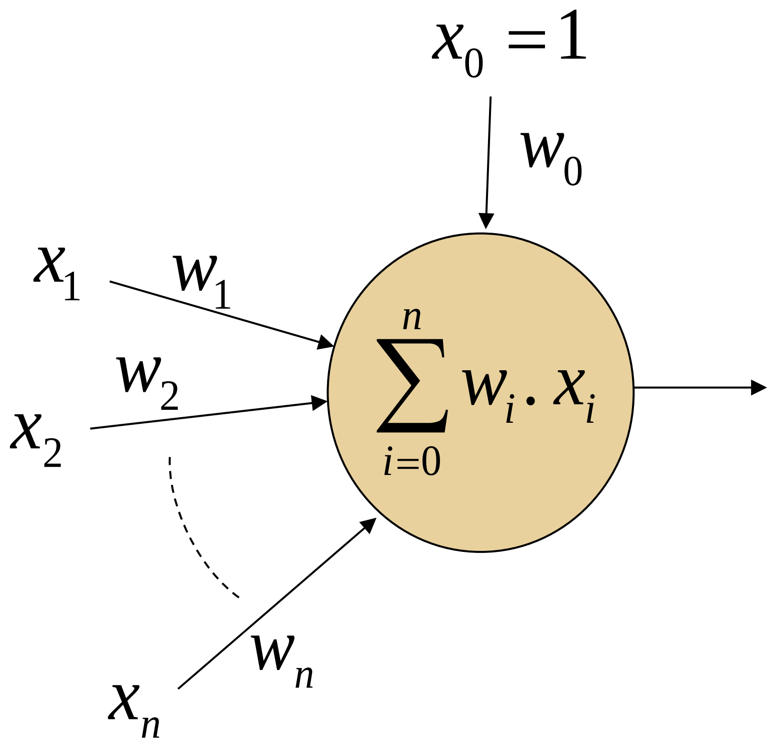
Combinador Lineal

- Resuelve un problema de **Regresión Lineal**
- Función de Error
  - Error cuadrático medio
- Técnica de optimización
  - Descenso de gradiente estocástico



$$y = \sum_{i=0}^{n} x_i . w_i$$

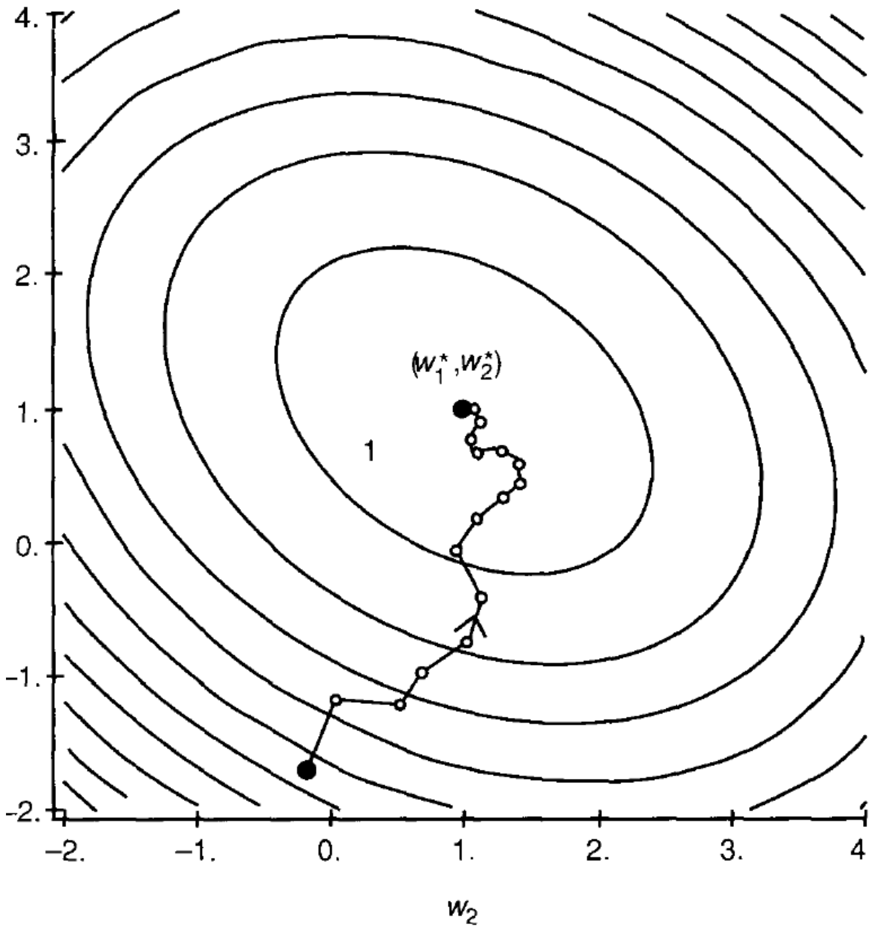
Técnica del descenso del gradiente estocástico

