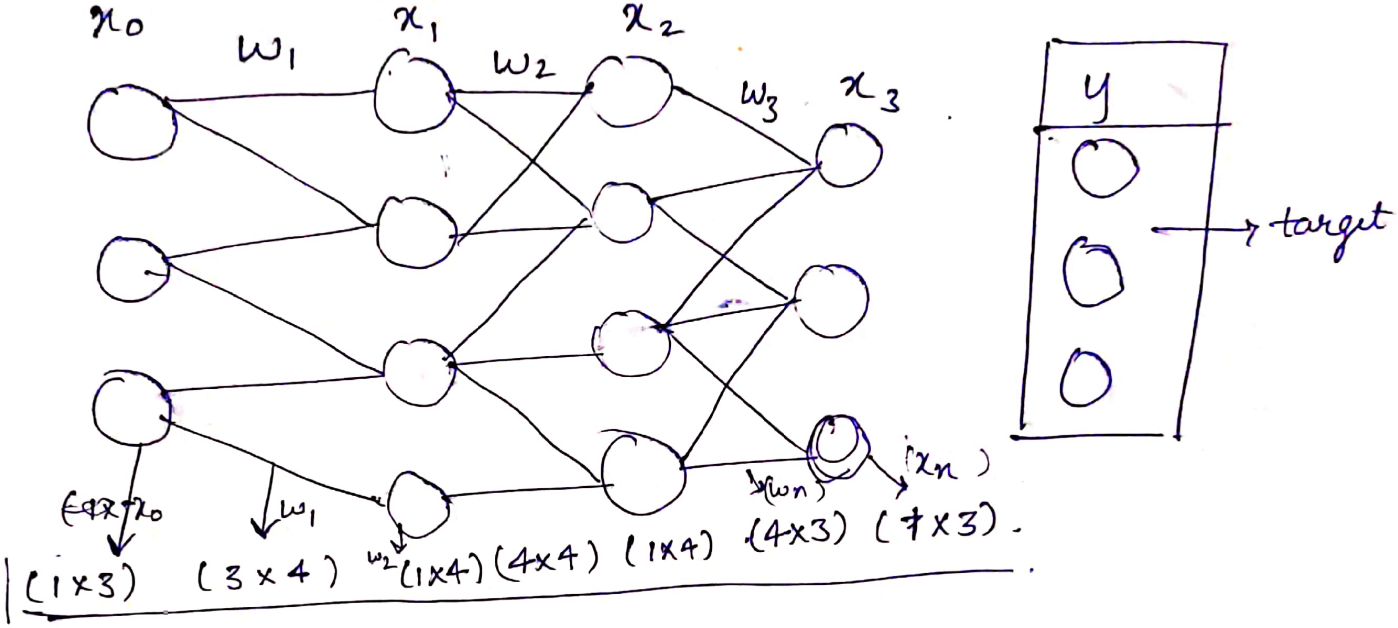

Lets assume our target to 'y'.

Here x_0 is a matrix (1×3).

w_1 is a matrix (3×4).

so ~~x_1 ...~~ $x_1 = f(w_0 x_0)$..

$\Rightarrow$ x_1 is $(1\times 3)\cdot(3\times 4)$ $\Rightarrow$ ~~(1×4)~~

$\Rightarrow$ x_i is (1×4) matrix.

Lets says our activation function is $f(x)$.

Then;

$$x_1 = f(w_1 x_0) + b_0 \longrightarrow (1)$$

$$x_2 = f(w_2 x_1) + b_1 \longrightarrow (2)$$

$$x_3 = f(w_3 x_2) + b_2 \longrightarrow (3)$$

Ignore b_0, b_1, b_2 because they are constants and will be added after back propogation step.

Lets assume the error function to be: $\frac{1}{2}(y - x_3)^2$

$$\boxed{E = \frac{1}{2}(y - x_3)^2} \quad \text{or} \quad \boxed{E = \frac{1}{2}(x_3 - y)^2}$$

Now, lets see how this error was affected by w_3. For that we need the derivative with respect to w_3.

$$\frac{\partial E}{\partial w_3} = \frac{\partial}{\partial w_3}\left(\frac{1}{2}(x_3 - y)^2\right)$$

$$= (x_3 - y) \cdot \frac{\partial x_3}{\partial w_3}$$

$$= (x_3 - y) \cdot \frac{\partial (f(w_3 x_2))}{\partial w_3}$$

$$= [(x_3 - y)\, f'(w_3 x_2)] \cdot \frac{\partial (w_3 x_2)}{\partial w_3}$$

$$= ((x_3 - y) \cdot f'(w_3 x_2)) \cdot x_2$$

$$\boxed{\frac{\partial E}{\partial w_3} = k_3 \cdot x_2} \rightarrow (4) \qquad \boxed{k_3 = (x_3 - y) \cdot f'(w_3 x_2)} \rightarrow (5)$$

Now lets do the backward pass for w_2.

$$\frac{\partial E}{\partial w_2} = (x_3 - y) \cdot \frac{\partial x_3}{\partial w_2}$$

$$= (x_3 - y) \cdot \frac{\partial (f(w_3 x_2))}{\partial w_2} = ((x_3 - y) \cdot f'(w_3 x_2)) \cdot \frac{\partial (w_3 x_2)}{\partial w_2}$$

$$= ((x_3 - y) \cdot f'(w_3 x_2))\, w_3 \cdot \frac{\partial x_2}{\partial w_2}$$

