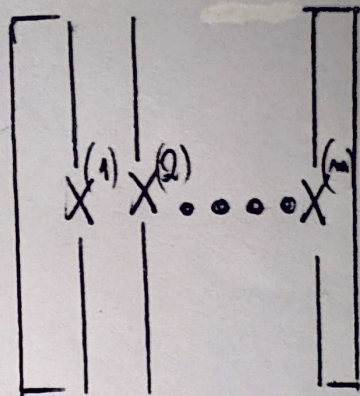
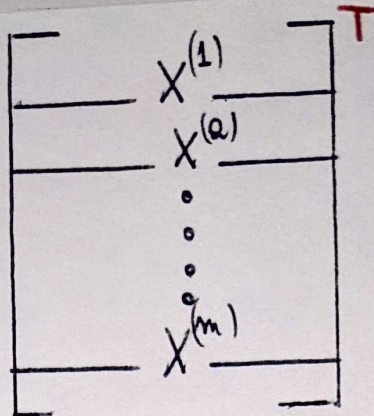


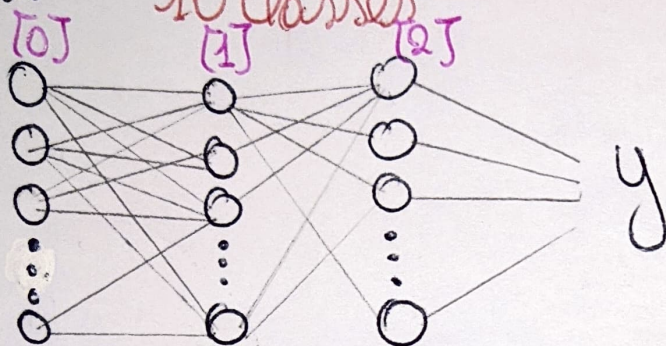
$\boxed{784}$   $m$  training Image  
 $(28 \times 28)$  pixel  $X =$



$\boxed{784} \Rightarrow 0, 1, 2, \dots, 9$

$28 \times 28$

10 classes



784

10

10

"input layer"

"first hidden layer"

"output layer"

Forward propagation:

$$A^{[0]} = X \quad (784 \times m)$$

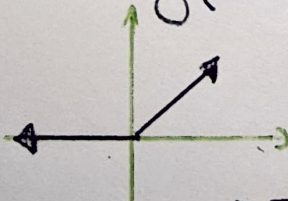
$$Z^{[1]} = W^{[1]} A^{[0]} + b^{[1]}$$

$10 \times m \quad 10 \times 784 \quad 784 \times m \quad 10 \times 1 \Rightarrow 10 \times m$

Activation Function:

Rectified Linear Unit

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

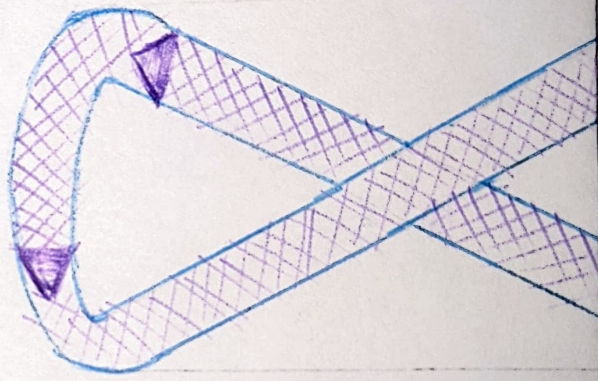


$$= \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

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

$10 \times m \quad 10 \times 10 \quad 10 \times m \quad 10 \times 1 \rightarrow 10 \times m$

$$A^{[2]} = \text{Softmax}(Z^{[2]})$$



*[Handwritten signature]*



Output layer

Softmax activation function

Probabilities

1.3  
5.1  
2.2  
0.7  
1.1

$$\frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

0.02  
0.09  
0.05  
0.01  
0.02

Backward propagation:

$$\begin{aligned} dz^{[2]} &= A^{[2]} - Y \\ dw^{[2]} &= \frac{1}{m} dz^{[2]} A^{[1]T} \\ db^{[2]} &= \frac{1}{m} \sum dz^{[2]} \end{aligned}$$

$$\begin{aligned} dz^{[1]} &= W^{[2]T} dz^{[2]} * g'(z^{[1]}) \\ dw^{[1]} &= \frac{1}{m} dz^{[1]} x^T \\ db^{[1]} &= \frac{1}{m} \sum dz^{[1]} \end{aligned}$$

