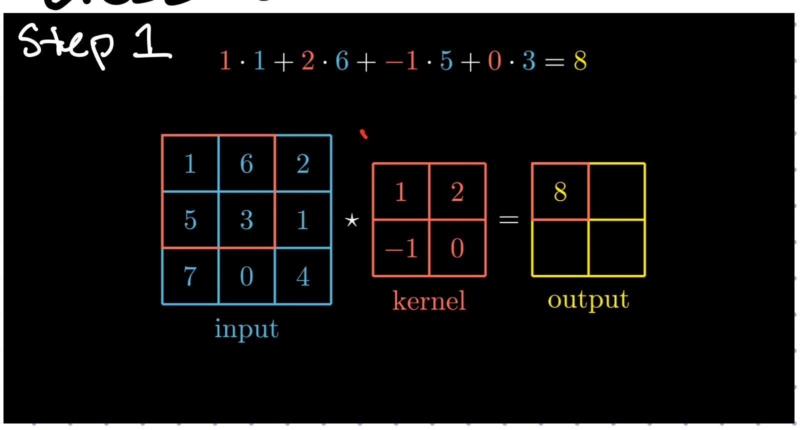
Overview:

1. Convolution operation
2. Convolutional layer
3. Reshape Layer
4. Binary Cross Entropy Loss
5. Sigmoid Activation

Convolution walk through

Cross-correlation



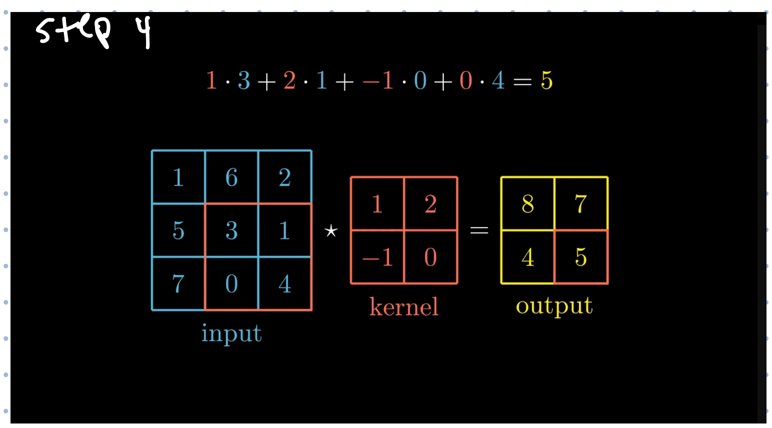
$$\begin{bmatrix}1 & 6\\ 5 & 3\end{bmatrix} \times \begin{bmatrix}1 & 2\\ -1 & 0\end{bmatrix}$$

$=$

$1 \cdot 1 + 6 \cdot 2 + 5 \cdot -1 + 3 \cdot 0$

$1 + 12 + -5 + 0$

8



$$\begin{bmatrix}3 & 1\\ 0 & 4\end{bmatrix} \times \begin{bmatrix}1 & 2\\ -1 & 0\end{bmatrix}$$

