Convolutional NN

Date: __/__/___

✦ Cross - Correlation

$1.1+2.6+-1.5+0.3=8$

$$\underbrace{\begin{bmatrix}1 & 6 & 2\\ 5 & 3 & 1\\ 7 & 0 & 4\end{bmatrix}}_{\substack{\text{input}\\ n\times n}} \times \underbrace{\begin{bmatrix}1 & 2\\ -1 & 0\end{bmatrix}}_{\substack{\text{Kernel}\\ k\times k}} = \underbrace{\begin{bmatrix}8 & 7\\ 4 & 5\end{bmatrix}}_{\substack{\text{output}\\ (n-k+1)}}$$

k

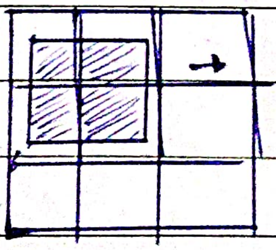
✦ Convolutional :

$$\text{Input} \times \underbrace{\begin{bmatrix}0 & -1\\ 2 & 1\end{bmatrix}}_{\text{rot180(Kernel)}} = \underbrace{\begin{bmatrix}7 & 5\\ 11 & 3\end{bmatrix}}_{\text{output}}$$

→ $\text{conv}(I,K) = I * \text{rot}180(K)$

→ $\underbrace{I * K}_{\text{Convolution}} = \underbrace{I \star \text{rot}180(K)}_{\text{Cross-Correlation}}$

- The convolution above which consists of the kernel sliding over the input and performing it's respective arithmetic operations is called "Valid" convolution
- The kernel has to fit in input.



- "full" convolution:

