Rightarrow x_1 w_{11} + x_2 w_{12} + x_3 w_{13} + x_4 w_{14}$

$\Rightarrow (-2.4)(0.1) + (1.2)(0.1) + (0.8)(0.1) + (1.1)(0.1)$

$\Rightarrow -0.24 + 0.12 + 0.08 + 0.11$

$\Rightarrow \cancel{0.07}\, -0.09$

$\Rightarrow x_1 w_{21} + x_2 w_{22} + x_3 w_{23} + x_4 w_{24}$

$\Rightarrow \cancel{0.07}\, -0.09$

Layer 1 $= 0.07$

-0.09 ~~0.07~~ for all.

~~a_1 = softmax~~ $= \dfrac{e^{-0.09}}{e^{0.09} + e^{0.09}}$ $\dfrac{e^{-0.09}}{e^{-0.09} + e^{-0.09}}$

$a_1 = \text{Relu}(0.09)$

$= 0$ $\quad \dfrac{(x+|x|)}{2} \rightarrow \dfrac{-0.09+0.09}{2}$

$= \dfrac{1.09}{}$

~~Layer 1~~

$z_1 = 0 - 0.09 \qquad a_1 = 0$

$z_2 = -0.09 \qquad a_2 = 0$

$z_3 = -0.09 \qquad a_3 = 0$

Layer 2

0 for all z_1 and z_2 values

New x values are a_1, a_2 & $a_3 = 0$.

Hideen relu $= 0$.

Layer 3

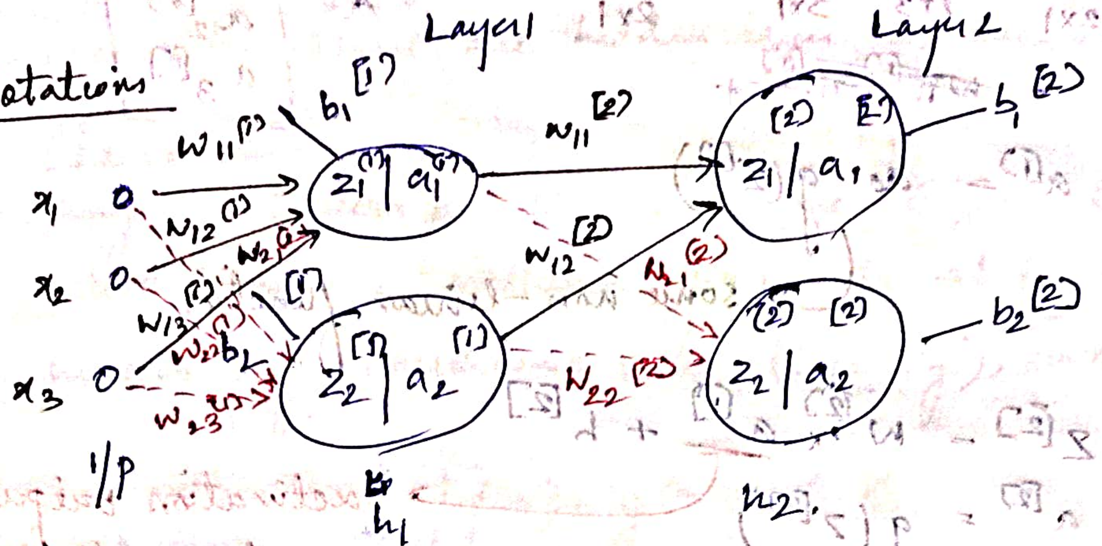
$$a_1 = \frac{e^0}{e^0 + e^0} = \frac{1}{2} = 0.5$$

$$a_2 = \frac{e^0}{e^0 + e^0} = \frac{1}{2} = 0.5$$

doing the softmax on a_1 & a_2

firing rate is non-linear.

Notations



$W^{[l]}$ — connections — matrix — no. rows corresponds to no. of neurons in l layer — columns — no. of neurons in $(l-1)$ layer

b — bias — vector.

z — $w()+b$ — vector.

a — $g(z)$ — vector.

for each layer.

element wise operation

$a^{(0)} = x$

$$W = \begin{bmatrix} w_{11}^{[1]} & w_{12}^{[1]} & w_{13}^{[1]} \\ w_{21}^{[1]} & w_{22}^{[1]} & w_{23}^{[1]} \end{bmatrix} \qquad b^{[1]} = \begin{bmatrix} b_1^{[1]} \\ b_2^{[2]} \end{bmatrix} \qquad z = \begin{bmatrix} z_1^{[1]} \\ z_2^{[1]} \end{bmatrix}$$

$$z_1^{[1]} = w_{11}x_1 + w_{12}x_2 + w_{13}x_3$$
$$= w_{11}^{[1]}a_1^{[0]} + w_{12}^{[1]}a_2^{[0]} + w_{13}^{[1]}a_3^{[0]} + b_1^{[1]}.$$

$$z_2^{[1]} = w_{21}^{[1]}a_1^{[0]} + w_{22}^{[1]}a_2^{[0]} + w_{23}^{[1]}a_3^{[0]} + b_2^{[1]}.$$

$$z_1^{[2]} = w_{11}^{[2]}a_1^{[1]} + w_{12}^{[2]}a_2^{[1]} + b_1^{[2]}$$
$$z_2^{[2]} = w_{21}^{[2]}a_1^{[1]} + w_{22}^{[2]}a_2^{[1]} + b_2^{[2]}.$$

$$w^{[1]} = \begin{bmatrix} w_{11}^{[1]} & w_{12}^{[1]} & w_{13}^{[1]} \\ w_{21}^{[1]} & w_{22}^{[1]} & w_{23}^{[1]} \end{bmatrix}_{2\times 3} \qquad w^{[2]} = \begin{bmatrix} w_{11}^{[2]} & w_{12}^{[2]} \\ w_{21}^{[2]} & w_{22}^{[2]} \end{bmatrix}$$

$$z^{[1]} = w^{[1]} a^{[0]} + b^{[1]}$$

$$z^{[1]} = w^{[1]} a^{[0]} + b^{[1]}$$
$2\times1 \quad 2\times3 \quad 3\times1 \quad 2\times1$

$$a = \begin{bmatrix} a_1^{[0]} \\ a_2^{[0]} \\ a_3^{[0]} \end{bmatrix}$$

$\cancel{= w_{11}^{[1]} a^{[0]} +}$

$$a^{[1]} = g(z^{[1]})$$

some non-linear function

$$z^{[2]} = w^{[2]} \times a^{[1]} + b^{[2]}$$
$$a^{[2]} = g(z^{[2]})$$

activation output from the previous layer.

$$z^{[3]} = w^{[3]} \times a^{[2]} + b^{[3]}$$
$$a^{[3]} = g(z^{[3]})$$

Ask sir: diff. a energy for diff. layer.