$=$

$3 \cdot 1 + 1 \cdot 2 + 0 \cdot -1 + 4 \cdot 0$

$3 + 2 + 0 + 0$

5

Size of Output

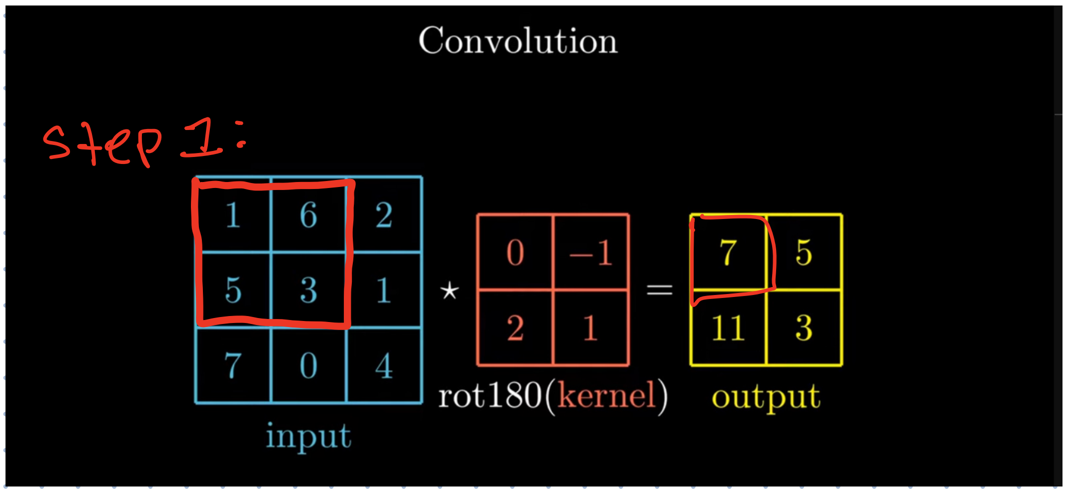
$Y = I - K + 1$

size of output is the size of input subtracted by size of the output plus 1

Example:

$2 = 3 - 2 + 1$

Convolution



step 1: $\begin{bmatrix} 1 & 6 \\ 5 & 3 \end{bmatrix} \times \begin{bmatrix} 0 & -1 \\ 2 & 1 \end{bmatrix} = 1\cdot 0 + 6\cdot -1 + 5\cdot 2 + 1\cdot 3$

$0 + -6 + 10 + 3$

$\boxed{7}$

Cross-Correlation $\neq$ Convolution

Math Notation

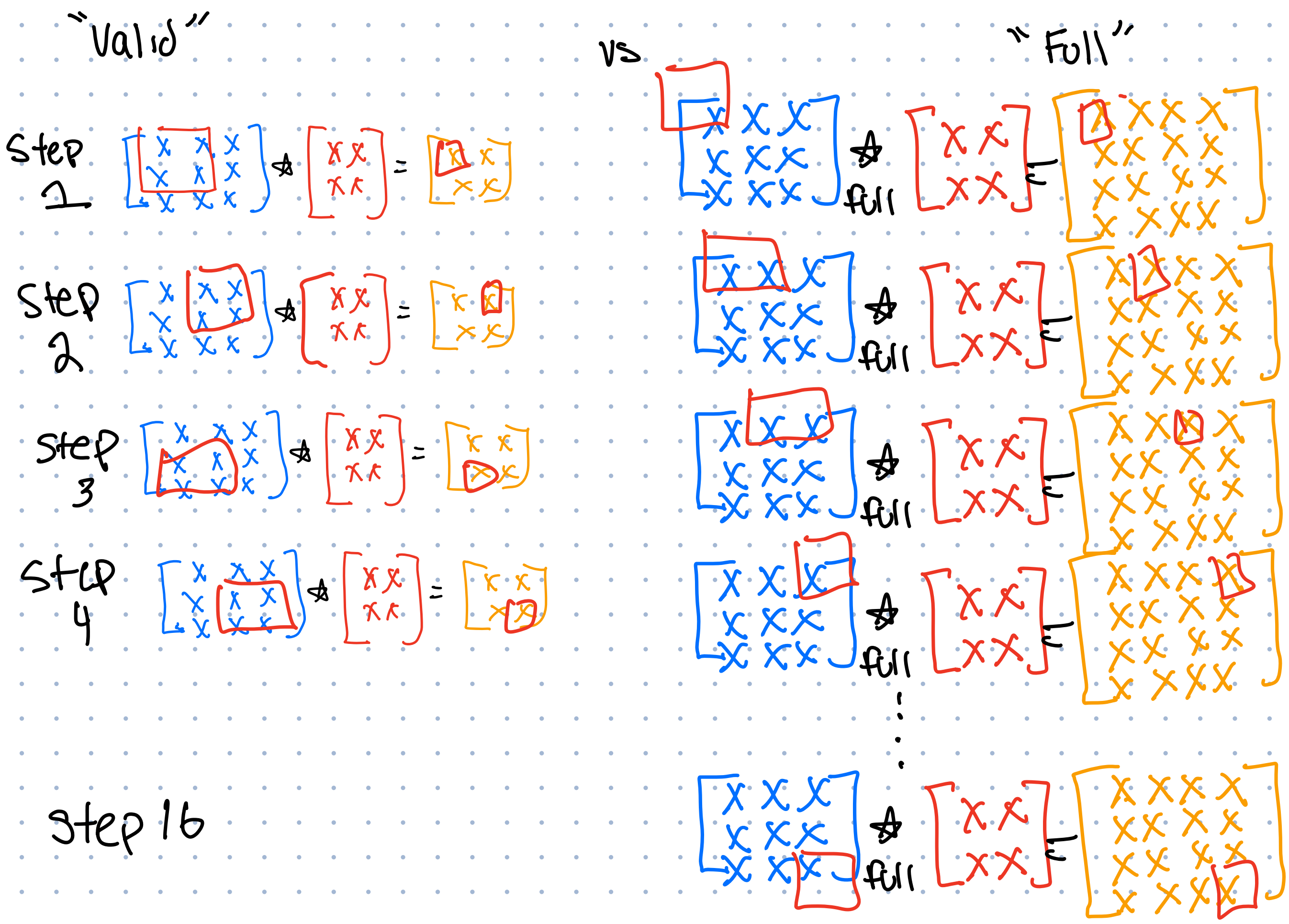
I = Input matrix
K = Kernal matrix
O = output matrix

