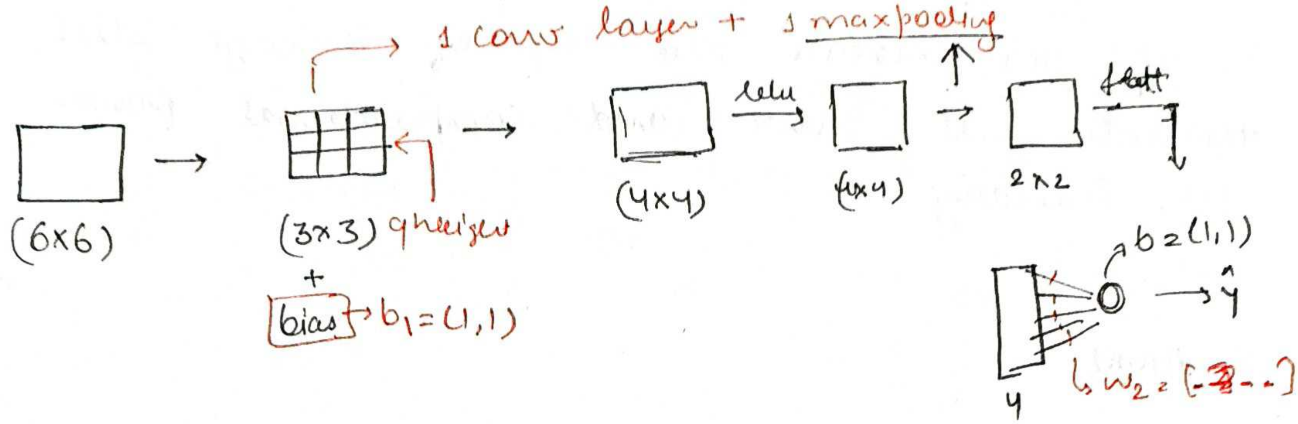
Backpropagation in CNN



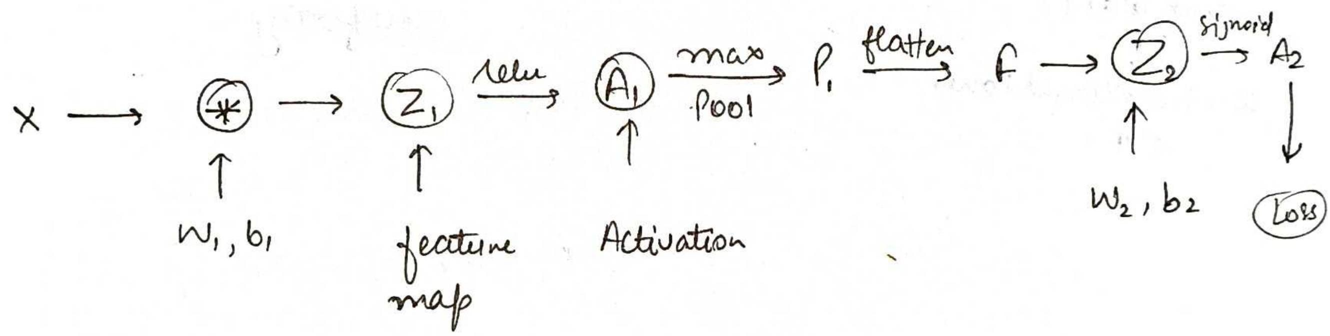
Loss → binary classification

$$\boxed{L = -y_i \log(\hat{y}_i) - (1-y_i)\log(1-\hat{y}_i)}$$

Trainable Parameter

$W_1 = (3,3)$ $\quad$ $W_2 = (1,4)$ $\Rightarrow$ 15 trainable parameter

$b_1 = (1,1)$ $\quad$ $b_2 = (1,1)$



Forward Propagation rough eqn

$Z_1 = \text{conv}(X, W_1) + b_1$

$A_1 = \text{relu}(Z_1)$

$P_1 = \text{maxpool}(A_1)$

$F = \text{flatten}(P_1)$

$Z_2 = F W_2 + b_2$

$A_2 = \sigma(Z_2)$

$F = (4,n)$, $W = (1,4)$ $\quad$ b_2

$(1,4)(4,1) = (1,1)\ (1,1)$

$Z_2 = (1,1)$

$A_2 = \sigma(Z_2) = (1,1)$

We have to apply gradient descent algo

$$\boxed{w_1 = w_1 - \eta \frac{\partial L}{\partial w_1}} \qquad \boxed{b_1 = b_1 - \eta \frac{\partial L}{\partial w_1}}$$

