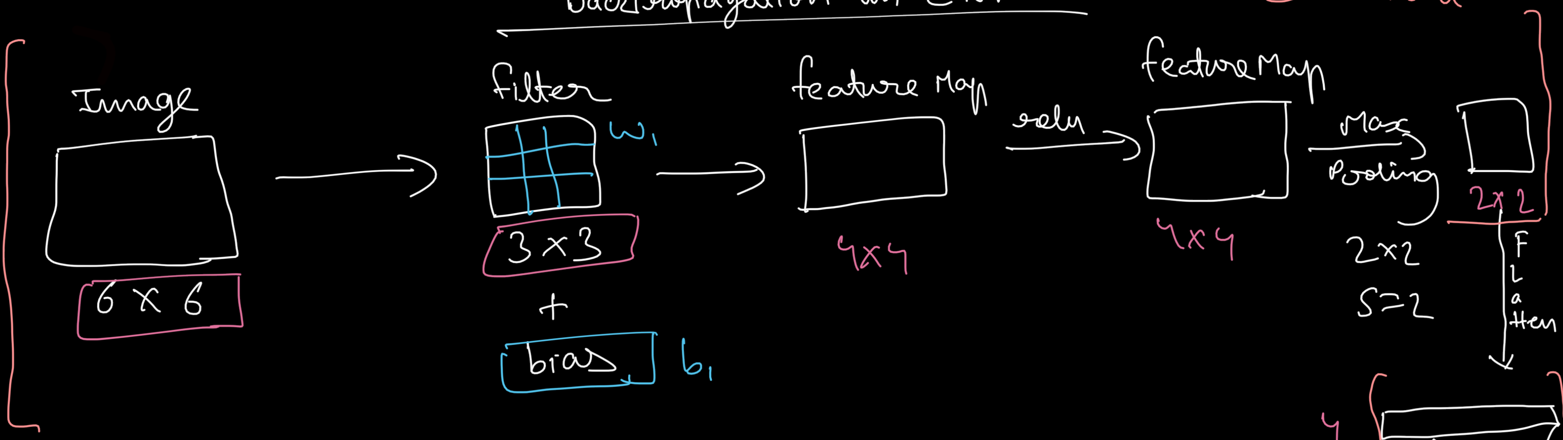
BackPropagation in CNN

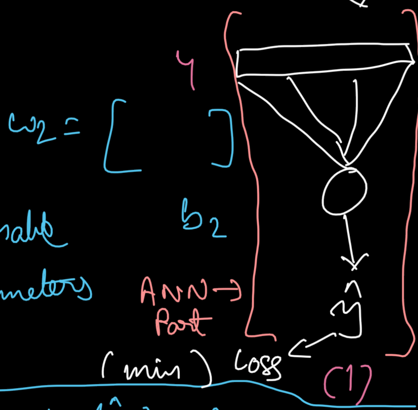
CNN Part



Trainable Parameters :

$w_1 = (3,3)$ $w_2 = (1,4)$

$b_1 = (1,1)$ $b_2 = (1,1)$ $\Rightarrow$ 15 trainable Parameters



$$\boxed{L = -Y_i \log(\hat{Y}_i) - (1-Y_i)\log(1-\hat{Y}_i)}$$

Loss → binary classification

Logical Flow

$$X \longrightarrow \circledast \longrightarrow Z_1 \xrightarrow{\text{relu}} A_1 \xrightarrow[\text{Pool}]{\text{max}} P_1 \xrightarrow{\text{Flatten}} F \longrightarrow Z_2 \xrightarrow{\text{Sigmoid}} A_2 \longrightarrow L$$