$$\mathrm{Conv}(I, K) = I \star \mathrm{rot180}(K)$$

$$I * K = I \star \mathrm{rot180}(K)$$

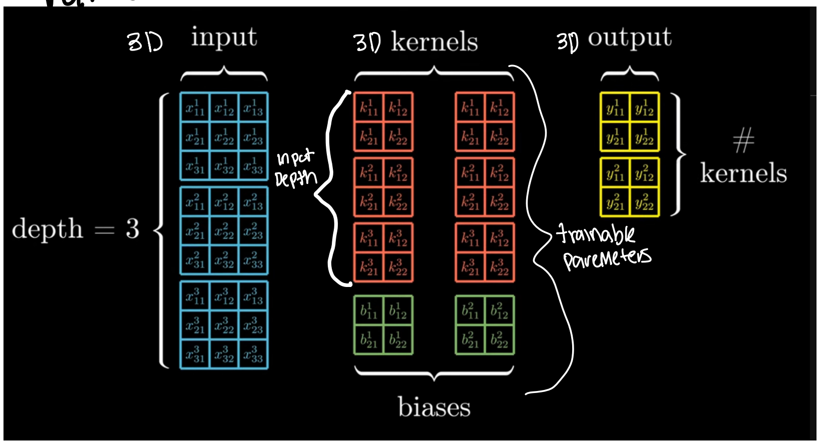
Convolution

Cross-Correlation

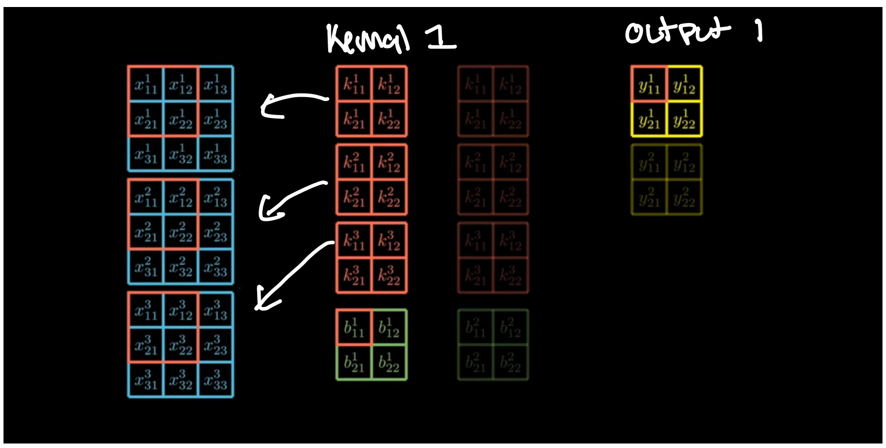


Convolutional Layer

Parts :