$$w(t+1) = w(t) + \Delta w(t)$$

$$w(t+1) = w(t) - \mu \nabla \xi(w(t))$$

- se utiliza

$$\xi = <\varepsilon_k^2> \approx \varepsilon_k^2 = (d_k - \sum_{i=0}^{N} x_{ik} w_i)^2$$



Técnica del descenso del gradiente estocástico

$$w(t+1) = w(t) + \Delta w(t)$$

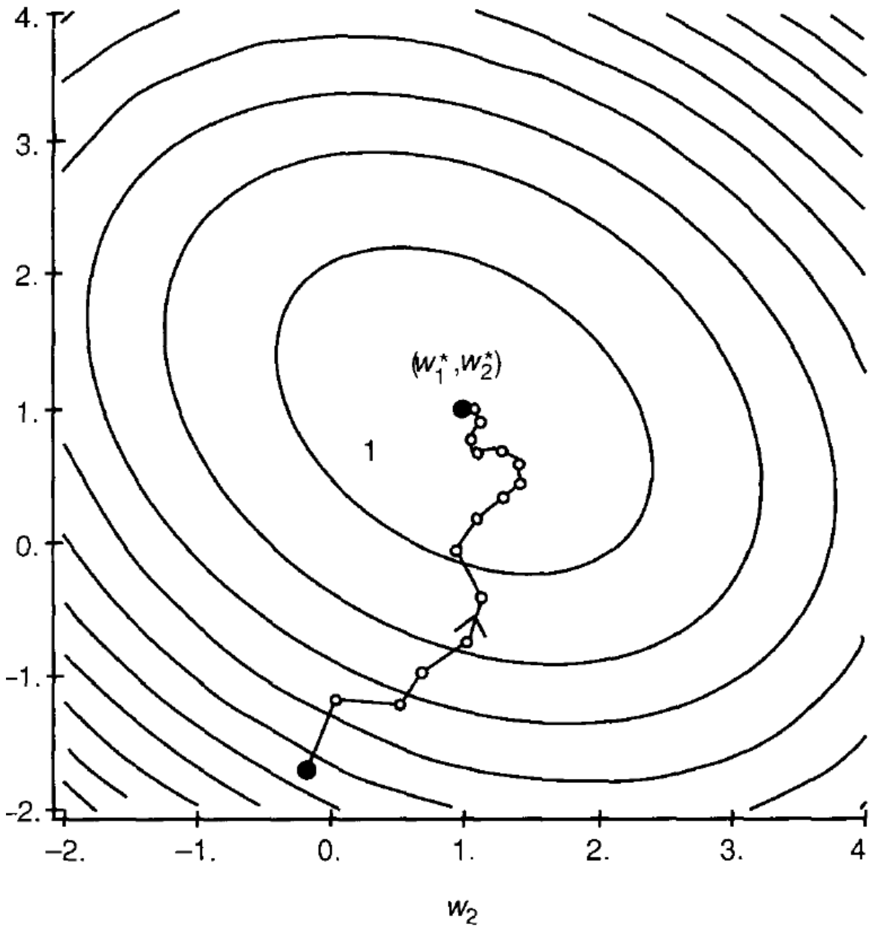
$$w(t+1) = w(t) - \mu \nabla \xi(w(t))$$