$\uparrow$

w_1, b_1

Forward Prop

$Z_1 = \text{Conv}(X, w_1) + b_1$

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

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

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

$Z_2 = w_2 F + b_2$

$A_2 = \sigma(Z_2)$

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

Gradient Descent

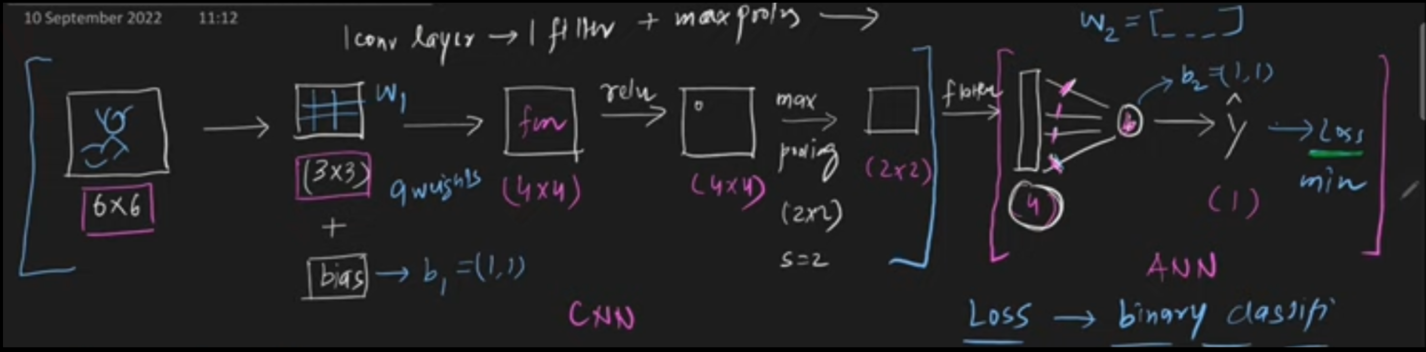
$w_1 = w_1 - \eta \frac{\partial L}{\partial w_1}$

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

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

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

End Goal to apply Gradient descent for trainable Parameters to get best value



$$X \longrightarrow \circledast \longrightarrow Z_1 \xrightarrow{\text{relu}} A_1 \xrightarrow[\text{Pool}]{\text{max}} P_1 \xrightarrow{\text{Flatten}} F \longrightarrow Z_2 \xrightarrow{\text{Sigmoid}} A_2 \longrightarrow L$$

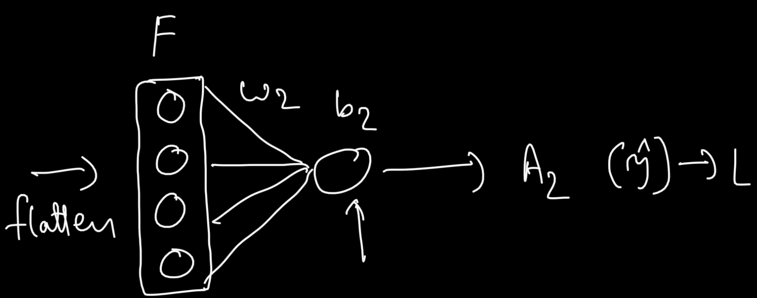
$\boxed{w_1, b_1}$ $\qquad$ $\boxed{w_2, b_2}$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial z_2} \times \frac{\partial z_2}{\partial w_2} \qquad \frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial z_2} \times \frac{\partial z_2}{\partial b_2}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial z_2} \times \frac{\partial z_2}{\partial F} \times \frac{\partial F}{\partial P_1} \times \frac{\partial P_1}{\partial A_1} \times \frac{\partial A_1}{\partial z_1} \times \frac{\partial z_1}{\partial w_1}$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial z_2} \times \frac{\partial z_2}{\partial F} \times \frac{\partial F}{\partial P_1} \times \frac{\partial P_1}{\partial A_1} \times \frac{\partial A_1}{\partial z_1} \times \frac{\partial z_1}{\partial b_1}$$

How to apply Backpropagation to
→ Convolution
→ flatten
→ Max Pooling



$$\boxed{\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial z_2} \times \frac{\partial z_2}{\partial w_2}}$$

$$\boxed{\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial z_2} \times \frac{\partial z_2}{\partial b_2}}$$

forward Prop eq^n

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

for a Single image

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

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

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

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

$$\frac{\partial A_2}{\partial z_2} = \sigma(z_2)[1-\sigma(z_2)] = a_2[1-a_2]$$

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

$$\boxed{\frac{\partial L}{\partial w_2}} = \frac{(a_2 - Y_i)}{\cancel{a_2(1-a_2)}} \times \cancel{a_2[1-a_2]} \times F = (a_2 - Y_i)F$$

