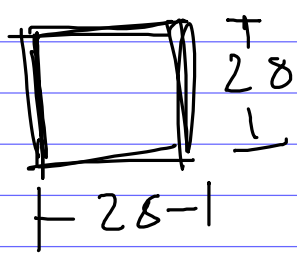
Neural Net Architecture

MNIST Images



784 Pixels (1 channel)

$$X = \begin{bmatrix} — & x_1 & — \\ — & x_2 & — \\ & \vdots & \\ — & x_m & — \end{bmatrix}$$

m x n

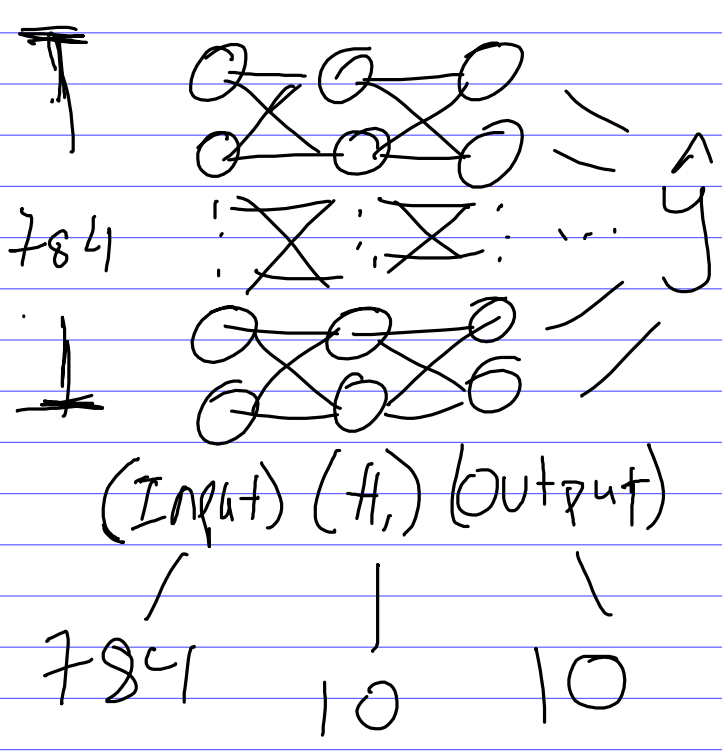
Samples (#)

Pixels (784)

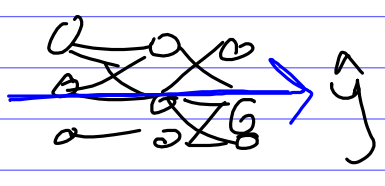
$$X = \begin{bmatrix} \cdots & x_1 & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & x_m & \cdots \end{bmatrix}$$

784+1 → +1 for labels

→ Transpose (X^T) for Graph



Forward Propagation



$A_0 = X^T$ (784 x m)

$\rightarrow Z_1 = W_1 A_0 + B_1$

(10 x m) (10 x 784)(784 x m) (10 x m)

$A_1 = g_1(Z_1) = ReLu(Z_1)$

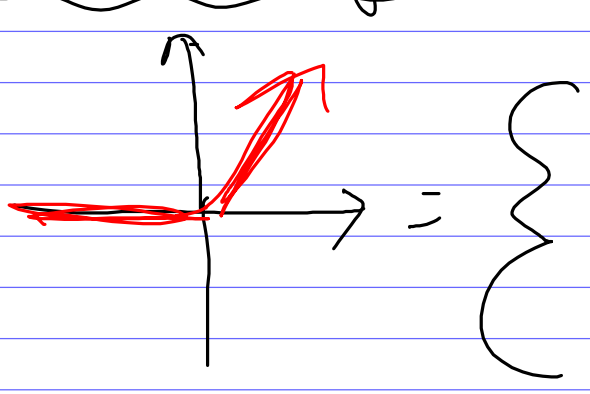
$\rightarrow Z_2 = W_2 A_1 + B_2$

(10 x m) (10 x 10)(10 x m) (10 x m)

$A_2 = g_2(Z_2) = SoftMax(Z_2)$

↳ Non-Linear (Rich Z_2)

ReLu Layer



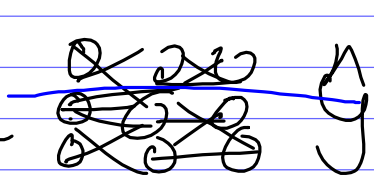
$= \begin{cases} 0 : X \leq 0 \\ X : X > 0 \end{cases}$

SoftMax Layer

classifications

$\begin{bmatrix} 1 \\ 2.4 \\ 3.1 \\ \vdots \end{bmatrix} \rightarrow \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \rightarrow \begin{bmatrix} 0.1 \\ .2 \\ .6 \\ \vdots \end{bmatrix}$

Back Propagation



Start w/ ŷ & move ← through NN

↳ Adjust W_is & LR

$dZ_2 = A_2 - Y$

$dW_2 = \frac{1}{m} dZ_2 A_1^T$

$dB_2 = \frac{1}{m} \sum dZ_2$

$dZ_1 = W_2^T dZ_2 \cdot g'(Z_1)$

$dW_1 = \frac{1}{m} dZ_1 X^T$

$dB_1 = \frac{1}{m} \sum dZ_1$

Updating Params

$W_1 := W_1 - \alpha dW_1$

$B_1 := B_1 - \alpha dB_1$

$W_2 := W_2 - \alpha dW_2$

$B_2 := B_2 - \alpha dB_2$

Overall

Forward P

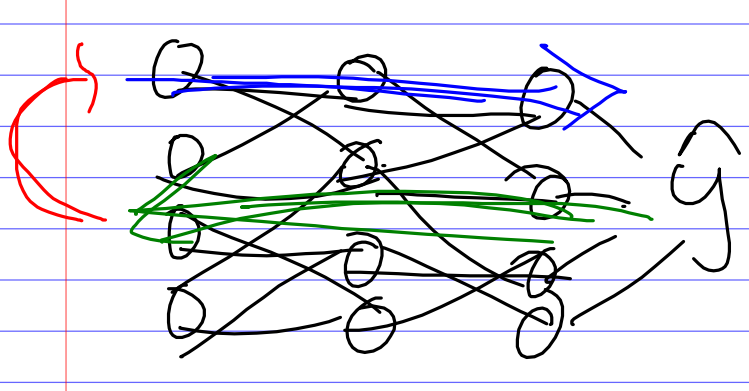
↓

Backward P

↓

Update

+1 Iterations



max Iter = N

or

Δerror ≤ TOL