output calculation



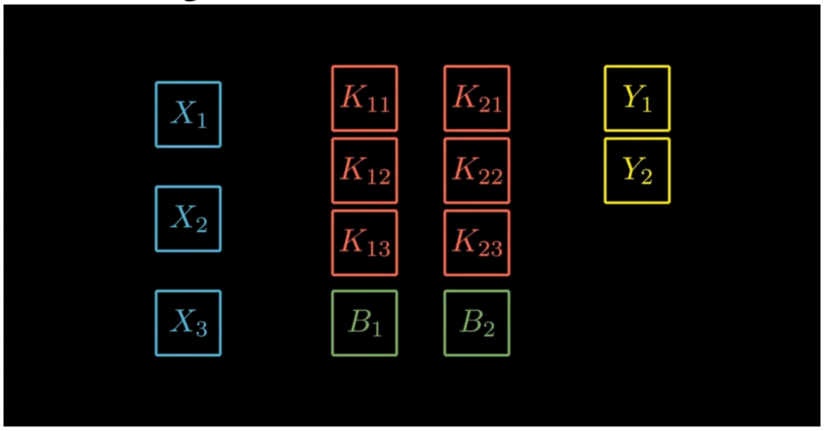
$$\begin{bmatrix} X^1_{11} & X^1_{12} \\ X^1_{21} & X^1_{22} \end{bmatrix} \star \begin{bmatrix} K^1_{11} & K^1_{12} \\ K^1_{21} & K^1_{22} \end{bmatrix} = O_1$$

$$\begin{bmatrix} X^2_{11} & X^2_{12} \\ X^2_{21} & X^2_{22} \end{bmatrix} \star \begin{bmatrix} K^1_{11} & K^1_{12} \\ K^1_{21} & K^1_{22} \end{bmatrix} = O_2$$

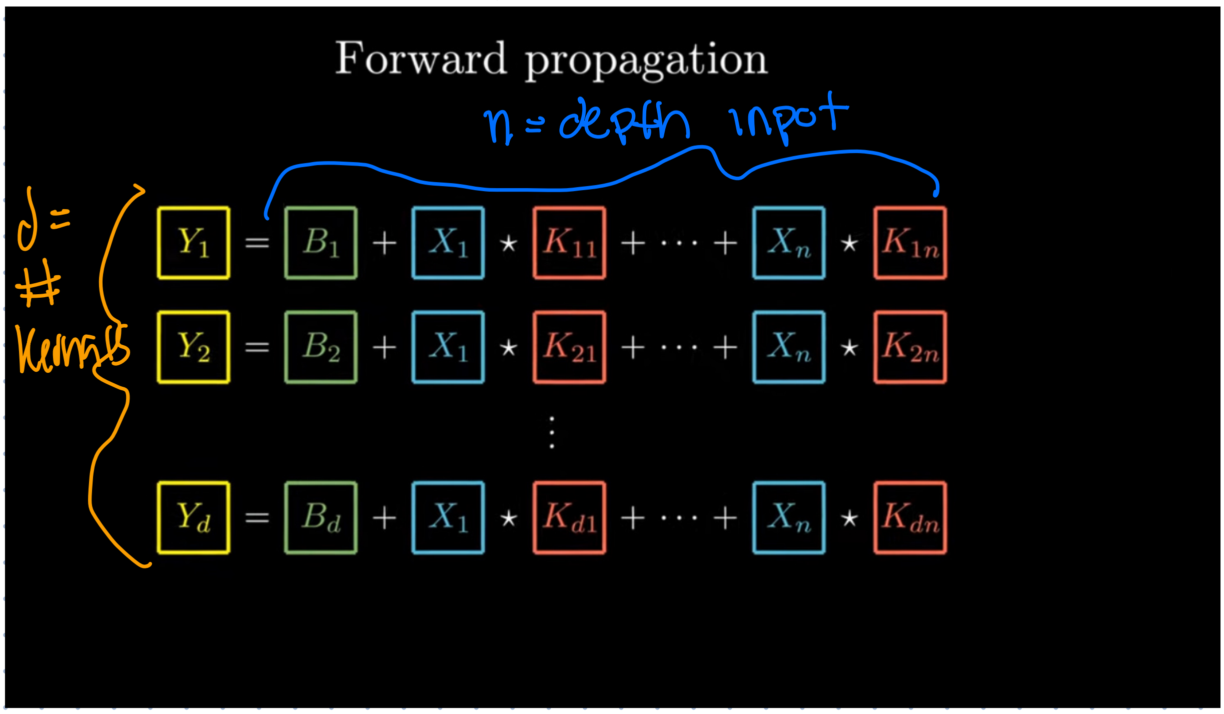
$$\begin{bmatrix} X^3_{11} & X^3_{12} \\ X^3_{21} & X^3_{22} \end{bmatrix} \star \begin{bmatrix} K^1_{11} & K^1_{12} \\ K^1_{21} & K^1_{22} \end{bmatrix} = O_3$$