$g(z) = z \longrightarrow$ Identity fn.

$$\begin{aligned}
a^{[3]} &= g(z^{[3]}) \\
&= z^{[3]} \\
&= w^{[3]} \times a^{[2]} + b^{[3]} \\
&= w^{[3]} g(z^{[2]}) + b^{[3]} \\
&= w^{[3]} z^{[2]} + b^{[3]} \\
&= w^{[3]} w^{[2]} a^{[1]} + b^{[2]} + b^{[3]} \\
&= w^{[3]} w^{[2]} g(z^{[1]}) + b^{[2]} + b^{[3]} \\
&= w^{[3]} w^{[2]} z^{[1]} + b^{[2]} + b^{[3]} \\
&= w^{[3]} w^{[2]} w^{[1]} a^{[0]} + b^{[1]} + b^{[2]} + b^{[3]} \\
&= \boxed{\tilde{W} x + \tilde{b}}
\end{aligned}$$

$\rightarrow$ linear regression

Representation power of NN

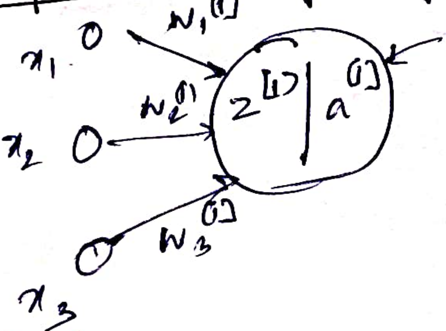
- NN with atleast one hidden layer ~~are~~ are universal **approximators**
- any continuous fn, $f(x)$, and some $\epsilon > 0$, there exists a neural network, $g(x)$, with at least one hidden layer (with some non-linearity), such that

$$\forall x, |f(x) - g(x)| \leq \epsilon$$

both functions mimic hence difference is zero.

Neural Network are non-convex

Perceptron - single layer NN.



$$a^{[1]} = g(z^{[1]})$$

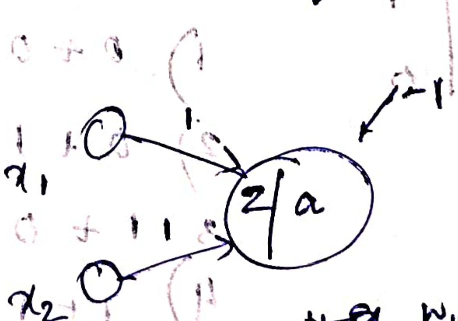
$$z^{[1]} = w_{11}^{[1]} x_1 + w_{12}^{[1]} x_2 + w_{13}^{[1]} x_3 + b_1^{[1]}$$

$$a^{[1]} = g(z^{[1]})$$

$$g = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z \leq 0 \end{cases}$$

AND

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1



$w_{11} = 1$

$w_{12} = 1$

$w_{11} x_1 + w_{12} x + (-1)$

$0 + 0 - 1 = -1$

$= 0 \quad \cancel{g(-1) = e^{-2} + e^{-2}}$

for $x_1=0$ $x_2=1$

$\cancel{w_1 x}$ $0+1-1=0$ $g(z)$

for $x_1=1, x_2=0$ if bias is θ = 1 so 1

$= 1+0-1=0$ $g(z)$

if bias is 0 ⇒ $1+0+0$
$= 1$ $g(z) = 1$

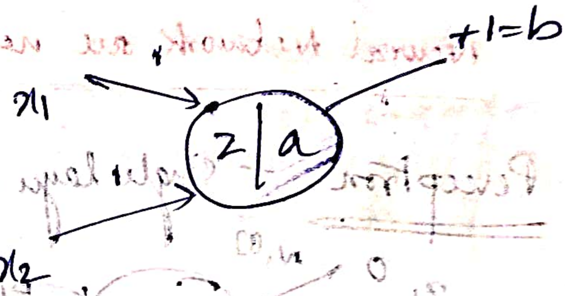
for $x_1=1, x_2=1$

⇒ $1+1-1$ ⇒ $2-1$

so $g(z)=1$.

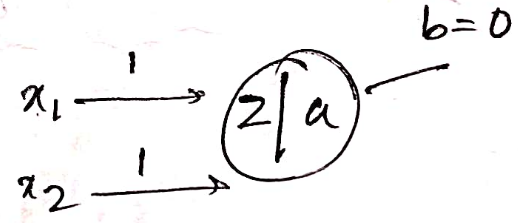
~~AND~~ OR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1



XOR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

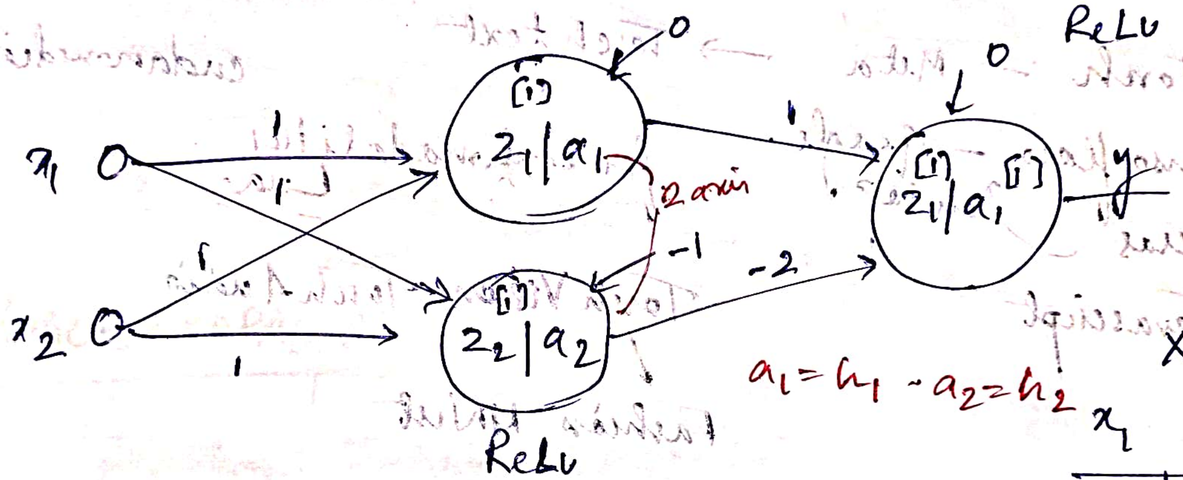


1) $0+0+0=0$ $g(z)=0$

2) $0+1+0=1$ $g(z)=1$

3) $1+0+0=1$ $g(z)=1$

4) $1+1+0=2$ $g(z)=1$



XOR

	x_1	x_2	y
①	0	0	0
②	0	1	1
③	1	0	1
④	1	1	0

$z_1 = w_{11}x_1 + w_{12}x_2 + b_1 = \cancel{1\times1+1}\ 0+0+0=0$

$= \cancel{0}$ $\quad g(z)=0$

② $\Rightarrow 0+1+0=1 \quad g(z)=1$

③ $\Rightarrow 1+0+0=1 \quad g(z)=1$

④ $\Rightarrow 1+1+0=2 \quad g(z)=1 \quad a_1$

total $z = 4$. $\quad g(z)=1$, $a_1=1$

$z_2 = 0+0-1=-1 \quad g(z)=0$

$= 0+1-1=0 \quad g(z)=0$

$= 1+0-1=0 \quad g(z)=0$

$= 1+1-1=1 \quad g(z)=1$

total $z = 1-1=0 \quad g(z)=0$, $a_2=0$

Computing the equation through several layers has is computational cost !!

