

Proyecto Concurrente Análisis

MNIST \rightarrow Set de análisis de imágenes

784
28x28

M entrenamiento
de imágenes

¡Importante realizar
un análisis con
10 neuronas
ocultas!

$$A = \begin{bmatrix} X_{(1)} \\ X_{(2)} \\ \vdots \\ X_{(m)} \end{bmatrix}$$

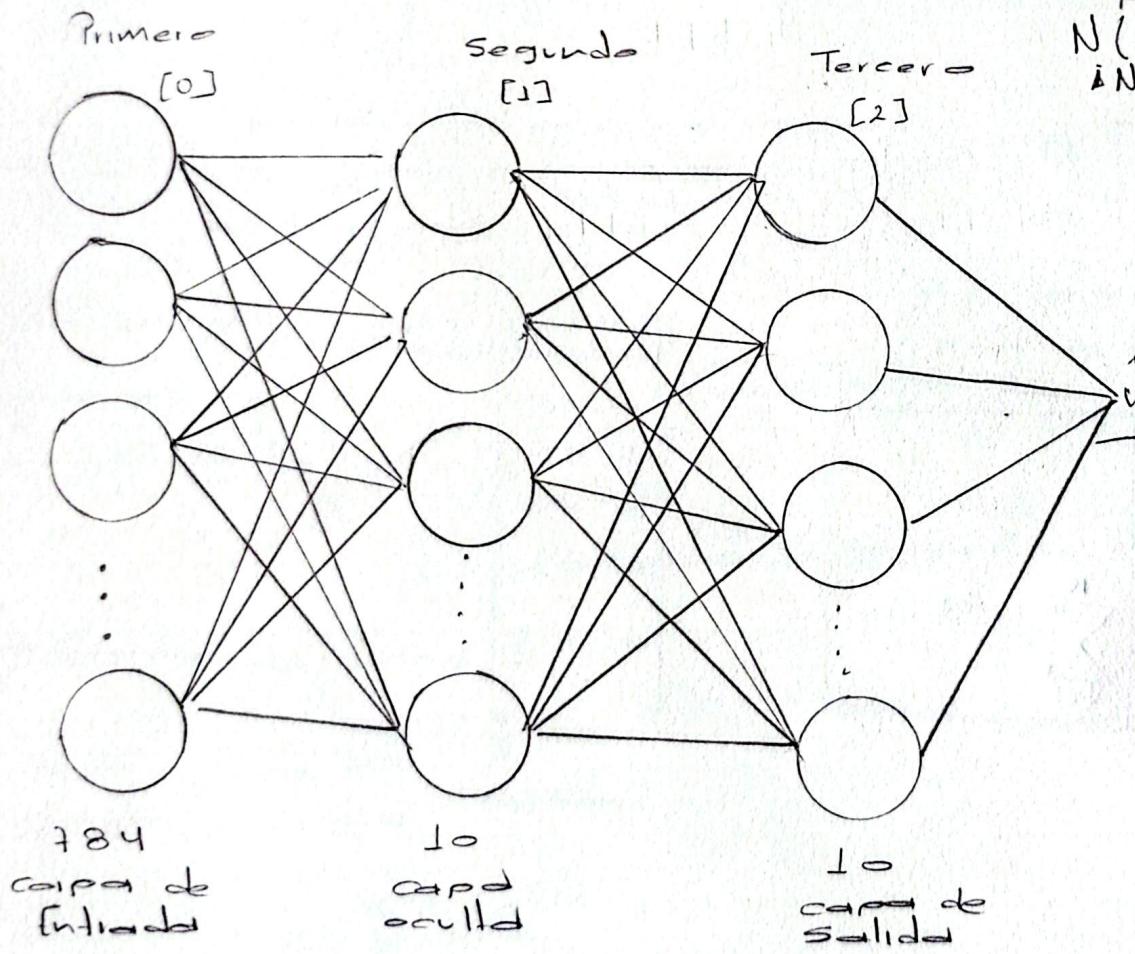
$$A^T = \begin{bmatrix} | & | & | \\ X_{(1)} & X_{(2)} & \dots & X_{(n)} \\ | & | & | \end{bmatrix}$$

784
28x28

\Rightarrow 0, 1, 2 ... 9

10 clases (10 neuronas)

Tener en
cuenta el
Proyecto
pide
 $N(256 - 1024)$
10 Neuronas
Ocultas!



