MNIST From Scratch using MNIST

784 → 28×28

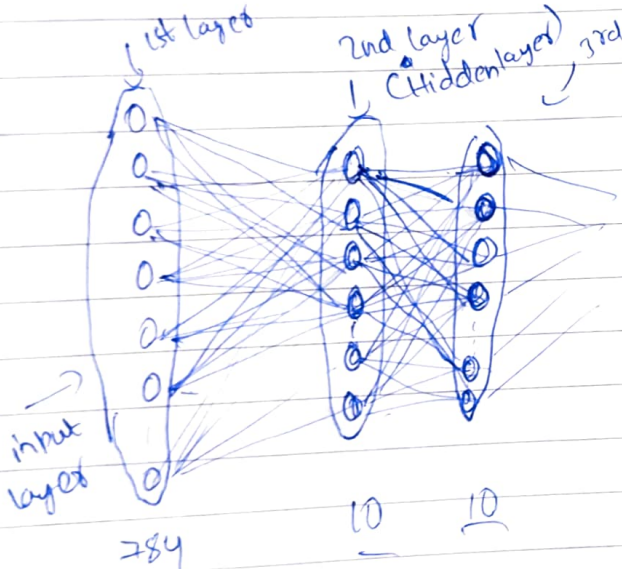
→ say we have m training images

→ each pixel is a value in [0, 1, ... 255]

0 ↑ completely Black

255 ↑ completely white

$$X = \begin{bmatrix} — & x^{(1)} & — \\ — & x^{(2)} & — \\ & \vdots & \\ — & x^{(m)} & — \end{bmatrix}^T = \begin{bmatrix} | & | & | & | & | \\ x^{(1)} & x^{(2)} & x^{(3)} & x^{(4)} & \cdots & x^{(m)} \\ | & | & | & | & | \end{bmatrix} \} 784 \text{ long}$$



input layer value

$A^{[0]} = X$

$Z^{[1]} = W^{[1]} A^{[0]} + b^{[1]}$

$10 \times m \quad 10 \times 784 \quad 784 \times m \quad 10 \times m$

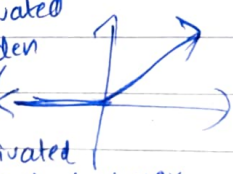
↑ unactivated

hidden layer values

Forward propagation

$A^{[1]} = ReLU(Z^{[1]})$

↑ activated hidden layer



$= \begin{cases} x \text{ if } x \geq 0 \\ 0 \text{ if } x < 0 \end{cases}$

unactivated ↑ output layer

$Z^{[2]} = W^{[2]} A^{[1]} + b^{[2]}$

activated output layer $A^{[2]} = \text{softmax}(Z^{[2]})$

Output layer

$\begin{bmatrix} 1.3 \\ 5.1 \\ 2.2 \\ 0.7 \\ 1.1 \end{bmatrix} \rightarrow \left[\frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \right] \rightarrow \begin{bmatrix} 0.02 \\ 0.90 \\ 0.05 \\ 0.01 \\ 0.02 \end{bmatrix}$

Softmax activation function

Back propagation

$dZ^{[2]} = A^{[2]} - Y$ (one hot encoded)

$10 \times m \quad 10 \times m \quad 10 \times m$

$dW^{[2]} = \frac{1}{m} dZ^{[2]} A^{[1]T}$

$10 \times 10 \quad 10 \times m \quad m \times 10$

$db^{[2]} = \frac{1}{m} \sum dZ^{[2]}$

$10 \times 1 \quad 10 \times 1$

$dZ^{[1]} = W^{[2]T} dZ^{[2]} .* g^{[1]'}(Z^{[1]})$

$10 \times m \quad 10 \times 10 \quad 10 \times m$

$dW^{[1]} = 1/m \; dZ^{[1]} A^{[0]T}$

$10 \times 784 \quad 10 \times m \quad m \times 784$

$db^{[1]} = \frac{1}{m} \sum dZ^{[1]}$

$10 \times 1 \quad 10 \times 1$

learning rate

$W^{[1]} := W^{[1]} - \alpha\, dW^{[1]}$

$b^{[1]} := b^{[1]} - \alpha\, db^{[1]}$

$W^{[2]} := W^{[2]} - \alpha\, dW^{[2]}$

$b^{[2]} := b^{[2]} - \alpha\, db^{[2]}$

} update params