$$O_1 + O_2 + O_3 + B = \boxed{Y^1_{11}}$$

Abstracting the matrixs :



Forward Propgration mode



Written in Summation form

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

$$\boxed{Y_1} = \boxed{B_1} + \boxed{X_1} \star \boxed{K_{11}} + \cdots + \boxed{X_n} \star \boxed{K_{1n}}$$

$$\boxed{Y_2} = \boxed{B_2} + \boxed{X_1} \star \boxed{K_{21}} + \cdots + \boxed{X_n} \star \boxed{K_{2n}}$$

$$\vdots$$

$$\boxed{Y_d} = \boxed{B_d} + \boxed{X_1} \star \boxed{K_{d1}} + \cdots + \boxed{X_n} \star \boxed{K_{dn}}$$

(A) Backward Propagation Overview

Given

Output of forward propagation

error

$$\frac{\partial E}{\partial Y_i}$$

To find

$$\frac{\partial E}{\partial k_{ij}}, \quad \frac{\partial E}{\partial B_i}$$

∂B_i — bias

∂k_{ij} — kernel

$$\frac{\partial E}{\partial X_j}$$

∂X_j → input

(1) $\frac{\partial E}{\partial k_{ij}}$: Looking at single term of foward propagation (Y_i) to generalize:

(A)

$$Y = B + X \star K$$

Expanded

$$\frac{\partial E}{\partial Y} = \begin{bmatrix} \frac{\partial E}{\partial y_{11}} & \frac{\partial E}{\partial y_{12}} \\ \frac{\partial E}{\partial y_{21}} & \frac{\partial E}{\partial y_{22}} \end{bmatrix} \longrightarrow \begin{bmatrix} \frac{\partial E}{\partial k_{11}} & \frac{\partial E}{\partial k_{12}} \\ \frac{\partial E}{\partial k_{21}} & \frac{\partial E}{\partial k_{22}} \end{bmatrix}$$

$$\begin{cases} y_{11} = b_{11} + k_{11}x_{11} + k_{12}x_{12} + k_{21}x_{21} + k_{22}x_{22} \\ y_{12} = b_{12} + k_{11}x_{12} + k_{12}x_{13} + k_{21}x_{22} + k_{22}x_{23} \\ y_{21} = b_{21} + k_{11}x_{21} + k_{12}x_{22} + k_{21}x_{31} + k_{22}x_{32} \\ y_{22} = b_{22} + k_{11}x_{22} + k_{12}x_{23} + k_{21}x_{32} + k_{22}x_{33} \end{cases}$$

To find:

expanded using chain rule

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

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

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

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

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

$$\because \frac{\partial}{\partial k_{11}}(y_{11}) = \frac{\partial}{\partial k_{11}}(k_{11}x_{11}) = x_{11}$$

$$\Rightarrow x_{11} = \frac{\partial y_{11}}{\partial k_{11}}$$

$\therefore$ we have $\dfrac{\partial E}{\partial K_{ij}} = X_j * \dfrac{\partial E}{\partial Y_i}$