- se utiliza

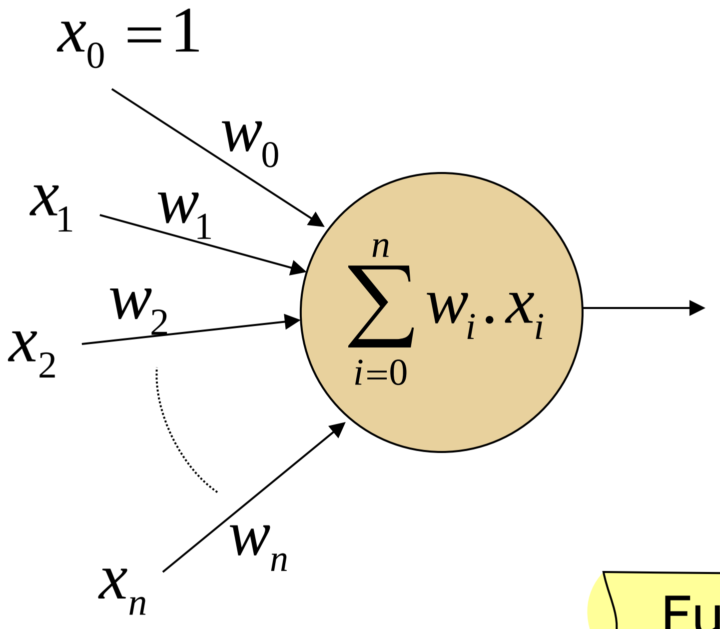
$$\xi = <\varepsilon_k^2> \approx \varepsilon_k^2 = (d_k - \sum_{i=0}^{N} x_{ik} w_i)^2$$

- veamos que

$$\nabla \varepsilon_k^2(t) = -2\varepsilon_k(t) x_k$$



Neurona General



$$y_k = f\left(\sum_{i=0}^{n} w_i . x_i\right)$$

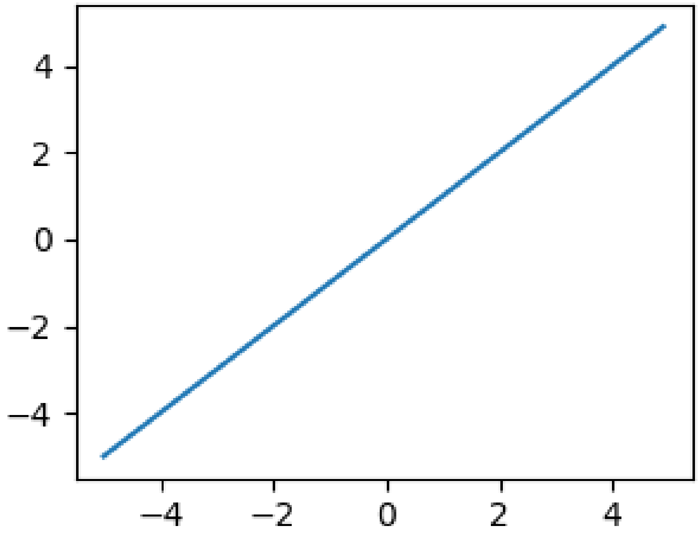
Función derivable

Función de Salida LINEAL

$$f(x) = x \qquad f'(x) = 1$$

desde Python

```
import numpy as np
import grafica as gr
from matplotlib import pyplot as plt

x = np.array(range(-50,50,1))/10.0
y = gr.evaluar('purelin', x)
plt.plot(x,y,'-')
```



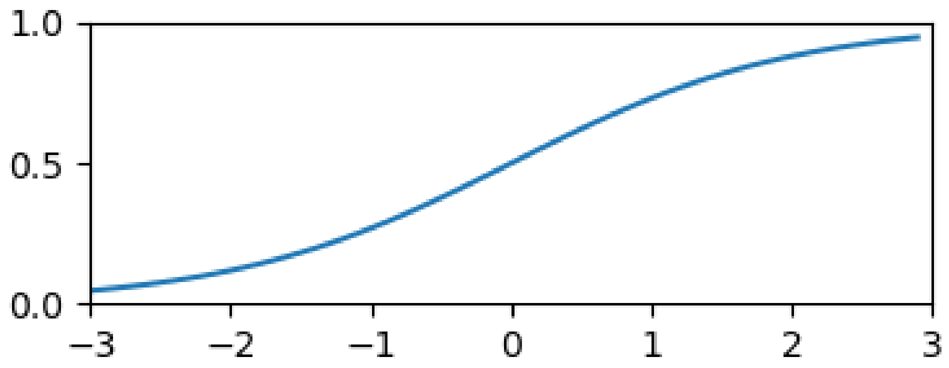
Función SIGMOIDE $\in$ (0,1)

$$f(x) = \frac{1}{1+e^{-x}}$$

$$f'(x) = f(x) * (1 - f(x))$$

desde Python

```
import numpy as np
import grafica as gr
from matplotlib import pyplot as plt

x = np.array(range(-30,30,1))/10.0
y = gr.evaluar('logsig', x)
plt.plot(x,y,'-')
plt.axis([-3, 3, 0, 1])
plt.show()
```



