

Deep Learning para visión

Convolutional Neural Networks (CNNs)



$$y = Lx$$

Perceptron: red de una sola neurona

entrada $x = [x_1, x_2, x_3, \dots, x_n]$

salida $y \in \mathbb{R}$

parameters $W = [w_1, w_2, \dots, w_n]$

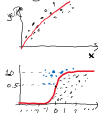
$$y = W^T x + b = \sum_{n=1}^N w_n x_n + b$$



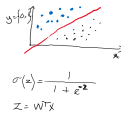
- incluir "b" como w_0
- no lineal

Podemos resolver:

- Regresión Lineal



- Clasificación Lineal



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

$$z = W^T x$$

Gradient Descent: entrenar = aprender parámetros

$$E = L(y, \hat{y}) \quad (y - \hat{y})^2 \rightarrow \text{coger todo el set de } M \text{ elementos de entrenamiento } E = \frac{1}{2} \sum_{i=1}^M (y_i - \hat{y}_i)^2$$

- Inicializar W aleatoriamente

- Calcular errores

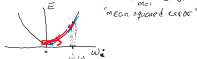
- Calcular $\frac{\partial E}{\partial w_i}$

- Ajustar w_i para disminuir error

$$w_i(t) = w_i(t-1) - \frac{\partial E}{\partial w_i}$$

valor de paso: Learning rate

$$w_i = w_i - \eta \frac{\partial E}{\partial w_i}, \quad \eta = 0.01 \dots 0.1$$

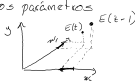


Gradient Descent (GD): se realiza al mismo tiempo para todos los parámetros

$$w_0(t) = w_0(t-1) - \eta \frac{\partial E}{\partial w_0(t-1)}$$

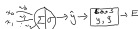
$$w_1(t) = w_1(t-1) - \eta \frac{\partial E}{\partial w_1(t-1)}$$

$$w_n(t) = w_n(t-1) - \eta \frac{\partial E}{\partial w_n(t-1)}$$



$$y = \sigma(z); \quad z = W^T x$$

$$L(y, \hat{y}) = (y - \hat{y})^2$$



$$\frac{\partial E}{\partial w_i} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \sigma(z)} \cdot \frac{\partial \sigma(z)}{\partial z} \cdot \frac{\partial z}{\partial w_i}$$

Back propagation
GD + Chain rule



x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

Función XOR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

¿Qué pasaría si?



no podemos activaciones no lineales?

¿qué pasa tengo 2 operaciones lineales una tras otra?

$$x W^{(1)} W^{(2)} = x W; \quad W = W^{(2)} W^{(1)}$$

Perceptrón multicapa (MLP)

Ejemplo de Backprop en MLP



$$x = [0.05, 0.1]$$

$$y = 1$$

$$w_1 = 0.15$$

$$w_2 = 0.20$$

$$w_3 = 0.25$$

$$w_4 = 0.30$$

$$w_5 = 0.50$$

$$\eta = 0.55$$

$$b_1 = 0.35$$

$$b_2 = 0.35$$

$$b_3 = 0.6$$

$$\eta = 0.1$$

$$s = \sum x_i w_i + b$$

$$a = \sigma(s)$$

$\sigma()$: sigmoide

$$E = \frac{1}{2} (y - \hat{y})^2$$

Proceso

1) Forward pass

2) Calcular errores

3) Calcular gradientes

4) Backward pass ajustes

Forward pass

$$s_1 = w_1 x_1 + w_2 x_2 + b_1$$

$$= (0.15)(0.05) + (0.25)(0.1) + 0.35$$

$$= 0.3825$$

$$a_1 = \sigma(s_1)$$

$$= 0.5995$$

$$s_2 = w_3 a_1 + w_4 x_2 + b_2$$

$$= (0.2)(0.05) + (0.2)(0.1) + 0.35$$

$$= 0.88969$$

$$a_2 = \sigma(s_2)$$

$$= 1$$

$$s_3 = w_5 a_1 + w_6 a_2 + b_3$$

$$= (0.5)(0.5995) + (0.55)(0.88969) + 0.6$$

$$= 1.2252$$

$$a_3 = \sigma(s_3)$$

$$= 0.773$$

$$\hat{y} = a_3$$

$$= 0.773$$

Error

$$E = \frac{1}{2} (y - \hat{y})^2$$

$$= 0.5 (1 - 0.773)^2$$

$$= 0.0258$$

Gradientes

$$\begin{aligned} \frac{\partial E}{\partial w_6} &= \frac{\partial E}{\partial a_3} \cdot \frac{\partial a_3}{\partial s_3} \cdot \frac{\partial s_3}{\partial w_6} \\ &= \frac{\partial}{\partial a_3} \left(\frac{1}{2} (y - a_3)^2 \right) \cdot \frac{\partial \sigma(s_3)}{\partial s_3} \cdot \frac{\partial (w_5 a_1 + w_6 a_2 + b_3)}{\partial w_6} \\ &= \frac{\partial}{\partial a_3} \left(\frac{1}{2} (y - a_3)^2 \right) \cdot \frac{\partial \sigma(s_3)}{\partial s_3} \cdot a_2 \\ &= (y - a_3) \cdot \sigma'(s_3) \cdot (1 - \sigma(s_3)) \cdot a_2 \\ &= (-1 - 0.773) \cdot (0.773) \cdot (1 - 0.773) \cdot (0.5995) \\ &= -0.0238 \end{aligned}$$