② $\dfrac{\partial E}{\partial B_i}$: Looking at single term of forward propagation (Y_i) to generalize:

$$Y = B + X * K$$

$$\frac{\partial E}{\partial Y} = \begin{bmatrix} \frac{\partial E}{\partial y_{11}} & \frac{\partial E}{\partial y_{12}} \\ \frac{\partial E}{\partial y_{21}} & \frac{\partial E}{\partial y_{22}} \end{bmatrix} \longrightarrow \begin{bmatrix} \frac{\partial E}{\partial b_{11}} & \frac{\partial E}{\partial b_{12}} \\ \frac{\partial E}{\partial b_{21}} & \frac{\partial E}{\partial b_{22}} \end{bmatrix}$$

$$\begin{cases} y_{11} = b_{11} + k_{11}x_{11} + k_{12}x_{12} + k_{21}x_{21} + k_{22}x_{22} \\ y_{12} = b_{12} + k_{11}x_{12} + k_{12}x_{13} + k_{21}x_{22} + k_{22}x_{23} \\ y_{21} = b_{21} + k_{11}x_{21} + k_{12}x_{22} + k_{21}x_{31} + k_{22}x_{32} \\ y_{22} = b_{22} + k_{11}x_{22} + k_{12}x_{23} + k_{21}x_{32} + k_{22}x_{33} \end{cases}$$

$$\Downarrow$$

$$\frac{\partial y_{11}}{\partial b_{11}} = \frac{\partial\,(b_{11} + k \cdot x + \ldots)}{\partial b_{11}} = 1$$

$$\frac{\partial E}{\partial b_{11}} = \frac{\partial E}{\partial y_{11}} \underbrace{\frac{\partial y_{11}}{\partial b_{11}}}_{1} + \frac{\partial E}{\partial y_{12}} \underbrace{\frac{\partial y_{12}}{\partial b_{11}}}_{0} + \frac{\partial E}{\partial y_{21}} \underbrace{\frac{\partial y_{21}}{\partial b_{11}}}_{0} + \frac{\partial E}{\partial y_{22}} \underbrace{\frac{\partial y_{22}}{\partial b_{11}}}_{0}$$

$\therefore$ we have $\dfrac{\partial E}{\partial B_i} = \dfrac{\partial E}{\partial Y_i}$

③ $\dfrac{\partial E}{\partial X_i}$

$$\frac{\partial E}{\partial x_{11}} = \frac{\partial E}{\partial y_{11}} \overbrace{k_{11}}^{\frac{\partial y_{11}}{\partial x_{11}}} + \frac{\partial E}{\partial y_{12}} \cdot \overbrace{\frac{\partial y_{12}}{\partial x_{11}}}^{0} + \frac{\partial E}{\partial y_{21}} \cdot \overbrace{\frac{\partial y_{21}}{\partial x_{11}}}^{0} + \frac{\partial E}{\partial y_{22}} \overbrace{\frac{\partial y_{22}}{\partial x_{11}}}^{0}$$

$$\frac{\partial E}{\partial x_{12}} = \frac{\partial E}{\partial y_{11}} k_{12} + \frac{\partial E}{\partial y_{12}} k_{11} \qquad \vdots$$

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

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

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

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

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

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

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

$$\begin{cases} y_{11} = b_{11} + k_{11}x_{11} + k_{12}x_{12} + k_{21}x_{21} + k_{22}x_{22} \\ y_{12} = b_{12} + k_{11}x_{12} + k_{12}x_{13} + k_{21}x_{22} + k_{22}x_{23} \\ y_{21} = b_{21} + k_{11}x_{21} + k_{12}x_{22} + k_{21}x_{31} + k_{22}x_{32} \\ y_{22} = b_{22} + k_{11}x_{22} + k_{12}x_{23} + k_{21}x_{32} + k_{22}x_{33} \end{cases}$$

$$\Rightarrow \begin{bmatrix} \frac{\partial E}{\partial y_{11}} & \frac{\partial E}{\partial y_{12}} \\ \frac{\partial E}{\partial y_{21}} & \frac{\partial E}{\partial y_{22}} \end{bmatrix} \overset{\star}{\underset{full}{}} \begin{bmatrix} k_{22} & k_{21} \\ k_{12} & k_{11} \end{bmatrix}$$

$$\Rightarrow \frac{\partial E}{\partial X} = \frac{\partial E}{\partial Y} \underset{full}{*} rot180(K) \quad [\text{full cross-correlation}]$$

$$= \frac{\partial E}{\partial Y} \underset{full}{*} K \quad [\text{full-convolution}]$$