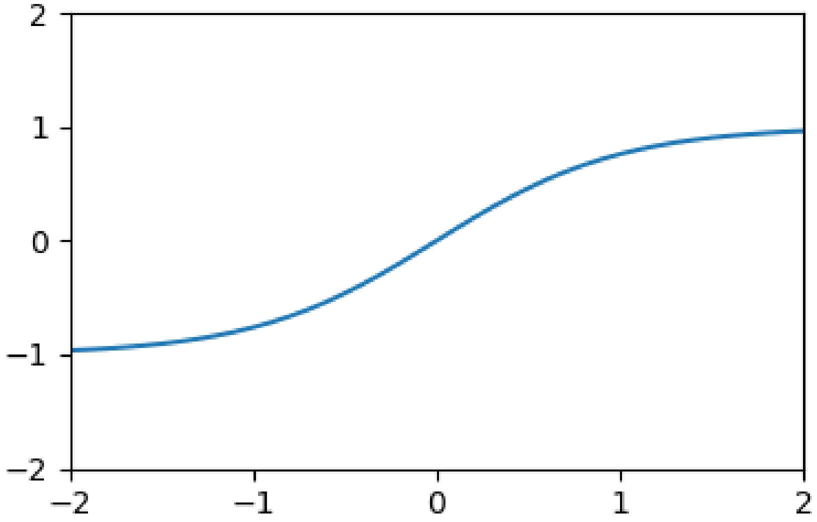
Función SIGMOIDE $\in$ (-1,1)

$$f(x) = \frac{2}{1+e^{-2x}} - 1$$

$$f'(x) = 1 - f(x)^2$$

desde Python

```
import numpy as np
import grafica as gr
from matplotlib import pyplot as plt

x = np.array(range(-30,30,1))/10.0
y = gr.evaluar('tansig', x)
plt.plot(x,y,'-')
plt.axis([-2, 2, -2, 2])
plt.show()
```



Ejemplo

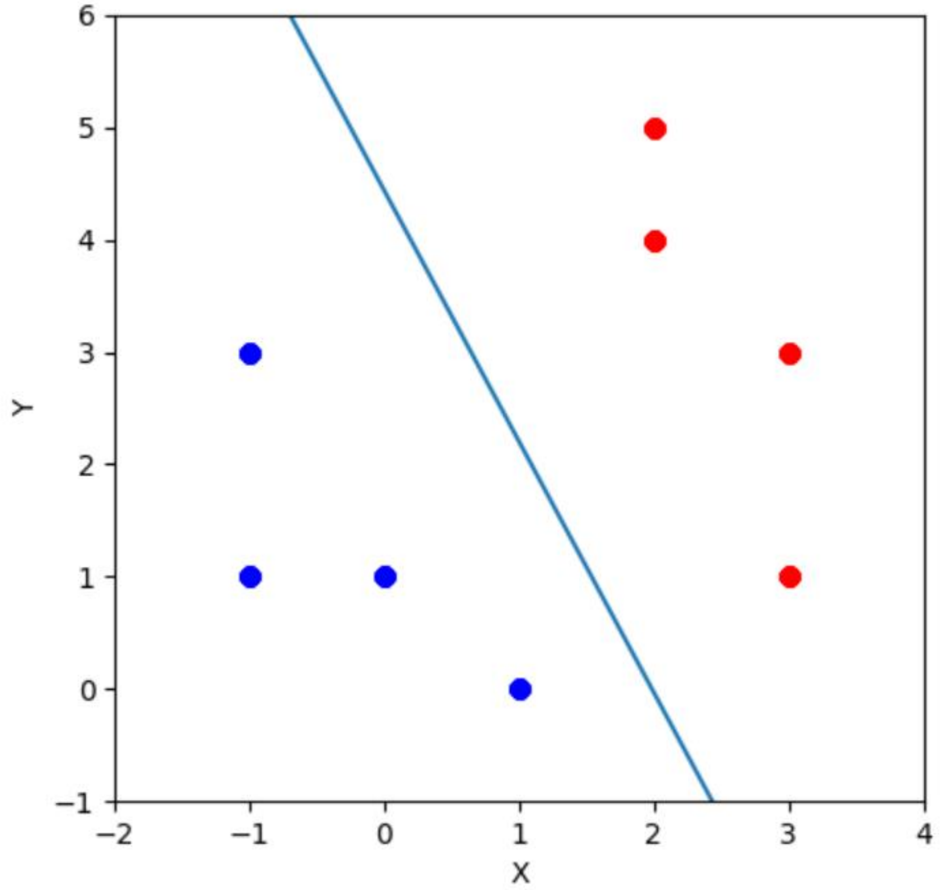
- Dados los siguientes conjuntos de puntos del plano