Random initialization is done as 0 will for wts → lead to zero in the equation.

Computational Graph

Wt initialization

i) $W_{ij} \sim$ uniform $([-0.1, .1])$

→ Uniform distribution

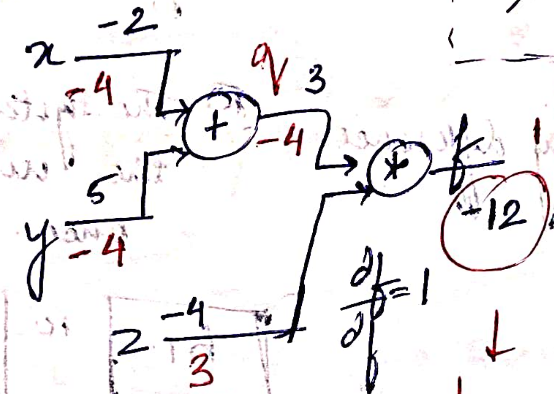
ii) $W_{ij} \sim N\left(0, \sqrt{\frac{2}{n^{[l]} + n^{[l-1]}}}\right)$

→ Xavier init

faster convergence

Let $f = (x+y)z$

multivariable func



$\frac{\partial f}{\partial f} = 1$

forward pass output

x has +ve impact on q. but q has negative impact on f.

Let $q = x + y$

$f = qz$ $\quad \frac{\partial f}{\partial q} = z$

$$\left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right] ?$$

$\frac{\partial q}{\partial x} = 1$ $\quad \frac{\partial q}{\partial y} = 1$ $\quad \frac{\partial f}{\partial z} = q$

$\Rightarrow \frac{\partial f}{\partial x} = \boxed{\frac{\partial f}{\partial q}} \cdot \boxed{\frac{\partial q}{\partial x}}$

$= z \cdot 1$

$= z \Rightarrow -4$

global gradient.

local gradient

$\Rightarrow \frac{\partial f}{\partial y} = \frac{\partial f}{\partial q} \cdot \frac{\partial q}{\partial y}$

$= z \cdot 1$

$= z \Rightarrow -4$

output q respect to x – local gradient

then global gradient → value that came to me

- start from back
- put the gradients
- compute the local & global gradients
- multiply them.

(*) $\frac{\partial f}{\partial z}$ - local grad × global grad

$\frac{\partial f}{\partial y} = \ldots$

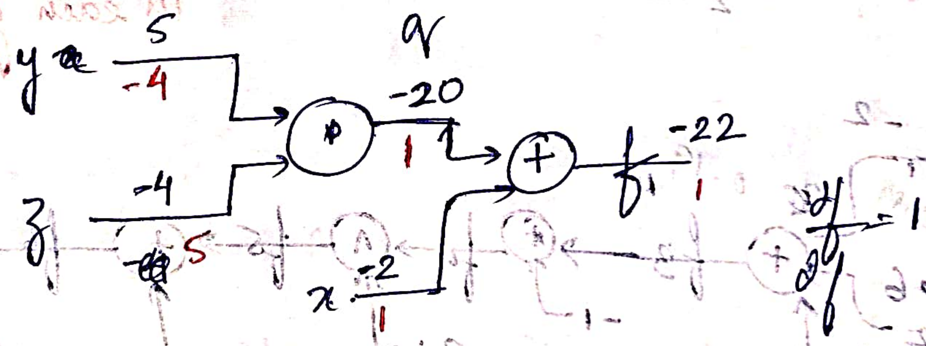
29/07/25 $f(x, y, z) = x + yz$

$x = -2,\ y = 5,\ z = -4$

1) Identify additional functions & Compute local derivatives.
2) Draw computational graph
3) Perform forward pass
4) Perform backward pass starting from the end of the circuit.

$f = x + yz$

let



let $q = yz$ $\quad f = q + x \longrightarrow \frac{\partial f}{\partial q} = \cancel{2}1 \Rightarrow \cancel{-20}$ $\quad \frac{\partial f}{\partial x} = \cancel{q}\,1 \Rightarrow \cancel{-20}$

$\frac{\partial q}{\partial y} = z \quad \frac{\partial q}{\partial z} = y$

$\Rightarrow -4 \quad \Rightarrow 5$

$\cancel{\frac{\partial q}{\partial x} = 0}$

(*) $\boxed{\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}}$

$\Rightarrow \cancel{\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \cdot \frac{\partial q}{\partial x} \Rightarrow 0}$ $\quad 0$

$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial q} \cdot \frac{\partial q}{\partial y} \Rightarrow (\cancel{2}\,1)(-4) = \cancel{8}\,-4$

$\frac{\partial f}{\partial z} = \frac{\partial f}{\partial q} \cdot \frac{\partial q}{\partial z} \Rightarrow (\cancel{2}\,1)(5) = \cancel{20}\,5$

$$f(w,x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$

$w_0 = 2$

$x_0 = -1$

$w_1 = 3$

$x_1 = -2$

$w_2 = -3$

~~let $q = w_0 x_0 + w_1 x_1 + w_2$~~

~~$f = \frac{1}{1+e^{-q}}$~~

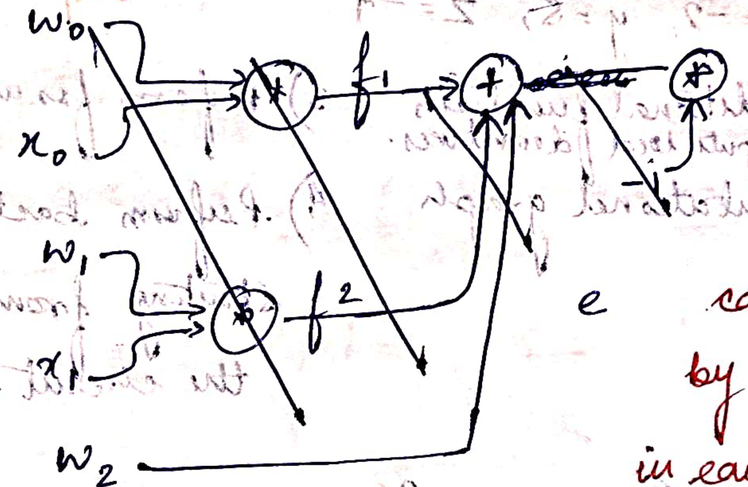
let $f_1 = w_0 x_0$, $f_2 = w_1 x_1$, $f_3 = \cancel{\frac{1}{1+e^{-q}}}\ f_1 + f_2 + w_2$