High Level Explanations of the rest of the Implementation

Note: these parts are not CNN specific

3) Reshape Layer

☆ has the functionality to reshape the data

↳ necessary b/c the output of the conv. is 3D

↳ we use dense layers typically @ end of network

☆ specifically, column vectors

→ uses numpy to reshape input → output shape & output gradient → input shape

4) Binary Cross Entropy Loss

↳ used for classification of only 2 digits of MNIST (we will be attempting a diff. version since we want to classify more than 2 digits if we can)

☆ uses log function

5) Sigmoid Activation

→ inputs bounded between Ø-1, since Bin. CEL. can't take negative inputs as it uses a log function

Cross Entropy Loss

$Y^* \rightarrow \begin{bmatrix} y_i^* \\ \downarrow \\ y_i^* \end{bmatrix}$ ✱ is the "right"/desired answer outputs

→ for binary, all y_i^* values can either be ∅ or 1 since we're only classifying 2

$Y \rightarrow \begin{bmatrix} y_i \\ \downarrow \\ y_i \end{bmatrix}$ ✱ is the actual outputs

✩ this is where we compute $\frac{dE}{dY}$ for the backward prop. of the last layer of neural network

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

$$\frac{\partial E}{\partial Y} = \begin{bmatrix} \frac{\partial E}{\partial y_1} \\ \frac{\partial E}{\partial y_2} \\ \vdots \\ \frac{\partial E}{\partial y_i} \end{bmatrix}$$

$$\begin{aligned}
\frac{\partial E}{\partial y_1} &= \frac{\partial}{\partial y_1}\left(-\frac{1}{n}\sum_{i=1}^{n} y_i^* log(y_i) + (1 - y_i^*)log(1 - y_i)\right) \\
&= \frac{\partial}{\partial y_1}\left(-\frac{1}{n}(y_1^* log(y_1) + (1 - y_1^*)log(1 - y_1)) - \underbrace{\cdots - \frac{1}{n}(y_n^* log(y_n) + (1 - y_n^*)log(1 - y_n))}\right) \\
&= \frac{\partial}{\partial y_1}\left(-\frac{1}{n}(y_1^* log(y_1) + (1 - y_1^*)log(1 - y_1))\right) \\
&= -\frac{1}{n}\left(\frac{y_1^*}{y_1} - \frac{1 - y_1^*}{1 - y_1}\right) \\
&= \frac{1}{n}\left(\frac{1 - y_1^*}{1 - y_1} - \frac{y_1^*}{y_1}\right)
\end{aligned}$$

here we derive by applying chain rule

all these terms disappear since they do not have y_1 terms, ∴ we will be performing a derivative

Sigmoid Activation

✲ gets applied multiple times throughout network

Sigmoid Function

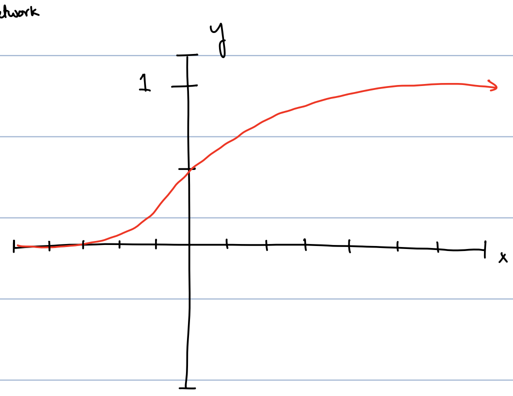
$$\sigma(x) = \frac{1}{1+e^{-x}}$$

Sigmoid Derivative

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

or

$$= \sigma(x)\left(1-\sigma(x)\right)$$



$\sigma(x)$

✲ as you can see, it is bounded between Ø & 1