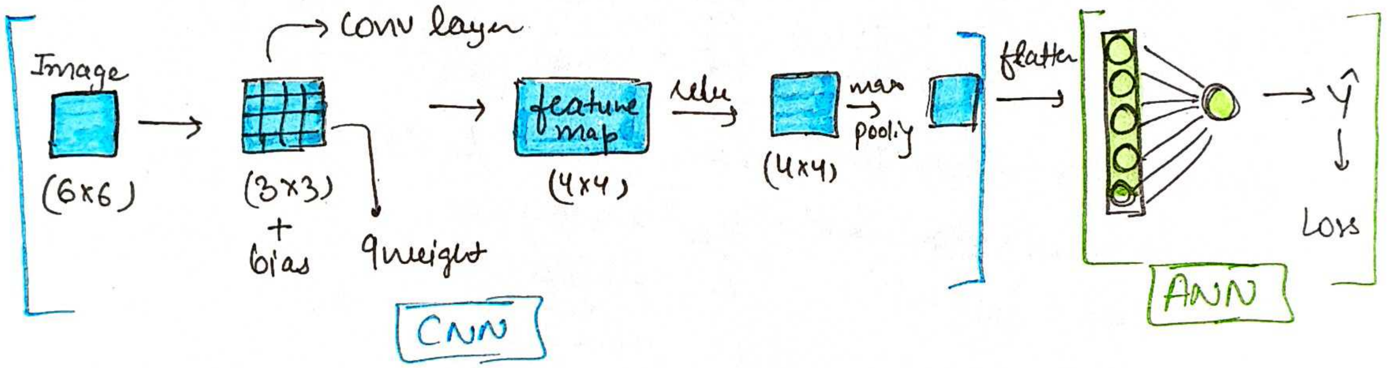
Objective: Find the value of w_1 → minimize the loss value
Find the value of b_1 → minimize the loss value

$$\boxed{w_1} = w_1 - \eta \boxed{\frac{\partial L}{\partial w_1}}$$

matrix

matrix

* For better understanding about back propagation let assume there are 2 parts → first (CNN)
→ second (ANN)



Gradient Descent

$$w_1 = w_1 - \eta \frac{\partial L}{\partial w_1} \qquad w_2 = w_2 - \eta \frac{\partial L}{\partial w_2}$$

$$b_1 = b_1 - \eta \frac{\partial L}{\partial b_1} \qquad b_2 = b_2 - \eta \frac{\partial L}{\partial b_2}$$

$$X \rightarrow \circledast \rightarrow Z_1 \xrightarrow{relu} A_1 \xrightarrow[pool]{max} P_1 \xrightarrow{Flatten} F \xrightarrow{\boxed{W_2 , b_2}} Z_2 \xrightarrow{Sigmoid} A_2(\hat{Y}) \rightarrow L$$

$$\uparrow$$

$$\boxed{W_1, b_1}$$

$\longleftarrow$ Start

$\frac{\partial L}{\partial W_2}$ → when W_2 change → how much L change
but W_2 and L are indirect ~~not~~ connected

$$\left[\frac{\partial L}{\partial W_2} = \boxed{\frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2}} \times \frac{\partial Z_2}{\partial W_2}\right]$$

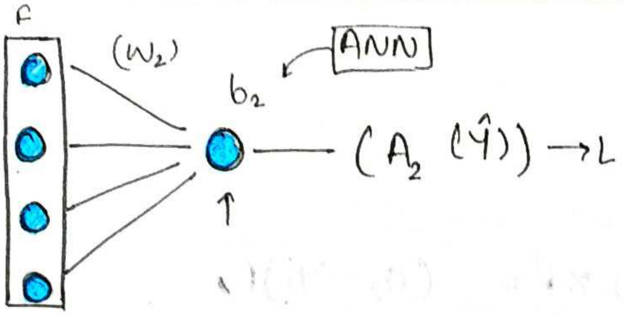
$$\left[\frac{\partial L}{\partial b_2} = \boxed{\frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2}} \times \frac{\partial Z_2}{\partial b_2}\right]$$

$$\left[\boxed{\frac{\partial L}{\partial W_1}} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial F} \times \frac{\partial F}{\partial P_1} \times \frac{\partial P_1}{\partial A_1} \times \frac{\partial A_1}{\partial Z_1} \times \frac{\partial Z_1}{\partial W_1}\right]$$

$$\left[\boxed{\frac{\partial L}{\partial b_1}} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial F} \times \frac{\partial F}{\partial P_1} \times \frac{\partial P_1}{\partial A_1} \times \frac{\partial A_1}{\partial Z_1} \times \frac{\partial Z_1}{\partial b_1}\right]$$

for finding these derivative we have to learn
How backpropagation on to apply

1. Convolution
2. Flatten
3. Max pooling



$$\frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial A_2}\frac{\partial A_2}{\partial Z_2}\frac{\partial Z_2}{\partial W_2}$$

