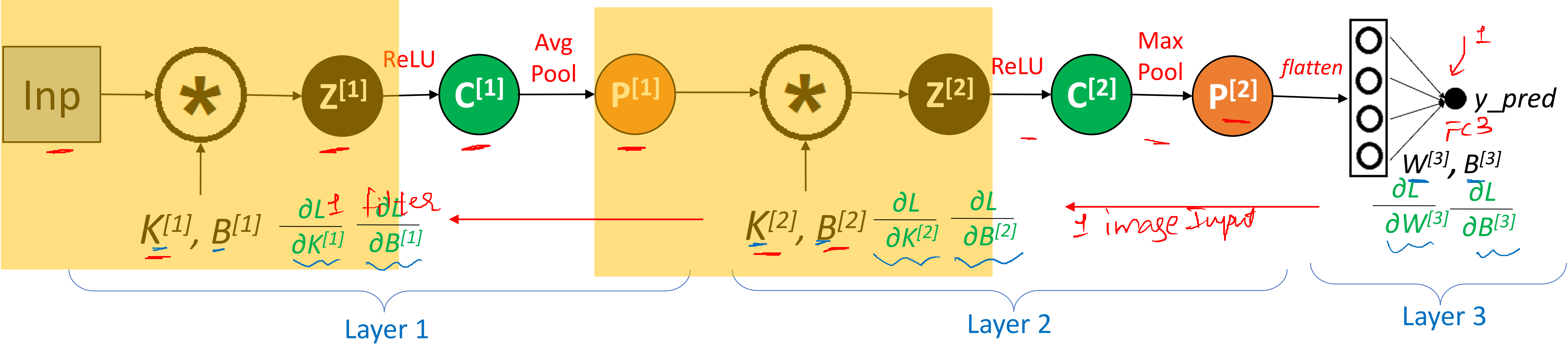
Backpropagation in CNN – Part 2



Forward Propagation

$Z^{[1]} = \textit{Conv}(Inp, K^{[1]}) + B^{[1]}$

$C^{[1]} = \textit{ReLU}(Z^{[1]})$

$P^{[1]} = \textit{AvgPool}(C^{[1]})$

$Z^{[2]} = \textit{Conv}(P^{[1]}, K^{[2]}) + B^{[2]}$

$C^{[2]} = \textit{ReLU}(Z^{[2]})$

$P^{[2]} = \textit{MaxPool}(C^{[2]})$

$f = \textit{flatten}(P^{[2]})$

$Z^{[3]} = W^{[3]} . f + B^{[3]}$

$a = y\_pred = \textit{sigmoid}(Z^{[3]})$

$$\text{Cost} = -\frac{1}{m} \sum_{i=1}^{m} [y_i * log(a_i) + (1 - y_i) * log(1 - a_i)]$$

Loss =

Weight Updation

learning rate

$$W_3 = W_3 - \alpha * \frac{\partial L}{\partial W_3}$$

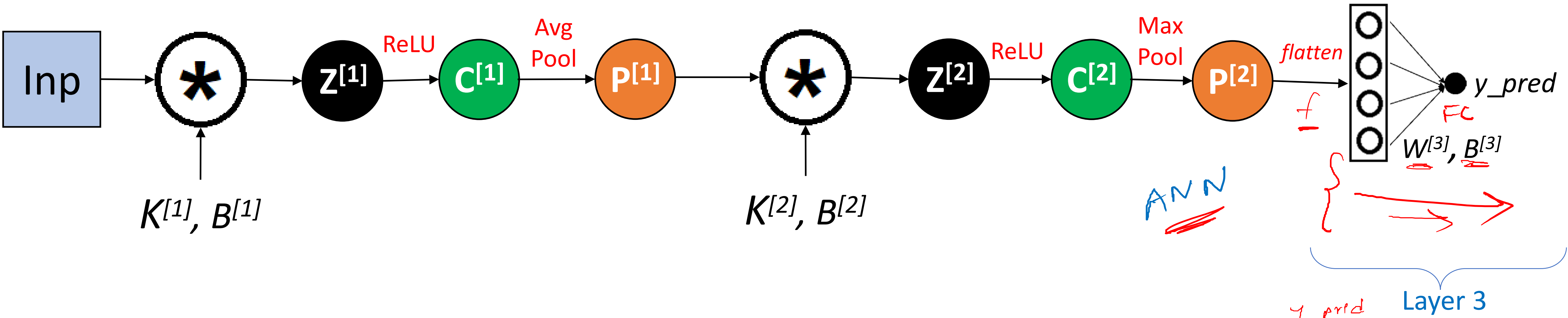
$$B_3 = B_3 - \alpha * \frac{\partial L}{\partial B_3}$$