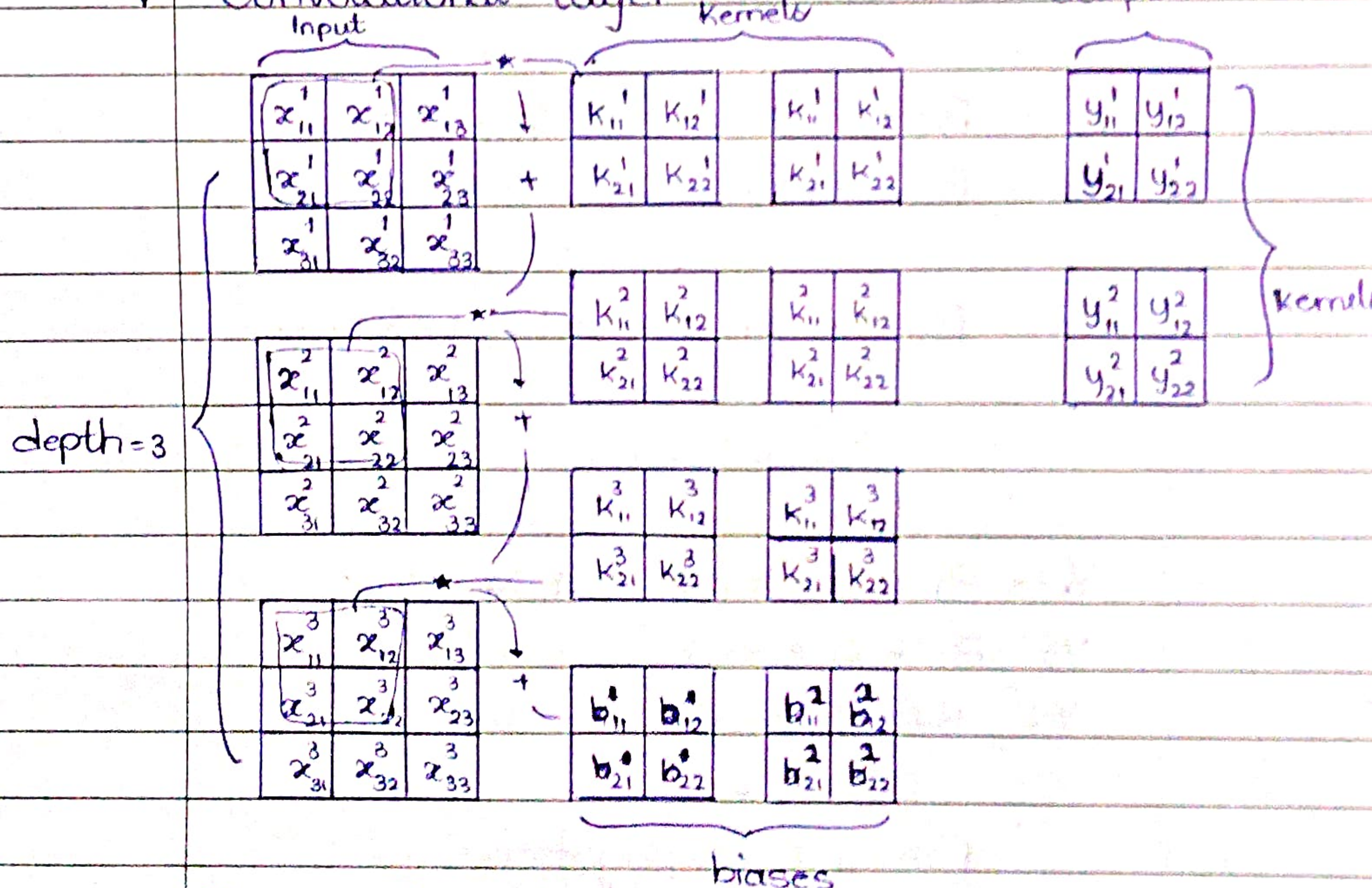
$-1.1+0.6 = -1$

$0.1 = 0$

$-1.6+0.2 = -6$

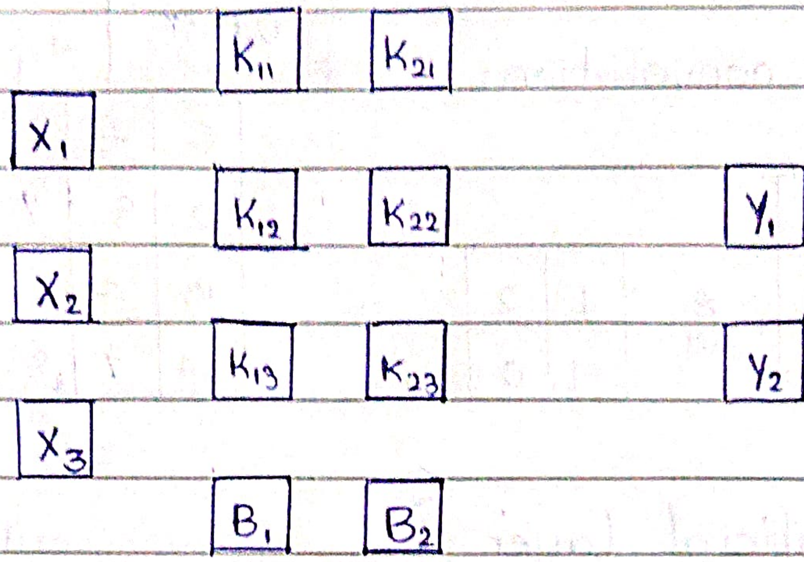
$$K \rightarrow \quad \begin{bmatrix}1&6&2\\5&3&1\\7&0&4\end{bmatrix} \underset{full}{\star} \begin{bmatrix}1&2\\-1&0\end{bmatrix} = \begin{bmatrix}0&-1&-6&-2\\2&8&7&1\\10&4&5&-3\\14&7&8&4\end{bmatrix}$$

✦ Convolutional layer



- Take each matrix in the first kernel and compute cross correlation with the input data. Sum up the three results and add first bias and this will produce the first output and repeat for each