$f_4 = -f_3$

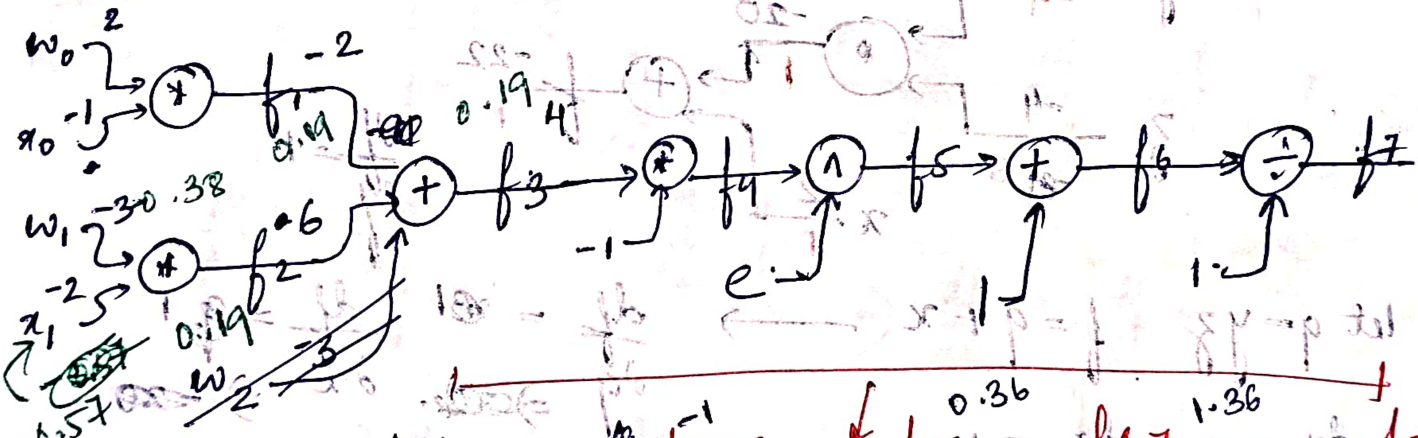
$f_5 = e^{f_4}$

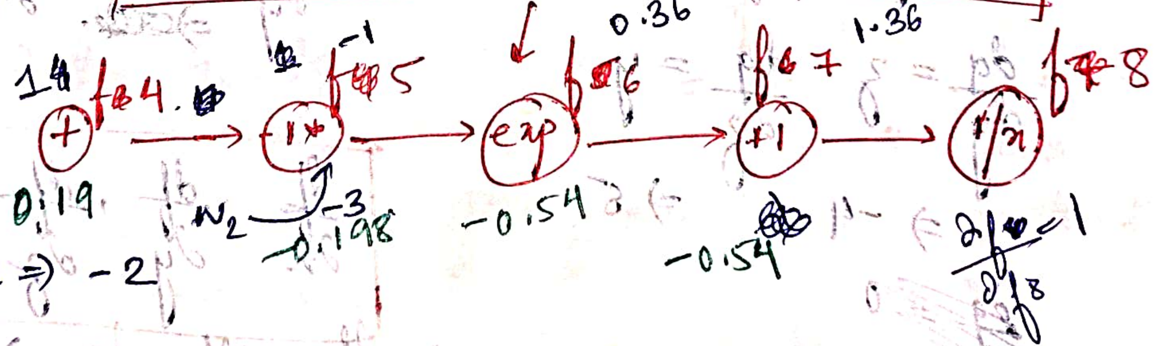
$f_6 = 1 + f_5$

$f_7 = \frac{1}{f_6}$



can be done by 2 inputs in each gate





$f_1 = 2 \times -1 \Rightarrow -2$

$f_2 = 6$

$f_3 = -2 + (6) + -3 \Rightarrow -2 + 6 - 3 \Rightarrow -2 - 9 \Rightarrow -11$

$f_3 = 4$ $\quad f_4 = w_2 + f_3 = -3 + 4 \Rightarrow 1$ $\quad f_5 = -1$

$f_6 = e^{-1} = 0.36$ $\quad f_7 = 1 + 0.36 = 1.36$ $\quad f_8 = \frac{1}{1.36 (f_7)} = 0.735$

$f_1 = w_0 x_0$ $\quad f_2 = w_1 x_1$ $\quad f_3 = f_1 + f_2$ $\quad f_4 = f_3 + w_2$

$f_5 = -f_4$ $\quad f_6 = e^{f_5}$ $\quad f_7 = 1 + f_6$ $\quad f_8 = \frac{1}{f_7}$

1 $\quad f_1 \Rightarrow \underbrace{\frac{\partial f_1}{\partial x_0} \Rightarrow w_0 \Rightarrow 2 \quad \frac{\partial f_1}{\partial w_0} = x_0 \Rightarrow -1}$

2 $\quad f_2 \Rightarrow \left.\begin{array}{l} \frac{\partial f_2}{\partial w_1} = x_1 \rightarrow -2 \\ \frac{\partial f_2}{\partial x_1} = w_1 \rightarrow -3 \end{array}\right\}$

3 $\quad \left.\begin{array}{l} f_3 = \frac{\partial f_3}{\partial f_1} = 1 \\ \frac{\partial f_3}{\partial f_2} = 1 \end{array}\right\}$

4 $\quad \left.\begin{array}{l} f_4 = \frac{\partial f_4}{\partial f_3} = 1 \\ f \Rightarrow \frac{\partial f_4}{\partial w_2} = 1 \end{array}\right\}$

5 $\quad f_5 \Rightarrow \frac{\partial f_5}{\partial f_4} = -1$

6 $\quad f_6 = \frac{\partial f_6}{\partial e^{f_5}} \Rightarrow e^{f_5} \Rightarrow e^{-1} \Rightarrow 0.36$