$$= (A_2 - Y)F$$

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

$\underbrace{(1,1)\ (1,4)}_{(1,4)} \rightarrow F^T$

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

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

$$\boxed{\frac{\partial L}{\partial b_2} = A_2 - Y}$$

Part 2

$$\frac{\partial L}{\partial w_1} = \boxed{\underbrace{\frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2}}_{(A_2 - y)}} \times \frac{\partial Z_2}{\partial F} \times \frac{\partial F}{\partial P_1} \times \frac{\partial P_1}{\partial A_1} \times \frac{\partial A_1}{\partial Z_1} \times \underset{5}{\frac{\partial Z_1}{\partial w_1}}$$

6 derivatives to calculate

$$\frac{\partial L}{\partial b_1} = \boxed{\frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2}} \times \underset{1}{\frac{\partial Z_2}{\partial F}} \times \underset{2}{\frac{\partial F}{\partial P_1}} \times \underset{3}{\frac{\partial P_1}{\partial A_1}} \times \underset{4}{\frac{\partial A_1}{\partial Z_1}} \times \underset{6}{\frac{\partial Z_1}{\partial b_1}}$$

$$\frac{\partial Z_2}{\partial F} = w_2$$

$\frac{\partial F}{\partial P_1} \Rightarrow$ No trainable Parameter in Flatten; In back propagation $\Rightarrow$ reshape (P_1.Shape) means from flatten bring it in shape of Pooling

$\frac{\partial P_1}{\partial A_1} \Rightarrow$ No trainable Parameter in Pooling layer

This operation is reverse of max pooling, where from 2x2 we are creating 4x4 but all numbers are zero rather than maximum ones, and in place of maximum ones errors are setted up.

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

Backprop on Convolution

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

X: $(3,3)$

w_1, b_1: $(2,2)$

$\left(\frac{\partial L}{\partial Z_1}\right) \to (2,2)$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial Z_1} \times \boxed{\frac{\partial Z_1}{\partial b_1}}$$

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

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \qquad w_1 = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$$

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

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

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

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

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

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial Z_1} \times \frac{\partial Z_1}{\partial b_1} = \frac{\partial L}{\partial Z_{11}} \times \frac{\partial Z_{11}}{\partial b_1} + \frac{\partial L}{\partial Z_{12}} \times \frac{\partial Z_{12}}{\partial b_1} + \frac{\partial L}{\partial Z_{21}} \times \frac{\partial Z_{21}}{\partial b_1} + \frac{\partial L}{\partial Z_{22}} \times \frac{\partial Z_{22}}{\partial b_1}$$

$$= \left(\frac{\partial L}{\partial Z_{11}} + \frac{\partial L}{\partial Z_{12}} + \frac{\partial L}{\partial Z_{21}} + \frac{\partial L}{\partial Z_{22}} \right)$$

$$= \text{Sum}\left(\frac{\partial L}{\partial Z_1}\right)$$

$$\frac{\partial L}{\partial b_1} = \text{Sum}\left(\frac{\partial L}{\partial Z_1}\right) \to \text{Scalar}$$

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

$$\frac{\partial L}{\partial w_1} = \begin{bmatrix} \frac{\partial L}{\partial w_{11}} & \frac{\partial L}{\partial w_{12}} \\ \frac{\partial L}{\partial w_{21}} & \frac{\partial L}{\partial w_{22}} \end{bmatrix} \qquad \frac{\partial L}{\partial z_1} = \begin{bmatrix} \frac{\partial L}{\partial z_{11}} & \frac{\partial L}{\partial z_{12}} \\ \frac{\partial L}{\partial z_{21}} & \frac{\partial L}{\partial z_{22}} \end{bmatrix}$$

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

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

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

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

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

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

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

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

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

$$x = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{33} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

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

$$\frac{\partial L}{\partial w_1} = \text{Conv}\left(x, \frac{\partial L}{\partial z_1}\right)$$

$$\frac{\partial L}{\partial w_1} = \text{Conv}\left(x, \frac{\partial L}{\partial z_1}\right)$$

$$\frac{\partial L}{\partial b_1} = \text{Sum}\left(\frac{\partial L}{\partial z_1}\right)$$