Propagación hacia delante (Forward Propagation).

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

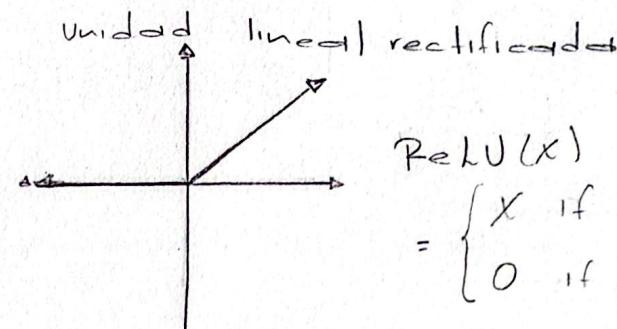
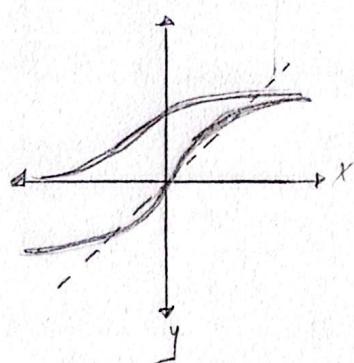
$$z^{[1]} = W^{[1]} A^{[0]} + b^{[1]}$$

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

Función de Activación \rightarrow Capa oculta

$$A^{[1]} = g(z^{[1]}) = \text{ReLU}(z^{[1]}) = \max(0, z^{[1]})$$

ReLU = Unidad lineal rectificada



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

$$z^{[2]} = W^{[2]} A^{[1]} + b^{[2]} \rightarrow \text{Capa de Salida}$$

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

$$A^{[2]} = \text{softmax}(z^{[2]})$$

Output layer

$$\begin{bmatrix} 1.3 \\ 5.1 \\ 2.2 \\ 0.7 \\ 1.1 \end{bmatrix}$$

softmax
función de activación

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

Probabilidades

$$\begin{bmatrix} 0.02 \\ 0.10 \\ 0.05 \\ 0.11 \\ 0.02 \end{bmatrix}$$

Propagación hacia atrás (Backward Propagation)

$$\text{Regla de la Cédula} \quad \frac{\partial L}{\partial W}$$

→ Entrar en la capa de salida

$$d_z^{[2]} = A^{[2]} - y$$

$\downarrow o \times m \quad \downarrow o \times m \quad \downarrow o \times m$

¡ Con esto podemos saber cuantos nos equivocamos en la predicción Cruda !

→ Gradientes de Capa [2] hacia atrás

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

$\downarrow o \times 1 \quad \downarrow m \times m \quad \downarrow o \times m \quad m \times 1$

$$db^{[2]} = \frac{1}{m} \sum_{cols} d_z^{[2]}$$

$\downarrow o \times 1$

→ final propagado a capa oculta ($d_z^{[1]}$)

$$d_z^{[1]} = (W^{[2]T} \cdot d_z^{[2]}) * g'(z^{[1]})$$

$\downarrow o \times m \quad \downarrow o \times o \quad \downarrow o \times m \quad \downarrow o \times m$

donde,

$$g'(z) = \begin{cases} 1, & \text{si } z > 0 \\ 0, & \text{si } z \leq 0 \end{cases}$$