7 $\quad f_7 = \frac{\partial f_7}{\partial f_6} = 1$

8 $\quad f_8 = \cancel{\frac{\partial f_8}{\partial (\frac{1}{f_7})}} \Rightarrow \cancel{\frac{\partial f_8}{\partial f_7}} \Rightarrow \frac{1}{f_7} \Rightarrow \frac{\partial f_8}{\partial f_7} = \cancel{f_7^2} \quad \frac{-1}{f_7^2}$

local ones

$\frac{\partial f_7}{\partial f_8} = \text{local} \times \text{global}$

$= \frac{\partial f}{\partial f_8} \times \frac{\partial f_7}{\partial f}$

Global

$\frac{\partial f_8}{\partial f_7} = \text{local} \times \text{global}$

$= \frac{-1}{f_7^2} \times 1 \Rightarrow \frac{-1}{(1.36)^2} \times 1 \Rightarrow -0.54$

$$\frac{\partial f_8}{\partial f_6} = \frac{\partial f_7}{\partial f_6} \times \frac{\partial f_8}{\partial f_7}$$

$= 1 \times -0.54$

$= -0.54$

$$\frac{\partial f_8}{\partial f_5} = \frac{\partial f_8}{\partial f_6} \times \frac{\partial f_6}{\partial f_5}$$

$= -0.54 \times 0.36$

$= -0.198$

$$\frac{\partial f_8}{\partial f_4} = \text{local} \times \text{global}$$

$$\Rightarrow \frac{\partial f_5}{\partial f_4} \times \frac{\partial f_8}{\partial f_5}$$

$\Rightarrow -1 \times -0.198$

$\Rightarrow 0.19$

$$\cancel{\frac{\partial f_8}{\partial f_3} = \frac{\partial f_4}{\partial f_3} \times \frac{\partial f_8}{\partial f_4}}$$

$\cancel{= 1 \times 0.19}$

$= 0.19$

$$\cancel{\frac{\partial f_8}{\partial f_2} = \frac{\partial f_3}{\partial f_2} \times \frac{\partial f_8}{\partial f_3}}$$

$\cancel{1 \times 0.19}$

$\Rightarrow 0.19$

$$\frac{\partial f_8}{\partial f_2} = \frac{\partial f_3}{\partial f_2} \times \frac{\partial f_8}{\partial f_3}$$

$\Rightarrow 0.19$

$$\cancel{\frac{\partial f_8}{\partial f_1} = \frac{\partial f_2}{\partial f_1} \times \frac{\partial f_8}{\partial f_2}}$$

$=$

$$\cancel{\frac{\partial f_4}{\partial f_4} = \frac{\partial f_5}{\partial f_4} \times \frac{\partial f_8}{\partial f_5}}$$

$= 0.19$

$$\frac{\partial f_8}{\partial f_3} = \frac{\partial f_4}{\partial f_3} \times \frac{\partial f_8}{\partial f_4}$$

$= 0.19$

$$\frac{\partial f_8}{\partial w_2} = \cancel{\frac{\partial f_2}{\partial w_2} \times \frac{\partial f_8}{\partial f_2}} \quad \frac{\partial f_4}{\partial w_2} \times \frac{\partial f_8}{\partial f_4}$$

$= \cancel{2} \times$

$\Rightarrow 1 \times 0.19$

$= 0.19$

$$\frac{\partial f_8}{\partial w_1} = \frac{\partial f_2}{\partial w_1} \frac{\partial f_8}{\partial f_2}$$

$= -2 \times 0.19$

$= -0.38$

$$\frac{\partial f_8}{\partial x_1} = \frac{\partial f_2}{\partial x_1} \times \frac{\partial f_8}{\partial f_2}$$

$$= w_1 \times 0.19$$

$$= -3 \times 0.19$$

$$= -0.57$$

$$\cancel{\frac{\partial f_8}{\partial w_0}} =$$

$$\frac{\partial f_8}{\partial f_1} = \frac{\partial f_3}{\partial f_1} \times \frac{\partial f_8}{\partial f_3}$$

$$= 1 \times 0.19$$

$$= 0.19$$

$$\frac{\partial f_8}{\partial w_0} = \frac{\partial f_1}{\partial w_0} \cdot \frac{\partial f_8}{\partial f_1}$$

$$= -1 \times 0.19$$

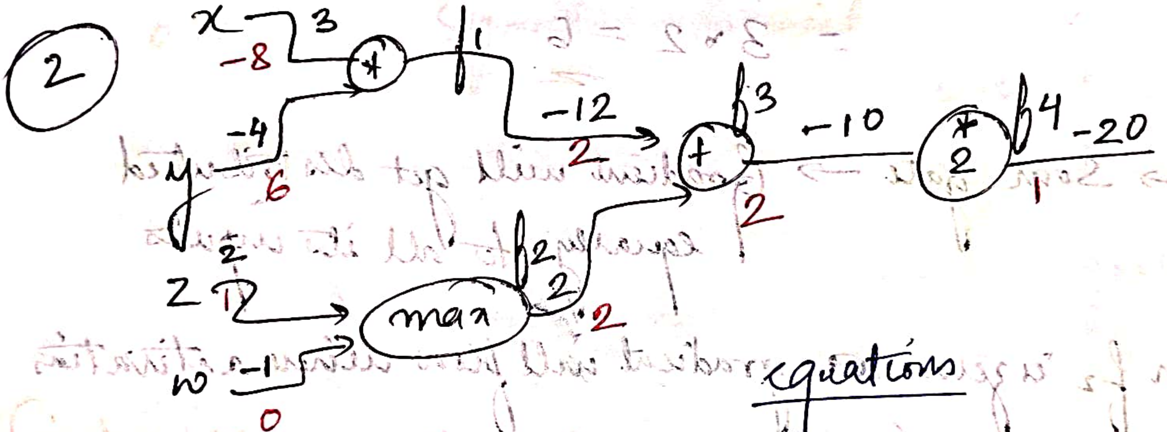
$$= \cancel{0.38} \quad -0.19$$

$$\frac{\partial f_8}{\partial x_0} = \frac{\partial f_1}{\partial x_0} \cdot \frac{\partial f_8}{\partial f_1}$$

$$= 2 \times 0.19$$

$$= 0.38$$

2



equations

$$f_1 = xy$$

$$\frac{\partial f_1}{\partial x} = y = -4 \qquad \frac{\partial f_1}{\partial y} = x \Rightarrow 3$$

$$f_2 = \max(z, w) = 1$$

$$f_3 = f_1 + f_2 \rightarrow \frac{\partial f_3}{\partial f_1} = 1 \qquad \frac{\partial f_3}{\partial f_2} = 1$$

$$f_4 = 2 f_3$$

$$\frac{\partial f_4}{\partial f_3} = 2$$

$$\frac{\partial f_2}{\partial z} = 1 \text{ if } z > w$$
$$0 \text{ if } z < w$$
undefined $\cdot\ z = w$

$$\frac{\partial f_2}{\partial w} = 1 \text{ if } w > z$$
$$0 \text{ if } w < z$$
undefined $\cdot\ w = 2$

