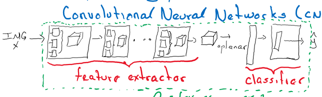
Deep Learning para visión

Convolutional Neural Networks (CNNs)



$\Omega = \{w_1, w_2, \dots w_N\}$

$y = \Omega x$

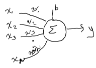
Perceptron: red de una sola neurona

entrada $\mathbf{x} = [x_1, x_2, x_3, \dots x_N]$

salida $y \in \mathbb{R}$

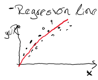
parametros $W = [w_1, w_2, \dots w_N]$
b

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

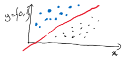


Podemos resolver:

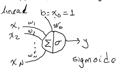
- Regresion lineal

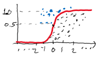


- Clasificacion lineal



- incluir "b" como w_0
- no lineal $b = x_0 = 1$





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

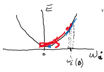
$z = W^T x$

Gradient Descent: entrenar = aprender parámetros

$E = \mathcal{L}(y, \hat{y})$ ej $(y - \hat{y})^2 \rightarrow$ sobre todo el set de M elementos de entrenamiento $E = \frac{1}{M} \sum_{m=1}^{M} (y - \hat{y})^2$

"mean squared error"

- Iniciar W aleatoriamente
- Calcular error
- Calcular $\frac{\partial E}{\partial w_i}$ $\overline{\partial w_i(t-1)}$
- Ajustar w_i para disminuir error



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

valor de paso: learnin rate

$$\underline{w_i = w_i - \eta \frac{\partial E}{\partial w_i}}, \quad \eta = 0, \dots 1 \approx 0.01$$

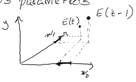
Gradient Descent (GD): se realiza al mismo tiempo para todos los parámetros

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

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

$\vdots$

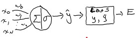
$w_N(t) = w_N(t-1) - \eta \frac{\partial E}{\partial w_N(t-1)}$



Ej)

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

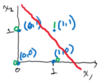
$\mathcal{L}(y, \hat{y}) = (y - \hat{y})^2$



$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \sigma(z)} \cdot \frac{\partial \sigma(z)}{\partial z} \cdot \frac{\partial z}{\partial w_1}$$

← **Backpropagation**

GD + Chain rule

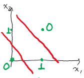


Funcion AND

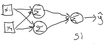
x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

Funcion XOR

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



¿Qué pasaría si?

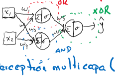


si "no ponemos activaciones no lineales"

¿Qué pasa tengo 2 operaciones lineales, una tras otra?

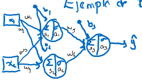
$X W^{(1)} \cdot W^{(2)} = XW$; $W = W^{(1)} W^{(2)}$

$\sigma(XW^{(1)}) X^{(2)}$



Perceptrón multicapa (MLP)

Ejemplo de Backprop en MLP ← Ejemplo



$\mathbf{X} = [0.05, 0.1]$

$y = 1$

$w_1 = 0.15$ $b_1 = 0.35$

$w_2 = 0.20$ $b_2 = 0.35$

$w_3 = 0.25$ $b_3 = 0.6$

$w_4 = 0.30$ $\eta = 0.1$

$w_5 = 0.50$

parametros w_6 0.55

$s = \sum_i x_i w_i + b$

$a = \sigma(s)$

$\sigma()$: sigmoide

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

Proceso

1) Forward pass
2) Calcular error
3) Calcular gradientes
4) Backward pass (ajustar

Forward pass

$s_1 = w_1 x_1 + w_3 x_2 + b_1$
$= (0.15)(0.05) + (0.25)(0.1) + 0.35$
$= 0.3825$

$a_1 = \sigma(s_1)$
$= 0.5945$

$s_2 = w_2 x_1 + w_4 x_2 + b_2$
$= (0.2)(0.05) + (0.3)(0.1) + 0.35$
$= 0.3964$

$a_2 = \sigma(s_2)$
$=$

$s_3 = w_5 a_1 + w_6 a_2 + b_3$
$= (0.5)(0.5945) +$
$(0.55)(0.5964) +$
0.6
$= 1.2252$

$a_3 = \sigma(s_3)$
$= 0.773$

$\hat{y} = a_3$
$= 0.773$

Error

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

$= 0.5(1 - 0.733)^2$

$= 0.0258$

$$\boxed{\sigma'(s) = \sigma(s)(1 - \sigma(s))}$$

Gradientes

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

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

$= \frac{\partial}{\partial a_3} \frac{1}{2}(y - a_3)^2 \cdot \frac{\partial}{\partial s_3} \sigma(s_3) \cdot \frac{\partial}{\partial w_6}(w_5 a_1 + w_6 a_2 + b_3)$