$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial A_2}\frac{\partial A_2}{\partial Z_2}\frac{\partial Z_2}{\partial b_2}$$

forward propagation eqⁿ

$$\left\{\begin{aligned} Z_2 &= W_2 f + b_2 \\ A_2 &= \sigma(Z_2) \end{aligned}\right\}$$

let assume, we are working on single image
so, we can write $\frac{\partial L}{\partial a_2}$ for single image

$$\frac{\partial L}{\partial a_2} = \frac{\partial}{\partial a_2}\left[-y_i \log(a_2) - (1-y_i)\log(1-a_2)\right]$$

$$= \frac{-y_i}{a_2} + \frac{(1-y_i)}{(1-a_2)} \Rightarrow \frac{-y_i(1-a_2) + a_2(1-y_i)}{a_2(1-a_2)}$$

$$= \frac{-y_i + \cancel{y_i a_2} + a_2 - \cancel{a_2 y_i}}{a_2(1-a_2)}$$

$$\boxed{\frac{\partial L}{\partial a_2} = \frac{(a_2 - y_i)}{a_2(1-a_2)}}$$

$$\boxed{\frac{\partial A_2}{\partial Z_2} = \sigma(Z_2)\left[1-\sigma(Z_2)\right] = a_2\left[1-a_2\right]}$$

$$\boxed{\frac{\partial z_2}{\partial w_2} = f} \qquad \boxed{\frac{\partial z_2}{\partial b_2} = 1}$$

$$\frac{\partial L}{\partial w_2} = \frac{(a_2 - y_i)}{\cancel{a_2(1-a_2)}} \times \cancel{a_2(1-a_2)} \times f = (a_2 - y_i)f$$

Now, replace $a_2 = A_2$, $y_i = Y$

$$\boxed{\frac{\partial L}{\partial w_2} = (A_2 - Y)f} \Rightarrow \boxed{\frac{\partial L}{\partial w_2} = (A_2 - Y)f^T}$$

$$\boxed{\frac{\partial L}{\partial b_2} = \frac{a_2 - y_i}{\cancel{a_2(1-a_2)}} \cancel{a_2(1-a_2)} \times 1 = (A_2 - Y)}$$

Shape Analysis:

We want shape of $\frac{\partial L}{\partial w_2}$ same as W_2. We discussed shape at starting (Trainable parameter topic)

$W_2 = (1,4) \longrightarrow$ want $\frac{\partial L}{\partial w_2}$ also $(1,4)$

$$\underbrace{\underbrace{(A_2 - Y)}_{(1,1)\ (1,1)}}_{(1,1)} \underbrace{f}_{(4,1) \xrightarrow{\text{Transpose}} (1,4)}$$

$(1,1)$ $(1,4)$

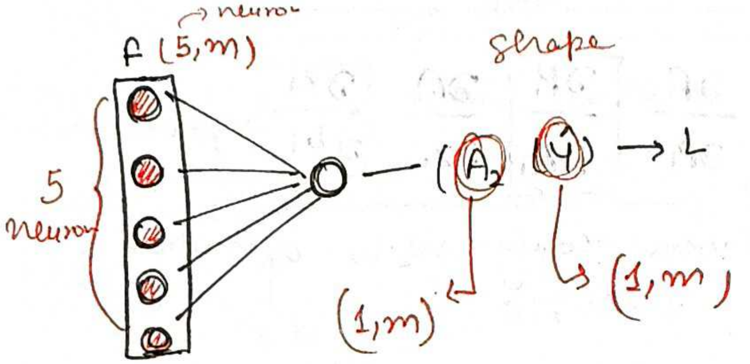
$\frac{\partial L}{\partial w_2} = [1,4]$

* All these shape discussed at starting topic of back prop

$$\boxed{\frac{\partial L}{\partial w_2} = (A_2 - Y)f^T}$$

* If we are using "Mini Batch propagation"
Sending multiple $x^{(m)}$ image (like 32, 64 image) together
And apply back prop on 32 images together.