From eqn (5):

$$= k_3 \cdot w_3 \cdot \frac{\partial x_2}{\partial w_2}$$

$$= k_3 w_3 \cdot \frac{\partial f(w_2 x_1)}{\partial w_2} = k_3 w_3 \cdot f'(w_2 x_1) \cdot \frac{\partial (w_2 x_1)}{\partial w_2}$$

$$\Rightarrow \boxed{\frac{\partial E}{\partial w_2} = (k_3 \cdot w_3 \cdot f'(w_2 x_1)) \cdot x_1}$$

$$\boxed{\frac{\partial E}{\partial w_2} = k_2 \cdot x_1.} \rightarrow (6)$$

$$\boxed{k_2 = (k_3 w_3 \cdot f'(w_2 x_1))}. \rightarrow (7)$$

similarly for backpass w.r.t $\underline{w_1}$:

$$\frac{\partial E}{\partial w_1} = (x_3 - y) \cdot \frac{\partial x_3}{\partial w_1}.$$

$$\frac{\partial E}{\partial w_1} = (x_3 - y) \cdot \frac{\partial (f(w_3 x_2))}{\partial w_1}. = ((x_3 - y) \cdot f'(w_3 x_2)) \cdot \frac{\partial (w_3 x_2)}{\partial w_1}$$

from (5):

$$\frac{\partial E}{\partial w_1} = k_3 \cdot w_3 \cdot \frac{\partial x_2}{\partial w_1}.$$

$$= k_3 \cdot w_3 \cdot \frac{\partial (f(w_2 x_1))}{\partial w_1}.$$

$$= \left(k_3 \; w_3 \cdot f'(w_2 x_1)\right) \cdot \frac{\partial (w_2 x_1)}{\partial w_1}.$$

From eqn 7

$$= \; k_2 \cdot w_2 \cdot \frac{\partial x_1}{\partial w_1}.$$

$$= \; k_2 \cdot w_2 \cdot \frac{\partial f(w_1 x_0)}{\partial w_1}.$$

$$= \; k_2 \, w_2 \cdot f'(w_1 x_0) \cdot \frac{\partial (w_1 x_0)}{\partial w_1}.$$

$$\boxed{\frac{\partial E}{\partial w_1} = (k_2 \cdot w_2 \cdot f'(w_1 x_0)) \cdot x_0}.$$

$$\Rightarrow \boxed{\frac{\partial E}{\partial w_1} = k_1 \cdot x_0.} \;(8) \qquad \boxed{k_1 = k_2 w_2 \cdot f'(w_1 x_0)} \;(9)$$

From eqns : 8, 6, 4 we see :

$$\frac{\partial E}{\partial w_3} = k_3 \cdot x_2 \;; \quad \frac{\partial E}{\partial w_2} = k_2 \, x_1 \;; \quad \frac{\partial E}{\partial w_1} = k_1 \, x_0 .$$

From eqn's : (9), (7), (5) we see :

$$k_3 = (x_3 - y) \cdot f'(w_3 x_2).$$

$$k_2 = k_3 w_3 \cdot f'(w_2 x_1)..$$

$$k_1 = k_2 w_2 \cdot f'(w_1 x_0).$$

So the trend is clear :

For the layer we calculate the "k_n"

$$\boxed{k_n = (x_n - y) \cdot f'(w_n x_{n-1})} \longrightarrow (10)$$

For any other layer "i" we calculate "k_i".

$$\boxed{k_i = w_{i+1} \cdot k_{i+1} \cdot f'(w_i x_{i-1})} \longrightarrow (11).$$

$$\boxed{\frac{\partial E}{\partial w_i} = k_i \cdot x_{i-1}} \longrightarrow (12) \ (\text{where } (0 <= i < n)).$$

CLARIFICATIONS:

① How do you use this ?

→ In a function like gradient descent.

$$\frac{\partial E}{\partial w_i} = k_i x_{i-1}$$

$$\boxed{w_i = w_i - \alpha w_i \cdot \frac{\partial E}{\partial w_i}}$$

$\alpha \to$ learning rate.

② How do I find derivative of matrix while calculating k_i ?

(Ans) you don't ; suppose $x_0 = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}_{1\times 3}$

$$w_1 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}_{3\times 2}$$

x_{po}- $\quad w_1 x_0 = \begin{bmatrix} 2 & 0 \end{bmatrix}$

~~Lets say~~ $\cancel{f(x) = \frac{1}{1+e^{-x}}}$

Lets say $f(x) = \frac{1}{1+e^{-x}}$ (sigmoid function)..

$f'(x) = f(x)(1-f(x))$. ($\because$ google).

Now;

$$x_1 = f(w_1 x_0) = \left[\frac{1}{1+e^{-2}} \quad \frac{1}{1+e^{0}} \right] = \left[\frac{e^2}{1+e^2} \quad \frac{1}{2} \right].$$

Now;

For $f'(w_1 x_0)$ we should expand $f'(x)$.

$$f'(x) = \frac{1}{(1+e^{-x})} \cdot \left(1 - \frac{1}{1+e^{-x}}\right).$$

$$f'(w_1 x_0) = \left[\frac{1}{1+e^{-2}} \left(1 - \frac{1}{1+e^{-2}}\right) \quad \frac{1}{1+e^{0}} \left(1 - \frac{1}{1+e^{0}}\right) \right]$$

$$= \left[\frac{1}{1+e^{-2}} \left(\frac{1+e^{-2}-1}{1+e^{-2}} \right) \quad \frac{1}{2} \times \frac{1}{2} \right].$$

$$\boxed{f'(w_1 x_0) = \left[\frac{e^{-2}}{(1+e^{-2})^2} \quad \frac{1}{4} \right]}$$

(Here you go for computers its very easy).