$= \frac{2}{2}(y - a_3) \frac{\partial}{\partial a_3} \cancel{(y - a_3)}' \cdot \sigma(s_3)(1 - \sigma(s_3)) \cdot a_2$

$= \underbrace{(y - a_3)(-1) \sigma(s_3)(1 - \sigma(s_3))}_{\delta_3} a_2$

$= -(1 - 0.733)(0.773)(1 - 0.773)(0.5963)$

$= -0.0238$

$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial a_3} \cdot \frac{\partial a_3}{\partial s_3} \cdot \frac{\partial s_3}{\partial w_5}$
$= \delta_3 \cdot a_1$
$= -0.0237$

$\frac{\partial E}{\partial b_3} = \frac{\partial E}{\partial a_3} \cdot \frac{\partial a_3}{\partial s_3} \cdot \frac{\partial s_3}{\partial b_3}$
$= \delta_3$
$= -0.0398$

$\frac{\partial E}{\partial w_1} = \underbrace{\frac{\partial E}{\partial a_3} \cdot \frac{\partial a_3}{\partial s_3}}_{\delta_3} \cdot \frac{\partial s_3}{\partial a_1} \cdot \frac{\partial a_1}{\partial s_1} \cdot \frac{\partial s_1}{\partial w_1}$

$= \delta_3 \cdot \frac{\partial}{\partial a_1}(w_5 a_1 + w_6 a_2 + b_3) \cdot \frac{\partial}{\partial s_1} \sigma(s_1) \cdot \frac{\partial}{\partial w_1}(w_1 x_1 + w_3 x_3 + b_1)$

$= \underbrace{\delta_3 \cdot w_5 \cdot \sigma(s_1)(1 - \sigma(s_1))}_{\delta_2} \cdot x_1$

$= (-0.0398)(0.5)(0.5945)(1 - 0.5945)(0.05)$

$= -0.0002$

$\frac{\partial E}{\partial w_3} = \delta_2 x_2$
$= -0.0005$

$\frac{\partial E}{\partial b_1} = \delta_2$
$= -0.0048$

$\frac{\partial E}{\partial w_2} = \underbrace{\delta_3 \cdot \frac{\partial s_3}{\partial a_2} \cdot \frac{\partial a_2}{\partial s_2}}_{\delta_1} \cdot \frac{\partial s_2}{\partial w_2}$

$= \delta_1 \cdot x_1$

$= -0.0003$

$\frac{\partial E}{\partial w_4} = \delta_1 \cdot x_2$
$= -0.0005$

$\frac{\partial E}{\partial b_2} = \delta_1$
$= -0.0053$

Backward Pass

$w_i = w_i - \eta \frac{\partial E}{\partial w_i}$

$w_1 = 0.15 - (0.1)(-0.0002) = 0.15002$ ✓

$w_2 = 0.2 - (0.1)(-0.0003) = 0.20003$ ✓

$w_3 = 0.25 - (0.1)(-0.0005) = 0.25005$ ✓

$w_4 = 0.3 - (0.1)(-0.0005) = 0.30005$ ✓

$w_5 = 0.5 - (0.1)(-0.0237) = 0.50237$ ✓

$w_6 = 0.55 - (0.1)(-0.0238) = 0.55238$ ✓

$b_1 = 0.3505$ ✓

$b_2 = 0.3505$ ✓

$b_3 = 0.604$ ✓

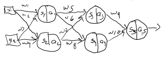
New Forwar Pass

$s_1 = 0.382986$ $s_2 = 0.390537$ $s_3 = 1.232132$

$a_1 = 0.594593$ $a_2 = 0.596412$ $a_3 = 0.774192$

$\hat{y} = 0.774192$

Q: Cómo hacemos para



$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial a_5} \cdot \frac{\partial a_5}{\partial s_5} \cdot \frac{\partial s_5}{\partial a_3} \cdot \frac{\partial a_3}{\partial s_3} \cdot \frac{\partial s_3}{\partial w_5}$ ✓ ok

$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial a_5} \cdot \frac{\partial a_5}{\partial s_5} \cdot \frac{\partial s_5}{\partial a_3} \cdot \frac{\partial a_3}{\partial s_3} \cdot \frac{\partial s_3}{\partial a_1} \cdot \frac{\partial a_1}{\partial s_1} \cdot \frac{\partial s_1}{\partial w_1} +$

$\frac{\partial E}{\partial a_5} \cdot \frac{\partial a_5}{\partial s_5} \cdot \frac{\partial s_5}{\partial a_4} \cdot \frac{\partial a_4}{\partial s_4} \cdot \frac{\partial a_1}{\partial s_1} \cdot \frac{\partial s_1}{\partial w_1}$ ✓