$$\begin{aligned} \frac{\partial E}{\partial w_5} &= \frac{\partial E}{\partial a_3} \cdot \frac{\partial a_3}{\partial s_3} \cdot \frac{\partial s_3}{\partial w_5} \\ &= \delta_3 \cdot a_1 \\ &= -0.0232 \end{aligned}$$

$$\begin{aligned} \frac{\partial E}{\partial b_3} &= \frac{\partial E}{\partial a_3} \cdot \frac{\partial a_3}{\partial s_3} \cdot \frac{\partial s_3}{\partial b_3} \\ &= \delta_3 \\ &= -0.0398 \end{aligned}$$

$$\begin{aligned} \frac{\partial E}{\partial w_1} &= \frac{\partial E}{\partial a_1} \cdot \frac{\partial a_1}{\partial s_1} \cdot \frac{\partial s_1}{\partial w_1} \\ &= \delta_1 \cdot \frac{\partial}{\partial a_1} \left(\frac{1}{2} (w_5 a_1 + w_6 a_2 + b_3)^2 \right) \cdot \frac{\partial \sigma(s_1)}{\partial s_1} \cdot \frac{\partial (w_1 x_1 + w_2 x_2 + b_1)}{\partial w_1} \\ &= \delta_1 \cdot w_5 \cdot \sigma'(s_1) \cdot (1 - \sigma(s_1)) \cdot x_1 \\ &= (-0.0398) \cdot (0.5) \cdot (0.5995) \cdot (1 - 0.5995) \cdot (0.05) \\ &= -0.0002 \end{aligned}$$

$$\begin{aligned} \frac{\partial E}{\partial w_2} &= \delta_1 \cdot x_2 \\ &= -0.0005 \end{aligned}$$

$$\begin{aligned} \frac{\partial E}{\partial b_1} &= \delta_1 \\ &= -0.0048 \end{aligned}$$

$$\begin{aligned} \frac{\partial E}{\partial w_4} &= \delta_3 \cdot \frac{\partial a_3}{\partial s_3} \cdot \frac{\partial s_3}{\partial w_4} \cdot \frac{\partial s_1}{\partial w_4} \\ &= \delta_1 \cdot x_1 \\ &= -0.0003 \end{aligned}$$

$$\begin{aligned} \frac{\partial E}{\partial w_4} &= \delta_1 \cdot x_2 \\ &= -0.0005 \end{aligned}$$

$$\begin{aligned} \frac{\partial E}{\partial b_2} &= \delta_1 \\ &= -0.0053 \end{aligned}$$

Backward Pass

$$w_1 = 0.15 - (0.1) \cdot (-0.0002) = 0.15002$$

$$w_2 = 0.2 - (0.1) \cdot (-0.0003) = 0.20003$$

$$w_3 = 0.25 - (0.1) \cdot (-0.0005) = 0.25005$$

$$w_4 = 0.3 - (0.1) \cdot (-0.0005) = 0.30005$$

$$w_5 = 0.5 - (0.1) \cdot (-0.0232) = 0.50232$$

$$w_6 = 0.55 - (0.1) \cdot (-0.0238) = 0.55238$$

$$b_1 = 0.3505$$

$$b_2 = 0.3505$$

$$b_3 = 0.604$$

New Forward Pass

$$s_1 = 0.382986$$

$$s_2 = 0.890537$$

$$s_3 = 1.232131$$

$$a_1 = 0.594593$$

$$a_2 = 0.594412$$

$$a_3 = 0.774192$$

$$\hat{y} = 0.774192$$

Q: Cómo hacemos para



$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial a_3} \cdot \frac{\partial a_3}{\partial s_3} \cdot \frac{\partial s_3}{\partial w_5} \cdot \frac{\partial s_1}{\partial w_5} \cdot \frac{\partial s_1}{\partial w_5} \quad \checkmark \text{ ok}$$

$$\frac{\partial E}{\partial w_6} = \frac{\partial E}{\partial a_3} \cdot \frac{\partial a_3}{\partial s_3} \cdot \frac{\partial s_3}{\partial w_6} \cdot \frac{\partial s_1}{\partial w_6} \cdot \frac{\partial s_1}{\partial w_6} \cdot \frac{\partial s_1}{\partial w_6} \quad \checkmark$$

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial a_3} \cdot \frac{\partial a_3}{\partial s_3} \cdot \frac{\partial s_3}{\partial w_5} \cdot \frac{\partial s_1}{\partial w_5} \cdot \frac{\partial s_1}{\partial w_5} \cdot \frac{\partial s_1}{\partial w_5} \quad \checkmark$$