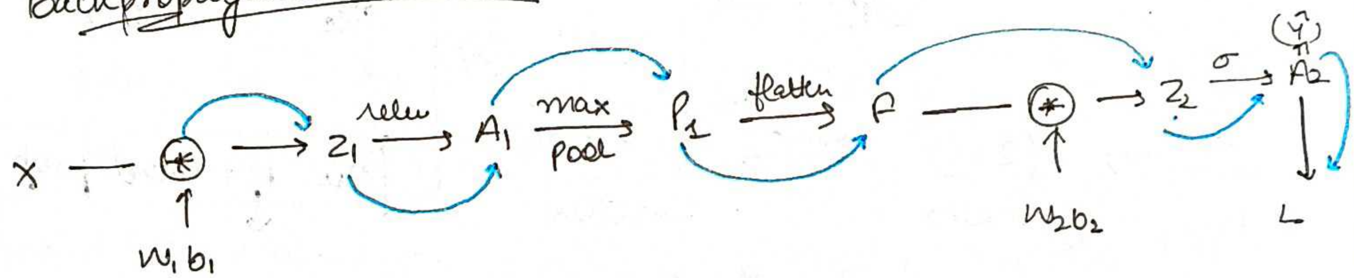
So, f will be (4, m) size



$$\boxed{\frac{\partial L}{\partial w_2} = \frac{1}{m}(A_2 - Y)f^T}$$

$$\boxed{\frac{\partial L}{\partial b_2} = \frac{1}{m}(A_2 - Y)}$$

Backpropagation Part 2



Forward Propagation

$Z_1 = \text{conv}(X, W_1) + b_1$

$A_1 = \text{relu}(Z_1)$

$P_1 = \text{maxpool}(A_1)$

$f = \text{flatten}(P_1)$

$Z_2 = W_2 f + b_2$

$A_2 = \sigma(Z_2)$

$$L = \frac{1}{m}\sum_{i=1}^{m}\left[-Y_i \log(A_2) - (1 - Y_i)\log(A_2)\right]$$

$$\boxed{w_1 = w_1 - \eta \frac{\partial L}{\partial w_1}} \qquad \boxed{b_1 = b_1 - \eta \frac{\partial L}{\partial b_1}}$$

$$\frac{\partial L}{\partial w_1} = \boxed{\frac{\partial L}{\partial A_2} \frac{\partial A_2}{\partial Z_2}} \frac{\partial Z_2}{\partial f} \frac{\partial f}{\partial P_1} \boxed{\frac{\partial P_1}{\partial A_1}} \frac{\partial A_1}{\partial Z_1} \frac{\partial Z_1}{\partial w_1}$$

$$\frac{\partial L}{\partial b_1} = \boxed{\frac{\partial L}{\partial A_2} \frac{\partial A_2}{\partial Z_2}} \frac{\partial Z_2}{\partial f} \frac{\partial f}{\partial P_1} \boxed{\frac{\partial P_1}{\partial A_1}} \frac{\partial A_1}{\partial Z_1} \boxed{\frac{\partial Z_1}{\partial b_1}} \rightarrow \text{back prop Convolution layer}$$

→ Already found when solving $W_2\, b_2$ derivative

→ $(A_2 - Y)$

$$\boxed{\frac{\partial Z_2}{\partial f} = W_2}$$

$\frac{\partial f}{\partial P_1}$ → not trainable parameter

[In forward prop we convert max pool value to flatten but in back prop convert flatten → max. pool size size]

$$\frac{\partial f}{\partial P_1} = \text{reshape}(P_1.\text{shape})$$

$$\frac{\partial P_1}{\partial A_1} \rightarrow \underset{\text{matrix}}{(2\times2)} \xrightarrow{\text{convert}} \underset{\text{matrix}}{(4\times4)}$$

In forward prop

↓

No. contribution → Loss

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$
$$\begin{bmatrix} 9 & 10 \\ 11 & 12 \end{bmatrix} \begin{bmatrix} 13 & 14 \\ 15 & 16 \end{bmatrix} \rightarrow \begin{bmatrix} 4 & 8 \\ 12 & 16 \end{bmatrix}$$

(4x4) (2x2)

$$\boxed{\frac{\partial L}{\partial A_2} \frac{\partial A_2}{\partial Z_2} \frac{\partial Z_2}{\partial f} \frac{\partial f}{\partial P_1}} \rightarrow (2\times2)$$

$$\begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}$$

(2x2)

convert

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & x_1 & 0 & x_2 \\ 0 & 0 & 0 & 0 \\ 0 & x_3 & 0 & x_4 \end{bmatrix}$$