$$\frac{\partial f_4}{\partial f_3} = \frac{\partial f_4}{\partial f_3} \times 1$$

$$= 2$$

$$\frac{\partial f_4}{\partial f_2} = \frac{\partial f_3}{\partial f_2} \cdot \frac{\partial f_4}{\partial f_3}$$

$$= 1 \times 2$$

$$= 2$$

$$\frac{\partial f_4}{\partial f_1} = \frac{\partial f_3}{\partial f_1} \cdot \frac{\partial f_4}{\partial f_3}$$

$$= 1 \times 2$$

$$\frac{\partial f_4}{\partial x} = \frac{\partial f_1}{\partial x} \cdot \frac{\partial f_4}{\partial f_1}$$

$$= -4 \cdot 2$$

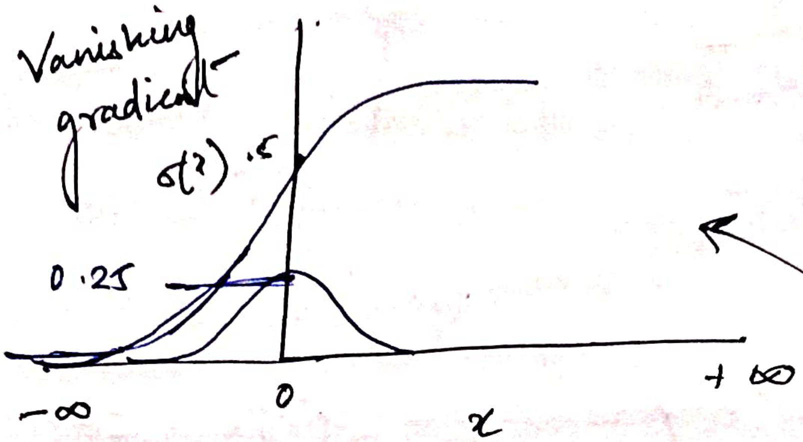
$$= -8$$

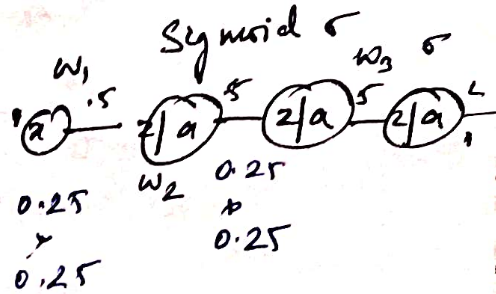
$$\frac{\partial f_4}{\partial y} = \frac{\partial f_1}{\partial y} \frac{\partial f_4}{\partial f_1}$$

$$= 3 \times 2 = 6$$

Observation:

Vanishing and exploding Gradient Problems





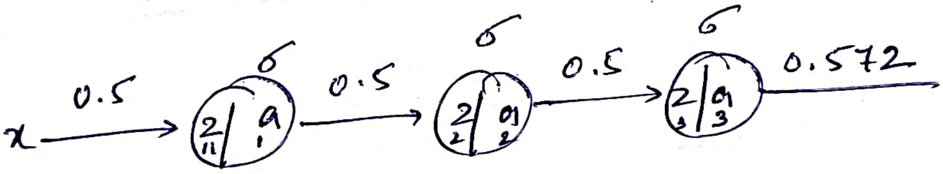
gradient is vanished as more decimal number will get multiplied and no learning will be there.

No updates as gradients will be non-existent

ReLU ⟶ Exploding gradient problem.

can happen potentially to ReLU

Relu can be solved by → Clipping the gradients → threshold.



$z_1 = 1 \times 0.5 = 0.5$

$a_1 = \sigma(z_1) = 0.622$ $\qquad \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-0.5}} = \frac{1}{1+0.606} = \frac{1}{1.606}$

$L = \frac{1}{2}(a_3 - y)^2$, $\cancel{\frac{dL}{da_3} = a_3 - y}$

$= 0.572$

$$\frac{dL}{da_3} = \frac{1}{2} \times 2 (a_3 - y)^1 = a_3 - y$$

~~$a_1 = 0.622$~~

~~$z_2 = 0.5 \times 0.622$~~

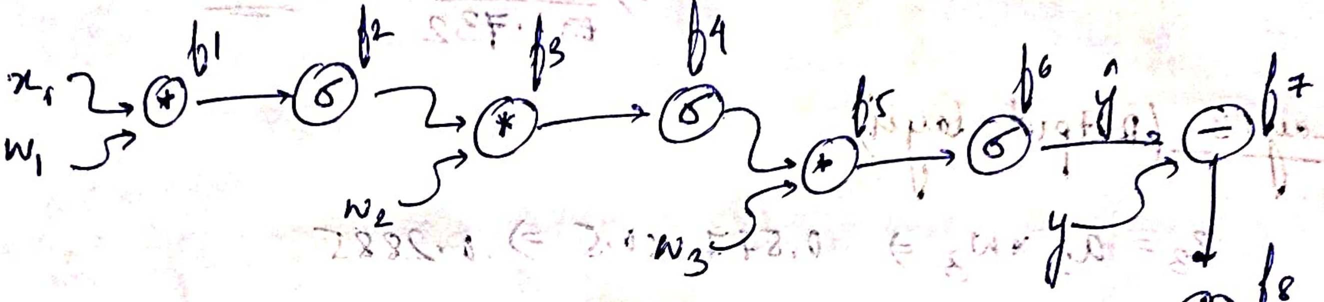
~~$= 0.311$~~

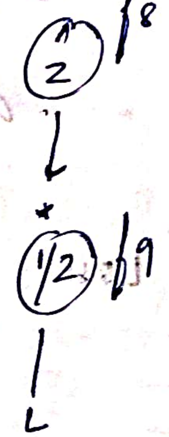
$$\frac{\partial L}{\partial a_3} = -0.428$$

$a_3 - y = 0.572 - 1$

$= -0.428$

$$\frac{dL}{da_3} = \cancel{0.572}\ 0.$$





$f1 = x_1 w_1$

$= 1 \times 0.5$

$= 0.5$

$f2 = \sigma(0.5)$

$$= \frac{1}{1+e^{-2}} = \frac{1}{1+0.606}$$

$$= \frac{1}{1.606}$$

$\Rightarrow 0.622$

$f3 = 0.622 \times 0.5(w_2)$

$= 0.311$

$f4 = \sigma(0.311)$

$$= \frac{1}{1+e^{-0.311}} = \frac{1}{1+0.732} \Rightarrow \frac{1}{1.732} \Rightarrow 0.577$$