$$K_2 = K_2 - \alpha * \frac{\partial L}{\partial K_2}$$

$$B_2 = B_2 - \alpha * \frac{\partial L}{\partial B_2}$$

$$K_1 = K_1 - \alpha * \frac{\partial L}{\partial K_1}$$

$$B_1 = B_1 - \alpha * \frac{\partial L}{\partial B_1}$$



Forward Propagation

$f = flatten(\ P^{[2]}\ )$

$Z^{[3]} = W^{[3]} \cdot f + B^{[3]}$

$a = y\_pred = sigmoid(\ Z^{[3]}\ )$

Backward Propagation

$$dZ^{[3]} = \frac{\partial L}{\partial Z^{[3]}} = (y\_pred - y)$$

$$dW^{[3]} = \frac{\partial L}{\partial W^{[3]}} = dZ^{[3]} * f^T$$

$$dB^{[3]} = \frac{\partial L}{\partial B^{[3]}} = dZ^{[3]}$$

$$df = \frac{\partial L}{\partial f} = W^{[3]T} * dZ^{[3]}$$

matrix multiplication

$\frac{\partial L}{\partial f}$

y-pred

$$dZ_3 = (A_3 - Y)$$

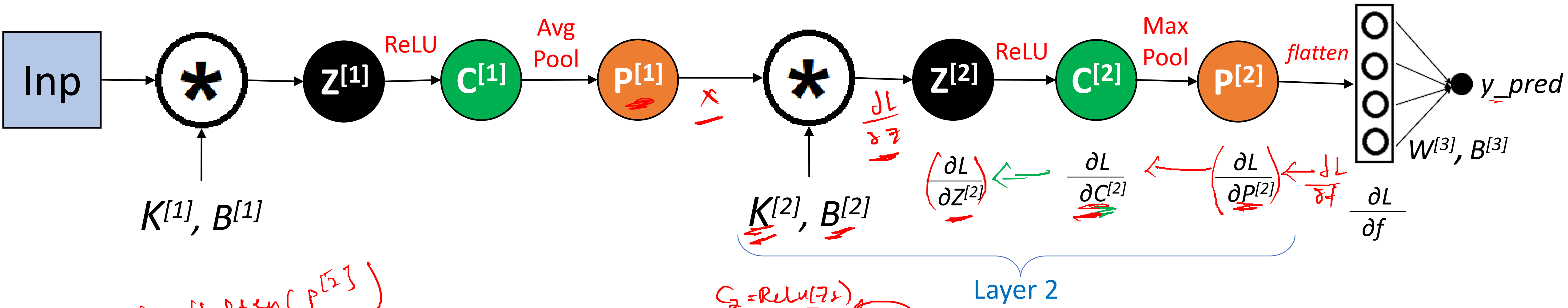
$$\frac{\partial cost}{\partial W_3} = dW_3 = \frac{1}{m} . dZ_3 . A_2^T$$

