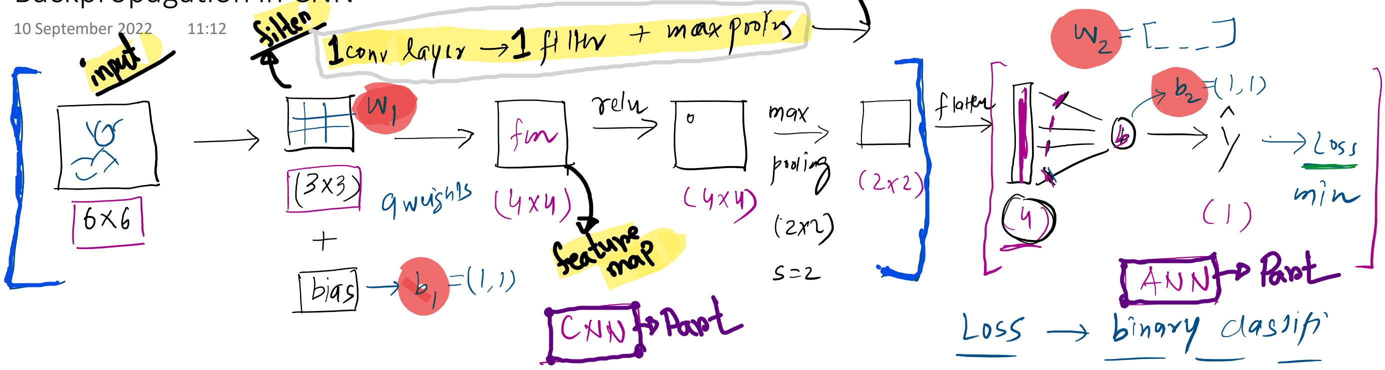
CNN Architecture — **Lec-47** — **Backpropagation in CNN-1**

Backpropagation in CNN

10 September 2022 11:12



Trainable Parameters

$W_1 = (3,3)$ $\quad W_2 = (1,4)$

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

= 15 trainable parameters

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

$A_2 = \hat{y}$

(Binary cross entrophy Loss function Lec(14))

Logical Flow

Forward Prop

$$Z_1 = \text{conv}(X, W_1) + b_1$$
$$A_1 = \text{relu}(Z_1)$$
$$P_1 = \text{maxpool}(A_1)$$

$$F = \text{flatten}(P_1)$$
$$Z_2 = W_2 F + b_2$$
$$A_2 = \sigma(Z_2)$$

Gradient Descent

$$W_1 = W_1 - \eta \frac{\partial L}{\partial W_1} \qquad W_2 = W_2 - \eta \frac{\partial L}{\partial W_2}$$

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

Loss is minimized

$$\boxed{\frac{\partial L}{\partial W_2}}$$

Logical flow:

$$X \rightarrow \circledast \rightarrow Z_1 \xrightarrow{\text{relu}} A_1 \xrightarrow[\text{pool}]{\text{max}} P_1 \xrightarrow{\text{flatten}} F \xrightarrow{W_2, b_2} Z_2 \xrightarrow{\text{sigmoid}} A_2(\hat{y}) \rightarrow L$$

W_1, b_1

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

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

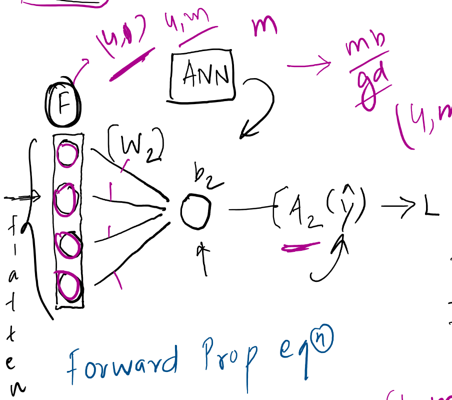
Backprop

(পরের ভিডিওতে দেখবো)

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

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

Backprop
- → Convolution
- → Flatten
- → Max pooling



→ mb gd (4,m)

$$\boxed{\frac{\partial L}{\partial w_2} = \boxed{\frac{\partial L}{\partial A_2}} \frac{\partial A_2}{\partial Z_2} \frac{\partial Z_2}{\partial w_2}} \qquad \boxed{\frac{\partial L}{\partial b_2} = \boxed{\frac{\partial L}{\partial A_2}} \frac{\partial A_2}{\partial Z_2} \frac{\partial Z_2}{\partial b_2}}$$

image কে Scalar(a) ধরো।

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

Forward Prop eqⁿ

$$\begin{cases} Z_2 = W_2 F + b_2 \\ A_2 = \sigma(Z_2) \end{cases}$$

(1,m)

$Y = (1,m)$

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

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

Differentiation of sigmoid:

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

W_2 update

Shape =

$$\boxed{\frac{\partial Z_2}{\partial W_2} = F} \qquad \frac{\partial Z_2}{\partial b_2} = 1$$

(1,4)

$$\boxed{\frac{\partial L}{\partial W_2}} = \frac{(a_2 - y_i)}{\cancel{a_2(1-a_2)}} \times \cancel{a_2(1-a_2)} \times F = (a_2 - y_i) F = (A_2 - Y) F^T$$

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

(1,1) (1,4)

(1,4)

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

m images

$$\boxed{\boxed{\frac{\partial L}{\partial W_2}} = (A_2 - Y) F^T} \qquad \boxed{\frac{\partial L}{\partial b_2} = (A_2 - Y)}$$

matrix multiplication এর জন্য Transpose করা লাগছে।

$\boxed{\frac{\partial L}{\partial w_2} = (A_2 - Y)F^T}$ $\boxed{\frac{\partial L}{\partial b_2} = (A_2 - Y)}$

$(1,m) - (1,m)$

$(1,m)\ (m,4) \rightarrow \boxed{(1,4)}$

1 image

batch of image

$W_2 \rightarrow (1,4)$