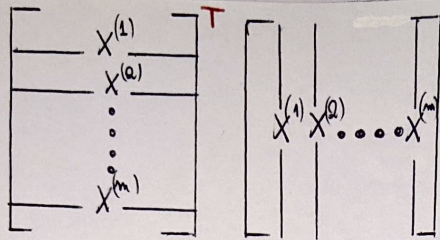
Parameter updates:

$$\begin{aligned} W^{[1]} &= W^{[1]} - \alpha dw^{[1]} \\ b^{[1]} &= b^{[1]} - \alpha db^{[1]} \\ W^{[2]} &= W^{[2]} - \alpha dw^{[2]} \\ b^{[2]} &= b^{[2]} - \alpha db^{[2]} \end{aligned}$$

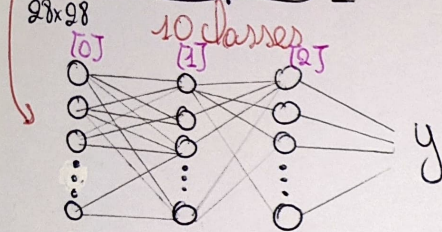
ossine



$\begin{bmatrix} 784 \\ 28 \times 28 \end{bmatrix}$  m training Image  $X =$



$\begin{bmatrix} 784 \\ 28 \times 28 \end{bmatrix} \Rightarrow 0, 1, 2, \dots, 9$



784  $\downarrow$  "input layer"  
10  $\downarrow$  "first hidden layer"  
10  $\downarrow$  "output layer"

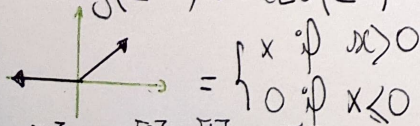
Forward propagation:

$$A^{[0]} = X \quad (784 \times m)$$

$$Z^{[1]} = W^{[1]} A^{[0]} + b^{[1]} \quad \begin{matrix} 10 \times m & 10 \times 784 & 784 \times m & 10 \times 1 \Rightarrow 10 \times m \end{matrix}$$

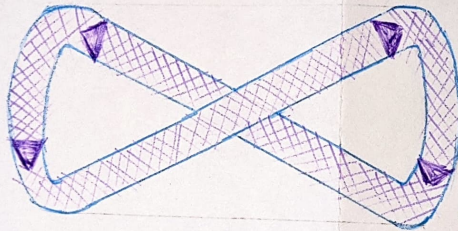
Activation Function: Rectified Linear Unit

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

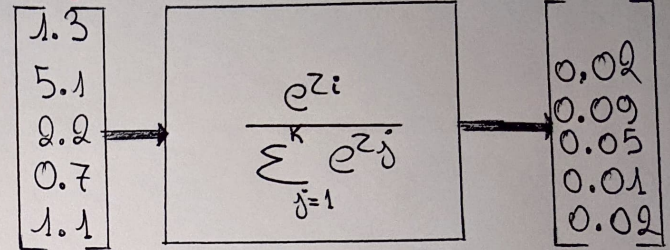


$$Z^{[2]} = W^{[2]} A^{[1]} + b^{[2]} \quad \begin{matrix} 10 \times m & 10 \times 10 & 10 \times m & 10 \times 1 \Rightarrow 10 \times m \end{matrix}$$

$$A^{[2]} = \text{Softmax}(Z^{[2]})$$



Output Layer      Softmax activation function      Probabilities



Backward propagation:

$$\begin{aligned} dz^{[2]} &= A^{[2]} - Y \\ dw^{[2]} &= \frac{1}{m} dz^{[2]} A^{[1]T} \\ db^{[2]} &= \frac{1}{m} \sum dz^{[2]} \end{aligned} \quad \left\{ \begin{aligned} dz^{[1]} &= W^{[2]T} dz^{[2]} \cdot g'(Z^{[1]}) \\ dw^{[1]} &= \frac{1}{m} dz^{[1]} X^T \\ db^{[1]} &= \frac{1}{m} \sum dz^{[1]} \end{aligned} \right.$$

Parameter updates:

$$\begin{aligned} W^{[1]} &= W^{[1]} - \alpha dw^{[1]} \\ b^{[1]} &= b^{[1]} - \alpha db^{[1]} \\ W^{[2]} &= W^{[2]} - \alpha dw^{[2]} \\ b^{[2]} &= b^{[2]} - \alpha db^{[2]} \end{aligned}$$

*Yasmine*