A_2

$$dB_3 = \frac{1}{m} . sum(dZ_3, 1)$$

dA_2

$$dZ_2 = (W_3^T . dZ_3) * f_2'(Z_2)$$



f = flatten(P[2])

$$dP^{[2]} = \frac{\partial L}{\partial P^{[2]}} = df.reshape(P^{[2]}.shape)$$

$C^{[2]} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ f = 2x2   4x4

$P^{[2]} = [4]$ 2x2

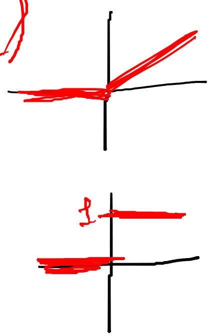
$\frac{\partial L}{\partial C} = 0$

$$\frac{\partial L}{\partial C^{[2]}} = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix} \qquad \frac{\partial L}{\partial P^{[2]}} = [2] \quad 2\times 2$$

$$dC^{[2]}_{mn} = \frac{\partial L}{\partial C^{[2]}_{mn}} = \begin{cases} \frac{\partial L}{\partial P_{xy}^{[2]}} & \text{, If } C_{mn} \text{ is the max element} \\ 0 & \text{, otherwise} \end{cases}$$

$C_2 = ReLU(Z_2)$

$$\frac{\partial L}{\partial Z} = \frac{\partial L}{\partial C_2}\left(\frac{\partial C_2}{\partial Z_2}\right)$$



$$Z^{[2]} = \begin{bmatrix} 1.02 & -0.25 \\ -3.45 & 2.2 \end{bmatrix}$$

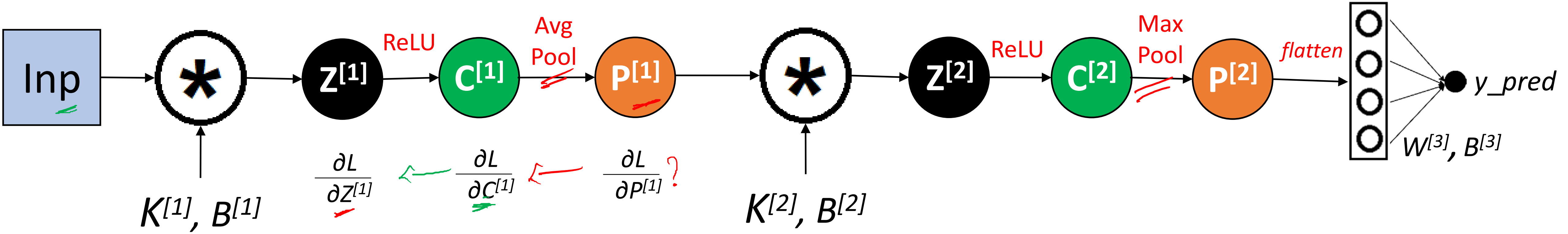
$$\frac{\partial C^{[2]}}{\partial Z^{[2]}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\frac{\partial C^{[2]}}{\partial Z_{mn}^{[2]}} = \begin{cases} 1 & \text{, If } Z_{mn} > 0 \\ 0 & \text{, If } Z_{mn} < 0 \end{cases}$$

$$dZ = \frac{\partial L}{\partial Z^{[2]}} = \left(\frac{\partial L}{\partial C^{[2]}}\right) * \left(\frac{\partial C^{[2]}}{\partial Z^{[2]}}\right)$$

$$dK^{[2]} = \frac{\partial L}{\partial K^{[2]}} = conv(P^{[1]}, dZ^{[2]})$$

$$dB^{[2]} = \frac{\partial L}{\partial B^{[2]}} = sum(dZ^{[2]})$$



$\frac{\partial L}{\partial x}$

$$\frac{\partial L}{\partial P^{[1]}} = conv(padded(dZ^{[2]}), 180^o\, rotated\, filter\, K^{[2]})$$

$$C^{[1]} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad P^{[1]} = [2.5]$$

1/4 2×2 4×4

$$\frac{\partial L}{\partial C^{[1]}} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \qquad \frac{\partial L}{\partial P^{[1]}} = [2]$$

$$dC^{[1]}_{mn} = \frac{\partial L}{\partial C^{[1]}_{mn}} = \left\{ \frac{1}{4} * \frac{\partial L}{\partial P_{xy}^{[2]}} \quad \text{for } x = \text{floor}(m/2),\ y = \text{floor}(n/2) \right.$$

$$\frac{\partial L}{\partial Z^{[1]}} = \frac{\partial L}{\partial C^{[1]}} * \frac{\partial C^{[1]}}{\partial Z^{[1]}}$$