Date: __/__/____



$$Y_1 = B_1 + X_1 * K_{11} + X_2 * K_{12} + X_3 * K_{13}$$

$$Y_2 = B_2 + X_1 * K_{21} + X_2 * K_{22} + X_3 * K_{23}$$

$$\vdots$$

$$Y_d = B_d + X_1 * K_{d1} + X_2 * K_{d2} + X_3 * K_{d3}$$

Generalised

$$Y_1 = B_1 + X_1 * K_{11} + \dots + X_n * K_{1n}$$

$$Y_2 = B_2 + X_1 * K_{21} + \dots + X_n * K_{2n}$$

$$\vdots$$

$$Y_d = B_d + X_1 * K_{d1} + \dots + X_n * K_{dn}$$

$\downarrow$

Forward Propagation

$$Y_i = B_i + \sum_{j=1}^{n} X_j * K_{ij}, \quad i = 1 \dots d$$

$n \rightarrow$ no. of inputs (features)
$j \rightarrow$ input index
$i \rightarrow$ output neuron index
$d \rightarrow$ no. of output neurons [More like colour channels) Input depth

$\frac{\partial E}{\partial Y_i} \longrightarrow \frac{\partial E}{\partial K_{ij}}, \frac{\partial E}{\partial B_i}$ → need them to update errors and biases

$\searrow \frac{\partial E}{\partial X_j}$ → need this to verify

→ $\partial E / \partial K_{ij}$

$$Y_i = B_i + \sum_{j=1}^{n} X_j \star K_{ij} \quad \text{[Forward pass]}$$

$$Y_i = B_i + X_1 \star K_{i1} + \dots + X_n \star K_{in}$$

$$Y_i = B_i + X_1 \star K_{i1} \quad \text{[Consideration]}$$

$$y_{11} = b_{11} + K_{11} x_{11} + K_{12} x_{12} + K_{21} x_{21} + K_{22} x_{22}$$

$$y_{12} = b_{12} + K_{11} x_{12} + K_{12} x_{13} + K_{21} x_{22} + K_{22} x_{23}$$

. . . .

✦ Chain rule

$$\frac{\partial E}{\partial K_{11}} = \frac{\partial E}{\partial y_{11}} \underbrace{\frac{\partial y_{11}}{\partial K_{11}}} + \frac{\partial E}{\partial y_{12}} \underbrace{\frac{\partial y_{12}}{\partial K_{11}}} + \frac{\partial E}{\partial y_{21}} \underbrace{\frac{\partial y_{21}}{\partial K_{11}}} + \frac{\partial E}{\partial y_{22}} \underbrace{\frac{\partial y_{22}}{\partial K_{11}}}$$

$$\frac{\partial E}{\partial K_{11}} = \frac{\partial E}{\partial y_{11}} x_{11} + \frac{\partial E}{\partial y_{12}} x_{12} + \frac{\partial E}{\partial y_{21}} x_{21} + \frac{\partial E}{\partial y_{22}} x_{22}$$

lly $$\frac{\partial E}{\partial K_{12}} = \frac{\partial E}{\partial y_{11}} x_{12} + \frac{\partial E}{\partial y_{12}} x_{13} + \frac{\partial E}{\partial y_{21}} x_{22} + \frac{\partial E}{\partial y_{22}} x_{23}$$

$$\frac{\partial E}{\partial K_{21}} = \frac{\partial E}{\partial y_{11}} x_{21} + \frac{\partial E}{\partial y_{12}} x_{22} + \frac{\partial E}{\partial y_{21}} x_{31} + \frac{\partial E}{\partial y_{22}} x_{32}$$

$$\frac{\partial E}{\partial K_{22}} = \frac{\partial E}{\partial y_{11}} x_{22} + \frac{\partial E}{\partial y_{12}} x_{23} + \frac{\partial E}{\partial y_{21}} x_{32} + \frac{\partial E}{\partial y_{22}} x_{33}$$

$$\Rightarrow \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \star \begin{bmatrix} \partial E/\partial y_{11} & \partial E/\partial y_{12} \\ \partial E/\partial y_{21} & \partial E/\partial y_{22} \end{bmatrix} \Rightarrow \frac{\partial E}{\partial K} = X \star \frac{\partial E}{\partial Y}$$

$Y = B + X \star K$

$\rightarrow$ ~~Y~~ + $\frac{\partial E}{\partial K} = X \star \frac{\partial E}{\partial Y}$

Can we differentiate a scalar with a matrix?