(4x4)

Only No. of Window contribute to find Loss. Rest of the no. are useless.

$$\frac{\partial A_1}{\partial Z_1} = \begin{cases} 1 & \text{if } Z_{1xy} > 0 \\ 0 & \text{if } Z_{1xy} < 0 \end{cases}$$

Backprop on convolution (b_1) $\quad \frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial Z_1} \times \frac{\partial Z_1}{\partial b_1}$

let assume for better understanding

let say

$$X \longrightarrow \circledast \longrightarrow Z_1$$

X: $(3,3)$; W_1, b_1: $(2,2)$; $\frac{\partial L}{\partial Z_1}$ $(2,2)$

$$\boxed{\frac{\partial L}{\partial A_2} \frac{\partial A_2}{\partial Z_2} \cdots \frac{\partial A_1}{\partial Z_1}} \rightarrow \frac{\partial L}{\partial Z_1} \rightarrow (2\times 2)$$

$$Z_1 = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$$

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \circledast W_1 = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} + b_1$$

$$Z_{11} = x_{11} w_{11} + x_{12} w_{12} + x_{21} w_{21} + x_{22} w_{22} + b_1$$

$$Z_{12} = x_{12} w_{11} + x_{13} w_{12} + x_{22} w_{21} + x_{23} w_{22} + b_1$$

$$Z_{21} = x_{21} w_{11} + x_{22} w_{12} + x_{31} w_{21} + x_{32} w_{22} + b_1$$

$$Z_{22} = x_{22} w_{11} + x_{23} w_{12} + x_{32} w_{21} + x_{33} w_{22} + b_1$$

$$\frac{\partial L}{\partial Z_1} = \begin{bmatrix} \frac{\partial L}{\partial Z_{11}} & \frac{\partial L}{\partial Z_{12}} \\ \frac{\partial L}{\partial Z_{21}} & \frac{\partial L}{\partial Z_{22}} \end{bmatrix}$$

Now we have to find $\frac{\partial z_1}{\partial b_1}$.

$\rightarrow z_1 \rightarrow$ differentiate $\rightarrow$ with b_1

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial z_1} \times \frac{\partial z_1}{\partial b_1} = \frac{\partial L}{\partial z_{11}} \times \frac{\partial z_1}{b_1} + \frac{\partial L}{\partial z_{12}} \times \frac{\partial z_{12}}{\partial b_1} + \frac{\partial L}{\partial z_{21}} \times \frac{\partial z_{21}}{\partial b_1} + \frac{\partial L}{\partial z_{22}} \times \frac{\partial z_{22}}{\partial b_1}$$

$$\frac{\partial z_1}{b_1} = \underbrace{x_{11} w_{11}}_{0} + \underbrace{w_{12} w_{12}}_{0} + \underbrace{x_{21} w_{21}}_{0} + \underbrace{x_{22} w_{22}}_{0} + \underbrace{b_1}_{1} = 1$$

Same as $\rightarrow$ every differentiate would be find

$$= \frac{\partial L}{\partial z_{11}}(1) + \frac{\partial L}{\partial z_{12}}(1) + \frac{\partial L}{\partial z_{21}}(1) + \frac{\partial L}{\partial z_{22}}(1)$$

$$\boxed{\frac{\partial L}{\partial b_1} = \text{sum}\left(\frac{\partial L}{\partial z_1}\right)} \rightarrow \text{Scaler}$$

for weights conv backprop

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z_1} \frac{\partial z_1}{\partial w_1}$$

$$\frac{\partial L}{\partial w_1} = \begin{bmatrix} \frac{\partial L}{\partial w_{11}} & \frac{\partial L}{\partial w_{12}} \\ \frac{\partial L}{\partial w_{21}} & \frac{\partial L}{\partial w_{22}} \end{bmatrix}$$

$$z_{11} = x_{11} w_{11} + x_{12} w_{12} + x_{21} w_{21} + x_{22} w_{22} + b_1$$

$$z_{12} = x_{12} w_{11} + \text{-------} + b_1$$

$$z_{21} = x_{21} w_{11} + \text{-------} + b_1$$

$$z_{22} = x_{22} w_{11} + \text{-------} + b_1$$

$$\frac{\partial L}{\partial z_1} = \begin{bmatrix} \frac{\partial L}{\partial z_{11}} & \frac{\partial L}{\partial z_{12}} \\ \frac{\partial L}{\partial z_{21}} & \frac{\partial L}{\partial z_{22}} \end{bmatrix}$$