$$\frac{\partial C^{[1]}}{\partial Z_{mn}^{[1]}} = \begin{cases} 1 & \text{, If } Z_{mn} > 0 \\ 0 & \text{, If } Z_{mn} < 0 \end{cases}$$

$$dZ^{[1]} = \frac{\partial L}{\partial Z^{[1]}} = \frac{\partial L}{\partial C^{[1]}} * \frac{\partial C^{[1]}}{\partial Z^{[2]}}$$

$$dK^{[1]} = \frac{\partial L}{\partial K^{[1]}} = conv(Inp, dZ^{[1]})$$

$$dB^{[1]} = \frac{\partial L}{\partial B^{[1]}} = sum(dZ^{[1]})$$

Backprop for Layer 3

$$dZ^{[3]} = \frac{\partial L}{\partial Z^{[3]}} = (y\_pred - y)$$

$$dW^{[3]} = \frac{\partial L}{\partial W^{[3]}} = dZ^{[3]}) * f$$

$$dB^{[3]} = \frac{\partial L}{\partial B^{[3]}} = dZ^{[3]}$$

$$df = \frac{\partial L}{\partial f} = W^{[3]T} * dZ^{[3]}$$

ANN

Backprop for Layer 2

$$dP^{[2]} = \frac{\partial L}{\partial P^{[2]}} = df.reshape(P^{[2]}.shape)$$

$$dC_{mn}^{[2]} = \frac{\partial L}{\partial C_{mn}^{[2]}} = \begin{cases} \frac{\partial L}{\partial P_{xy}^{[2]}} & \text{, If } C_{mn} \text{ is the max element} \\ 0 & \text{, otherwise} \end{cases}$$

$$\frac{\partial C_{mn}^{[2]}}{\partial Z_{mn}^{[2]}} = \begin{cases} 1 & \text{, If } Z_{mn} > 0 \\ 0 & \text{, If } Z_{mn} < 0 \end{cases}$$

$$dZ^{[2]} = \frac{\partial L}{\partial Z^{[2]}} = \frac{\partial L}{\partial C^{[2]}} * \frac{\partial C^{[2]}}{\partial Z^{[2]}}$$

$$dK^{[2]} = \frac{\partial L}{\partial K^{[2]}} = conv(P^{[1]}, dZ^{[2]})$$

$$dB^{[2]} = \frac{\partial L}{\partial B^{[2]}} = sum(dZ^{[2]})$$

Backprop for Layer 1

$$dC_{mn}^{[1]} = \frac{\partial L}{\partial C_{mn}^{[1]}} = \left\{ \frac{1}{4} * \frac{\partial L}{\partial P_{xy}^{[2]}} \right. \quad \text{for x=floor(m/2), y = floor(n/2)}$$

$$\frac{\partial C_{mn}^{[1]}}{\partial Z_{mn}^{[1]}} = \begin{cases} 1 & \text{, If } Z_{mn} > 0 \\ 0 & \text{, If } Z_{mn} < 0 \end{cases}$$

$$dZ^{[1]} = \frac{\partial L}{\partial Z^{[1]}} = \frac{\partial L}{\partial C^{[1]}} * \frac{\partial C^{[1]}}{\partial Z^{[2]}}$$

$$dK^{[1]} = \frac{\partial L}{\partial K^{[1]}} = conv(Inp, dZ^{[1]})$$

$$dB^{[1]} = \frac{\partial L}{\partial B^{[1]}} = sum(dZ^{[1]})$$

for (

Weight Updation

$$W_3 = W_3 - \alpha * \frac{\partial L}{\partial W_3}$$

$$B_3 = B_3 - \alpha * \frac{\partial L}{\partial B_3}$$

$$K_2 = K_2 - \alpha * \frac{\partial L}{\partial K_2}$$

$$B_2 = B_2 - \alpha * \frac{\partial L}{\partial B_2}$$

$$K_1 = K_1 - \alpha * \frac{\partial L}{\partial K_1}$$

$$B_1 = B_1 - \alpha * \frac{\partial L}{\partial B_1}$$