No, we can't, so ~~so~~ let's try to do ~~to do~~ with an Kernel ~~element~~ element.

$$\frac{\partial E}{\partial K_{21}} = \frac{\partial E}{\partial Y_1}\frac{\partial Y_1}{\partial K_{21}} + \frac{\partial E}{\partial Y_2}\frac{\partial Y_2}{\partial K_{21}} + \dots + \frac{\partial E}{\partial Y_d}\frac{\partial Y_d}{\partial K_{21}}$$

Looks fine, right? Noooooooo.......

We can't differentiate a matrix(Y) with a matrix.

Consider, $Y_2 = B_2 + X_1 \star K_{21} + \dots + X_n \star K_{2n}$

$$\frac{\partial E}{\partial K_{21}} = X_1 \star \frac{\partial E}{\partial Y_2}$$

$\longrightarrow$ Generalized

$$\frac{\partial E}{\partial K_{ij}} = X_j \star \frac{\partial E}{\partial Y_i}$$

$\rightarrow \frac{\partial E}{\partial B_i}$ [Take reference to prev one]

$Y_i = B_i + X_1 \star K_{i1}$

$$\frac{\partial E}{\partial b_{11}} = \frac{\partial E}{\partial y_{11}} ; \quad \frac{\partial E}{\partial b_{12}} = \frac{\partial E}{\partial y_{12}} ; \quad \frac{\partial E}{\partial b_{21}} = \frac{\partial E}{\partial y_{21}}$$

$$\frac{\partial E}{\partial b_{22}} = \frac{\partial E}{\partial y_{22}}$$

$$\rightarrow \frac{\partial E}{B_1} = \frac{\partial E}{\partial Y_1}$$

$$\rightarrow \frac{\partial E}{B_i} = \frac{\partial E}{\partial Y_i}$$

$$+ \quad \frac{\partial E}{\partial x_{11}} = \frac{\partial E}{\partial y_{11}}\frac{\partial y_{11}}{\partial x_{11}} + \frac{\partial E}{\partial y_{12}}\frac{\partial y_{12}}{\partial x_{11}} + \frac{\partial E}{\partial y_{21}}\frac{\partial y_{21}}{\partial x_{11}} + \frac{\partial E}{\partial y_{22}}\frac{\partial y_{22}}{\partial x_{11}}$$

$$\frac{\partial E}{\partial x_{11}} = \frac{\partial E}{\partial y_{11}} K_{11}$$

$$\frac{\partial E}{\partial x_{12}} = \frac{\partial E}{\partial y_{11}} K_{12} + \frac{\partial E}{\partial y_{12}} K_{11}$$

$$\frac{\partial E}{\partial x_{13}} = \frac{\partial E}{\partial y_{12}} K_{12}$$

$$\frac{\partial E}{\partial x_{21}} = \frac{\partial E}{\partial y_{11}} K_{21} + \frac{\partial E}{\partial y_{21}} K_{11}$$

$$\frac{\partial E}{\partial x_{22}} = \frac{\partial E}{\partial y_{11}} K_{22} + \frac{\partial E}{\partial y_{12}} K_{21} + \frac{\partial E}{\partial y_{21}} K_{12} + \frac{\partial E}{\partial y_{22}} K_{11}$$

$$\frac{\partial E}{\partial x_{23}} = \frac{\partial E}{\partial y_{12}} K_{22} + \frac{\partial E}{\partial y_{22}} K_{12}$$

$$\Rightarrow \begin{bmatrix} \frac{\partial E}{\partial y_{11}} & \frac{\partial E}{\partial y_{12}} \\ \frac{\partial E}{\partial y_{21}} & \frac{\partial E}{\partial y_{22}} \end{bmatrix} \underset{full}{\star} \underbrace{\overbrace{\begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}}^{rot180}}$$

$$\frac{\partial E}{\partial x_{31}} = \frac{\partial E}{\partial y_{21}} K_{21}$$

$$\frac{\partial E}{\partial x_{32}} = \frac{\partial E}{\partial y_{21}} K_{22} + \frac{\partial E}{\partial y_{22}} K_{21}$$

$$\frac{\partial E}{\partial y_{ij}} \underset{full}{\star} rot180([K_{ij}])$$

$$\frac{\partial E}{\partial x_{33}} = \frac{\partial E}{\partial y_{22}} K_{22}$$

$$\Rightarrow \frac{\partial E}{\partial X} = \frac{\partial E}{\partial Y} \underset{full}{\star} rot180(K)$$