$f5 = 0.577 \times 0.5$

$= 0.2885$

Step 1 Forward Pass

$f6 = \sigma(f5)$

$$= \frac{1}{1+e^{-0.2885}}$$

$= 0.574$

Layer 1

$z_1 = x_1 w_1$

$\Rightarrow 1 \times 0.5 \Rightarrow 0.5$

$a_1 = \sigma(z_1)$

$$= \frac{1}{1+e^{-2}} \Rightarrow \frac{1}{1+0.606} \Rightarrow \frac{1}{1.606}$$

$\Rightarrow 0.622$

For Vanishing use Sigmoid

Layer 2

$$Z_2 = a_1 \times w_2 \Rightarrow 0.622 \times 0.5 \Rightarrow 0.311$$

$$\cancel{a_2}\ a_2 = \sigma(z_2) \Rightarrow \frac{1}{1+e^{-z}} \Rightarrow \frac{1}{1+e^{-0.311}}$$

$$\Rightarrow \frac{1}{1.732} \Rightarrow 0.577$$

Layer 3 (Output layer)

$$Z_3 = a_2 \times w_3 \Rightarrow 0.577 \times 0.5 \Rightarrow 0.2885$$

$$a_3 = \sigma(z_3) \Rightarrow \frac{1}{1+e^{-0.288}} \Rightarrow \frac{1}{1.749} \Rightarrow 0.572$$

Loss: $L = \frac{1}{2}(a_3 - y)^2$

$$= \frac{1}{2}(0.572 - 1)^2 \Rightarrow 0.092. \Big\}$$ need to check

Gradient Calculation $\rightarrow$ $\frac{dL}{da_3} = a_3 - y \Rightarrow 0.572 - 1$

$$\Rightarrow -0.428$$

Output layer:

$$\frac{dL}{dz_3} = \frac{dL}{da_3} \cdot \sigma'(z_3) = \frac{dL}{da_3} \cdot a_3(1-a_3)$$

$$\cancel{= -0.428}$$

$$z_3(1-z_3) \Rightarrow \cancel{0.2885(1-0.2885)}$$

$$\cancel{= 0.2885 \cdot 0.7115 \Rightarrow 0.205}$$

$$\frac{dL}{dz_3} \Rightarrow \cancel{(-0.428)(0.205)} \Rightarrow \cancel{-0.09}$$

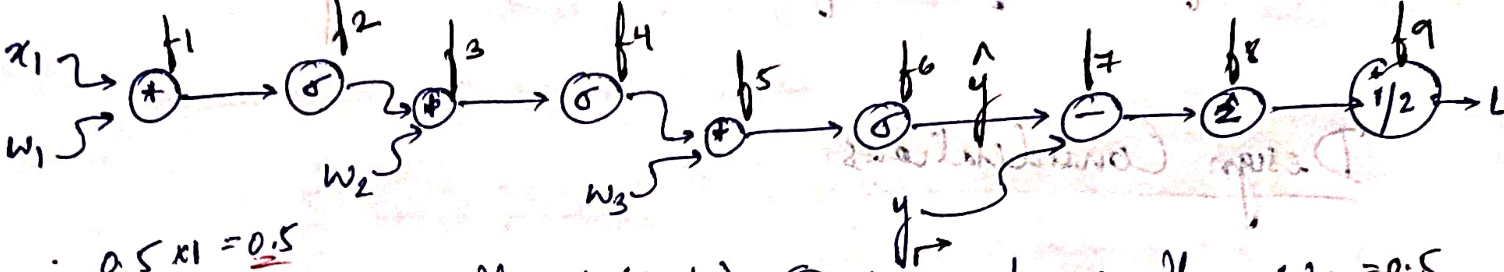
$$a_3(1-a_3) = 0.572(1-0.572) \Rightarrow 0.572 \times 0.428$$

$$\Rightarrow 0.244$$

$$\frac{dL}{dz_3} = 0.244 \times -0.428 \Rightarrow -0.104.$$

~~2nd layer~~ $\frac{dL}{da_2} = \frac{\cancel{a_2 - y}}{da_2}$ Use computational graph.

$\cancel{\frac{df_9}{df_9} = 1}$ $\quad \frac{\partial f_9}{\partial f_8} = \cancel{\frac{dL}{\partial L} = 1}$ $\quad \frac{\partial f_9}{\partial L} = \cancel{\frac{dL}{df_9} = 1 \times} \frac{\partial L}{\partial}$



① $f_1 = x_1 w_1 = 0.5 \times 1 = 0.5$ ② $f_2 = \sigma f_1 \Rightarrow \frac{\partial f_2}{\partial f_1} = f_1(1-f_1)$ ③ $f_3 = w_2 f_2 \Rightarrow \frac{\partial f_3}{\partial f_2} = w_2 = 0.5$

$\frac{\partial f_1}{\partial x} = w_1$, $\frac{\partial f_1}{\partial w_1} = x$

④ $f_4 = \sigma(f_3)$ $\quad \frac{\partial f_3}{\partial w_2} = f_2 = 0.25$

$\frac{\partial f_4}{\partial f_3} = f_3(1 - f_3)$

⑥ $f_6 = \sigma(f_5)$ ⑦ $f_7 = (f_6 - y)$

⑤ $f_5 = w_3 f_4 \Rightarrow \frac{df_5}{df_4} = w_3$, $\frac{df_5}{dw_3} = f_4$ $\quad \frac{df_6}{df_5} = f_5(1-f_5)$ $\quad \frac{df_7}{df_6} = \cancel{1 - 0.576} = \cancel{0.428} = 1$

⑧ $f_8 = (f_7)^2 \Rightarrow \frac{df_8}{df_7} = 2(f_7)$ ⑨ $f_9 = \frac{f_8}{2} \Rightarrow \frac{df_9}{df_8} = 0.5$

values in the next page

Backpropagation

$\cancel{\frac{dL}{dL} = 1}$ $\quad \frac{df_9}{df_9} = 1$ $\quad \frac{df_9}{df_8} = \frac{df_9}{df_8} \times \frac{dL}{df_9}$

$= 0.5 \times 1 \Rightarrow \cancel{0.25}\ 0.5$

$\frac{dL}{df_7} = \frac{dL}{df_8} \times \frac{df_8}{df_7}$

$\Rightarrow 0.25 \times 2(f_7) \Rightarrow 0.5 \times 0.428 \times 2 \Rightarrow 0.428$

$\frac{dL}{df_6} = \frac{dL}{df_7} \times \frac{df_7}{df_6} = 0.428 \times$

Batch norm → Batch Normalization.

Normalization per batch.

Normalizing the inputs first

Trained during any [illegible]

Introduce before Activation functions

- nn.Linear( )
- nn.Batchnorm( )
- nn.ReLu( ).

Mini Batch GD.