* → Multiplicación elemento a elemento

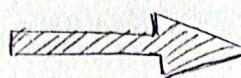
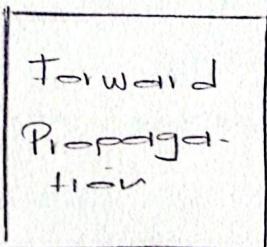
→ Gradientes Capa 1

$$dW^{[1]} = \frac{1}{m} d_z^{[1]} X^T$$

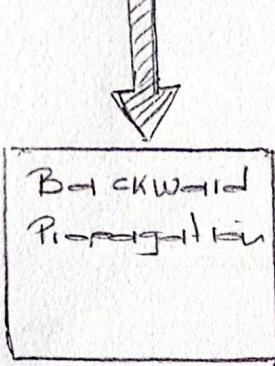
$\downarrow o \times 784 \quad \downarrow o \times m \quad m \times 784$

$$db^{[1]} = \frac{1}{m} \sum_{cols} d_z^{[1]}$$

$\downarrow o \times 1 \quad \downarrow m \times m \quad \downarrow o \times 1$



$\text{Softmax}(z^{[2]})$



Actualizamos los parámetros

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

α = hiperparámetro
(radio de aprendizaje)

Nota: ¡Importante!

Al siguiente es importante no lo confundir con el anterioridad, lo que es una función de Perdida, esta función se ejecuta en el intermedio de forward Propagation y Backward propagation, o sea esta en ese intermedio de las dos.

→ Loss Function (Perdida)

Matemáticamente esta minimizando el softmax

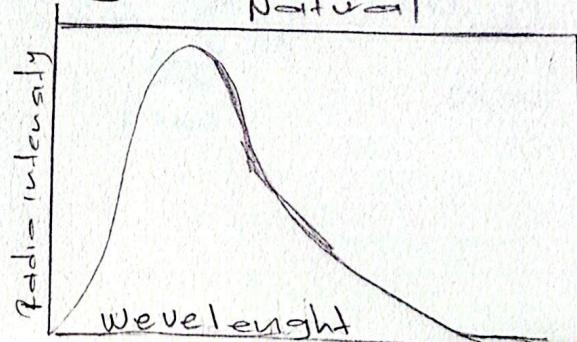
$$L = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n y_{j,i} \log(A_{j,i})$$

$$L(y, \hat{y}) = -\frac{1}{m} \sum_{j=1}^n y_j \cdot \log(\hat{y}_j)$$

$$\frac{\partial L}{\partial z^{[2]}} = \text{Predicción} - \text{Realidad}$$

$$\frac{\partial z^{[2]}}{\partial z} = A^{[2]} - y$$

$$\begin{aligned} y &(\text{Realidad}) = [0, 1] \\ \hat{y} &(\text{Predicción}) = [0, 0.8] \\ \log &= \text{logaritmo Natural} \end{aligned}$$



Notaciones y equivalencias en los apuntes

Concepto	Notación Apuntes	Notación PDF Proyecto	Forma
Input batch	$X = A^{[0]}$	X	$(784 \times m)^T$ $(784 \times m)$
Hidden Weights	W_1	$W^{[1]}$	(10×784) (N)
Nota de adicción: m es el numero de ejemplos y N numero de Neuronas, adicar que en el análisis lo realice con las neuronas en la capa oculta.			
Hidden Bias	b_1	$b^{[1]}$	(10×1)
Hidden Output	A_1	$a^{[1]}$	$(10 \times m)$
Output Weights	W_2	$W^{[2]}$	(10×10)
Output Output	A_2	$a^{[2]} = y$	$(10 \times m)$
Labels	y	y	$(1 \times m)$ (enteros) $(1 \rightarrow m)$ $(10 \times m)$ (One-Hot)

m = tamaño de carga actual (Número de filas)

1. Nivel físico (MNIST)

- Total de entrenamiento: 60.000 imágenes
- Total de Prueba: 10.000 imágenes

$$m = 60.000$$

- Matriz X de tamaño (784×60.000)