Date: __/__/____

Convolution [check introductory pg]

$$\Rightarrow \frac{\partial E}{\partial x} = \frac{\partial E}{\partial y} \underset{\text{full}}{*} K$$

$$\Rightarrow \frac{\partial E}{\partial x_j} = \sum_{i=1}^{d} \frac{\partial E}{\partial y_i} \underset{\text{full}}{*} K_{ij}, \; j = 1, \dots, d$$

Backpropagation

- $\frac{\partial E}{\partial K_{ij}} = X_j \star \frac{\partial E}{\partial y_i}$
- $\frac{\partial E}{\partial B_i} = \frac{\partial E}{\partial y_i}$
- $\frac{\partial E}{\partial x_j} = \sum_{i=1}^{n} \frac{\partial E}{\partial y_i} \underset{\text{full}}{*} K_{ij}$

Binary Entropy Loss :

$$\underset{\text{(pred)}}{y^*} = \begin{bmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_i^* \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \end{bmatrix}$$

$$E = -\frac{1}{n} \sum_{i=1}^{n} y_i^* \log(y_i) + (1 - y_i^*) \log(1 - y_i)$$

$\frac{\partial E}{\partial Y}$ => Need to be passed to the last layer of NN

$$\frac{\partial E}{\partial y_1} = \frac{\partial}{\partial y_1}\left(\frac{-1}{n}\sum_{i=1}^{n} y_i^* \log(y_i) + (1-y_i^*)\log(1-y_i)\right)$$

$$= \frac{\partial}{\partial y_1}\left(\frac{-1}{n}\left(y_1^* \log(y_1) + (1-y_1^*)\log(1-y_1)\right)\right)$$

$$= \frac{-1}{n}\left(\frac{y_1^*}{y_1} - \frac{1-y_1^*}{1-y_1}\right)$$

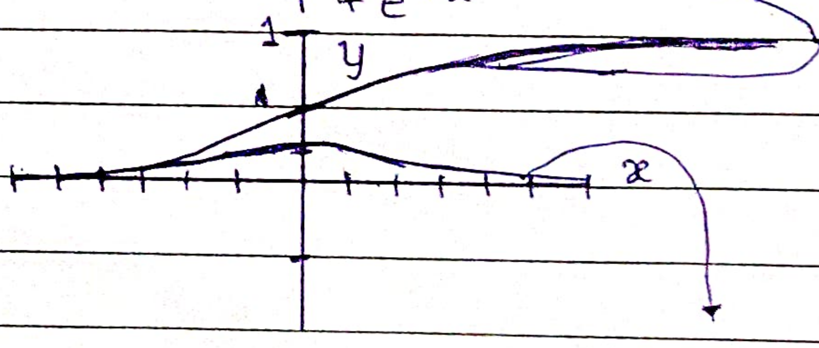
$$= \frac{1}{n}\left(\frac{1-y_1^*}{1-y_1} - \frac{y_1^*}{y_1}\right)$$

→ Generalized

$$\frac{\partial E}{\partial y_i} = \frac{1}{n}\left(\frac{1-y_i^*}{1-y_i} - \frac{y_i^*}{y_i}\right)$$

Sigmoid activation:

$$\sigma(x) = \frac{1}{1+e^{-x}} \qquad 0 \le \sigma(x) \le 1$$



$$\sigma'(x) = \frac{e^{-x}}{(1+e^{-x})^2} = \sigma(x)\left(1-\sigma(x)\right)$$