Backpropagation

$\frac{df_9}{df_9} = 1$ $\quad \frac{df_9}{df_8} = 1 \times 0.5 = 0.5$ $\quad \frac{df_9}{df_7} = \frac{df_8}{df_7} \times \frac{df_9}{df_8}$

$= 2(f_7) \times 0.5$

$= 2(0.428) \times 0.5$

$= 0.428$

$f_1 = 0.8$ $\quad f_2 = \sigma f_1 \Rightarrow \sigma f_1(1-\sigma f_1)$

$\Rightarrow 0.622(1-0.622)$

$\Rightarrow 0.235$

$f_3 = w_2 f_2 \Rightarrow \frac{df_3}{df_2} = 0.5$

$\frac{df_3}{dw_2} = 0.25$

$f_4 \Rightarrow \sigma f_3$

$\frac{df_4}{df_3} = \sigma f_3(1-\sigma f_3)$

$= 0.577(1-0.577)$

$= 0.244$

$f_5 = w_3 = 0.5$

$f_4 = 0.244$

$f_6 = \sigma f_5(1-\sigma f_5) \Rightarrow 0.574(1-0.574)$

$= \cancel{0.244}\ 0.245$

$f_7 = 0.428$

$f_8 = 2(f_7) = 0.856$

$f_9 = 0.5$

$\frac{df_9}{df_6} = \frac{df_7}{df_6} \times \frac{df_9}{df_7}$

$1 = \cancel{0.428} \times 0.428$

$= \cancel{0.1831}\ 0.428$

$\frac{df_9}{df_5} = \frac{df_6}{df_5} \times \frac{df_9}{df_6}$

$= 0.245 \times 0.1831$

$= 0.04448$

$\simeq 0.045$

$$\frac{df_9}{df_4} = \frac{df_5}{df_4} \times \frac{df_9}{df_5} \Rightarrow 0.5 \times 0.045 \Rightarrow 0.0225$$

$$\frac{df_9}{df_3} = \frac{df_4}{df_3} \times \frac{df_9}{df_4} \Rightarrow 0.244 \times 0.0225 \Rightarrow 0.00549$$

$$\simeq 0.0055$$

$$\frac{df_9}{df_2} = \frac{df_3}{df_2} \times \frac{df_9}{df_3} \Rightarrow 0.5 \times 0.0055 \Rightarrow 0.002745$$

$$\frac{df_9}{df_1} = \frac{df_2}{df_1} \times \frac{df_9}{df_2} \Rightarrow 0.235 \times 0.00274 \Rightarrow 0.0006450$$

$$\frac{df_9}{df_6} = \frac{df_7}{df_6} \times \frac{df_9}{df_7} \Rightarrow 1 \times 0.428 \Rightarrow 0.428$$

$$\frac{df_9}{df_5} = \frac{df_6}{df_5} \times \frac{df_9}{df_6} \Rightarrow 0.245 \times 0.428 = 0.1048$$

$$\simeq 0.105$$

$$\frac{df_9}{df_4} = \frac{df_5}{df_4} \times \frac{df_9}{df_5} \Rightarrow 0.5 \times 0.105 = \cancel{0.0525}\ 0.0525$$

$$\frac{df_9}{df_3} = \frac{df_4}{df_3} \times \frac{df_9}{df_4} \Rightarrow 0.0525 \times 0.244 \Rightarrow 0.01281$$

$$\simeq 0.013$$

$$\frac{df_9}{df_2} = \frac{df_3}{df_2} \times \frac{df_9}{df_3} \Rightarrow 0.5 \times 0.013 \Rightarrow 0.00064$$

$$\frac{df_9}{df_1} = \frac{df_2}{df_1} \times \frac{df_9}{df_2} \Rightarrow 0.235 \times 0.00064 \Rightarrow 0.000150$$

Vanishing Gradient Final value

Exploding Gradient

~~$f_1 = x w_1$~~ Taking the weights as 5

$f_1 = x w_1 \Rightarrow 1 \times 5$

$f_2 = ReLU(f_1) \Rightarrow \frac{5+|5|}{2} = 5$ $\quad \frac{df_2}{df_1} = 1$

$f_3 = w_2 f_2 \Rightarrow 5 \times 5 = 25$ $\quad \frac{df_3}{df_2} = w_2 = 5$

$f_4 = ReLU(f_3) = \frac{25+25}{2} = 25$ $\quad \frac{df_4}{df_3} = 1$

$f_5 = w_3 f_4 \Rightarrow 5 \times 25 \Rightarrow 125$

$f_6 = ReLU(f_5) = 125$ $\quad \frac{df_5}{df_4} = w_3 = 5$

$f_7 = f_6 - y \Rightarrow 125 - 1 \Rightarrow 124.$

$f_8 = (f_7)^2 \Rightarrow (124)^2 \Rightarrow 15376$ $\quad \frac{df_6}{df_5} = 1 \quad \frac{df_7}{df_6} = 1$

$f_9 = 15376 \times 0.5 \Rightarrow 7688$

$\frac{df_9}{df_8} = 0.5$ $\quad \frac{df_8}{df_7} = 2f_7 = 2 \times 124 = 248$

Calculations :-

$\frac{df_9}{df_9} = 1$ $\quad \frac{df_9}{df_8} = \frac{df_9}{df_8} \quad 1 \times 0.5 \Rightarrow \underline{0.5}$

$\frac{df_9}{df_7} = \frac{df_8}{df_7} \times \frac{df_9}{df_8} \Rightarrow 248 \times 0.5 \Rightarrow \underline{124}$

$\frac{df_9}{df_6} = \frac{df_7}{df_6} \times \frac{df_9}{df_7} \Rightarrow 1 \times 124 \Rightarrow \underline{124}$ $\quad \frac{df_9}{df_5} = \frac{df_6}{df_5} \times \frac{df_9}{df_6} = 1 \times 124 = \underline{124}$

$\frac{df_9}{df_4} = \frac{df_5}{df_4} \times \frac{df_9}{df_5} = 5 \times 124 \Rightarrow \underline{620}$

$$\frac{df_9}{df_3} = \frac{df_4}{df_3} \times \frac{df_9}{df_4} \Rightarrow 1 \times 620 \Rightarrow 620$$

$$\frac{df_9}{df_2} = \frac{df_3}{df_2} \times \frac{df_9}{df_3} \Rightarrow 5 \times 620 \Rightarrow 3100$$

$$\frac{df_9}{df_1} = \frac{df_2}{df_1} \times \frac{df_9}{df_2} \Rightarrow 1 \times 3100 \Rightarrow 3100$$

$$\frac{df_9}{dx} \Rightarrow \cancel{3100 \times} \frac{df_1}{dx} \times \frac{df_9}{df_1} \Rightarrow 5 \times 3100 \Rightarrow 15500$$

Final Exploding Gradient.