- w_{11} using with every $z_{11}, z_{12}, z_{21}, z_{22}$

So we have to find]

Same as $w_{22} \rightarrow z_{11}, z_{12}$ ---

$w_{21} \rightarrow z_{11}, z_{12}$ —

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial L}{\partial Z_{11}} \times \boxed{\frac{\partial Z_{11}}{\partial w_{11}}} + \frac{\partial L}{\partial Z_{12}} \times \boxed{\frac{\partial Z_{12}}{\partial w_{11}}} + \frac{\partial L}{\partial Z_{21}} \times \boxed{\frac{\partial Z_{21}}{\partial w_{11}}} + \frac{\partial L}{\partial Z_{22}} \times \boxed{\frac{\partial Z_{22}}{\partial w_{11}}}$$

$$\frac{\partial L}{\partial w_{12}} = \frac{\partial L}{\partial Z_{11}} \times \boxed{\frac{\partial Z_{11}}{\partial w_{12}}} + \frac{\partial L}{\partial Z_{12}} \times \boxed{\frac{\partial Z_{12}}{\partial w_{12}}} + \frac{\partial L}{\partial Z_{21}} \times \boxed{\frac{\partial Z_{21}}{\partial w_{12}}} + \frac{\partial L}{\partial Z_{22}} \times \boxed{\frac{\partial Z_{22}}{\partial w_{12}}}$$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial L}{\partial Z_{11}} \times \boxed{\frac{\partial Z_{11}}{\partial w_{21}}} + \frac{\partial L}{\partial Z_{12}} \times \boxed{\frac{\partial Z_{12}}{\partial w_{21}}} + \frac{\partial L}{\partial Z_{21}} \times \boxed{\frac{\partial Z_{21}}{\partial w_{21}}} + \frac{\partial L}{\partial Z_{22}} \times \boxed{\frac{\partial Z_{22}}{\partial w_{21}}}$$

$$\frac{\partial L}{\partial w_{22}} = \frac{\partial L}{\partial Z_{11}} \times \boxed{\frac{\partial Z_{11}}{\partial w_{22}}} + \frac{\partial L}{\partial Z_{12}} \times \boxed{\frac{\partial Z_{12}}{\partial w_{22}}} + \frac{\partial L}{\partial Z_{21}} \times \boxed{\frac{\partial Z_{21}}{\partial w_{22}}} + \frac{\partial L}{\partial Z_{22}} \times \boxed{\frac{\partial Z_{22}}{\partial w_{22}}}$$

We find derivatin $\frac{\partial Z_{11}}{\partial w_{11}}, \frac{\partial Z_{12}}{\partial w_{11}}$ ... cause $w_{11}, w_{12}, w_{21}, w_{22}$ in every $Z_{11}, Z_{12}, Z_{21}, Z_{22}$

$Z_{11} = x_{11} w_{11} + \overbrace{x_{12} w_{12} \ldots}^{0} = x_{11}$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial L}{\partial Z_{11}} x_{11} + \frac{\partial L}{\partial Z_{12}} x_{12} + \frac{\partial L}{\partial Z_{21}} x_{21} + \frac{\partial L}{\partial Z_{22}} x_{22}$$

$$\frac{\partial L}{\partial w_{12}} = \frac{\partial L}{\partial Z_{11}} x_{12} + \frac{\partial L}{\partial Z_{12}} x_{13} + \frac{\partial L}{\partial Z_{21}} x_{22} + \frac{\partial L}{\partial Z_{22}} x_{23}$$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial L}{\partial Z_{11}} x_{21} + \frac{\partial L}{\partial Z_{12}} x_{22} + \frac{\partial L}{\partial Z_{21}} x_{31} + \frac{\partial L}{\partial Z_{22}} x_{32}$$

$$\frac{\partial L}{\partial w_{22}} = \frac{\partial L}{\partial Z_{11}} x_{22} + \frac{\partial L}{\partial Z_{12}} x_{23} + \frac{\partial L}{\partial Z_{21}} x_{32} + \frac{\partial L}{\partial Z_{22}} x_{33}$$

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \qquad \frac{\partial L}{\partial Z} = \begin{bmatrix} \frac{\partial L}{\partial Z_{11}} & \frac{\partial L}{\partial Z_{12}} \\ \frac{\partial L}{\partial Z_{21}} & \frac{\partial L}{\partial Z_{22}} \end{bmatrix}$$

$\frac{\partial L}{\partial Z}$ put on X