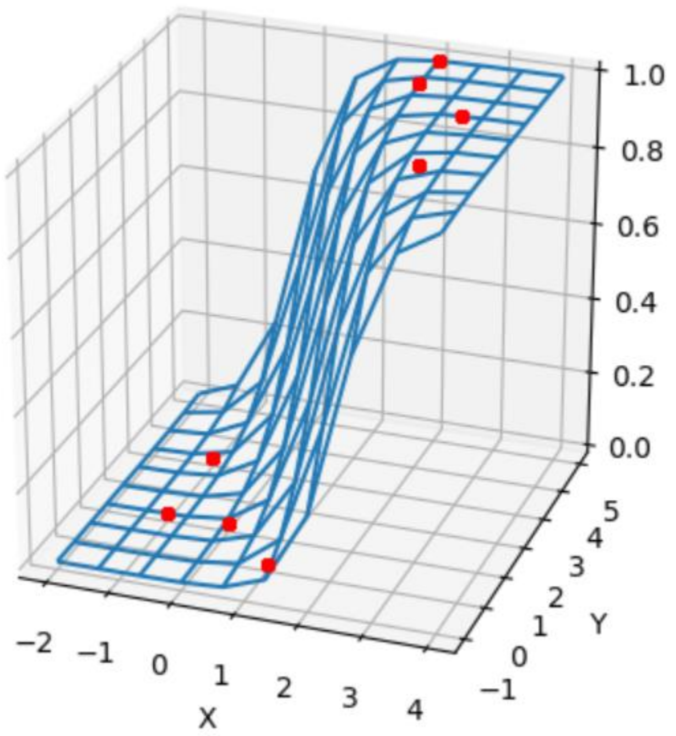
  A = {(2,2), (1,0), (0,1), (-1,1)}

  B = {(3,1), (3,3), (2,4), (2,5)}

  - Utilice una **neurona no lineal** para clasificarlos
  - Representar gráficamente la solución propuesta.

A = {(-1,3), (1,0), (0,1), (-1,1)}
B = {(3,1), (3,3), (2,4), (2,5)}





neuronaNoLineal.py

Entrenamiento de una neurona no lineal

- Seleccionar el valor de α
- Inicializar los pesos W y b con valores random
- Mientras (la variación del ECM sea mayor a la cota prefijada)
  - Para cada ejemplo
    - Ingresar el ejemplo a la red.
    - Calcular el error $\varepsilon = (y - \hat{y})$ y $\frac{\partial \varepsilon^2}{\partial w} = \frac{\partial (y-\hat{y})^2}{\partial w}$
    - Actualizar los pesos de la red

$$w_i = w_i - \alpha \frac{\partial \varepsilon^2}{\partial w_i}$$

¿Cómo sería la derivada del error si la neurona no es lineal?

$$\frac{\partial \varepsilon^2}{\partial w} = \left[\frac{\partial (y-\hat{y})^2}{\partial w_0}; \quad \dots; \quad \frac{\partial (y-\hat{y})^2}{\partial w_n} \right]$$

$$\frac{\partial \varepsilon^2}{\partial w_j} = -2(y-\hat{y}) \underbrace{\frac{\partial f}{\partial (neta)}}_{f'} \frac{\partial (neta)}{\partial w_j}$$

$$\frac{\partial (neta)}{\partial w_j} = \frac{\partial (\sum_{i=0}^{n} w_i x_i)}{\partial w_j} = x_j$$

$$\boxed{\frac{\partial \varepsilon^2}{\partial w_j} = -2(y-\hat{y}) f'(neta) x_j}$$

Entrenamiento de una neurona no lineal

- Seleccionar el valor de α
- Inicializar los pesos W y b con valores random
- Mientras (la variación del ECM sea mayor a la cota prefijada)
  - Para cada ejemplo
    - Ingresar el ejemplo a la red.
    - Calcular $\frac{\partial \varepsilon^2}{\partial w_i} = -2 * \varepsilon * f'(neta) * x_i$
    - Actualizar los pesos de la red

$$w_i = w_i - \alpha \frac{\partial \varepsilon}{\partial w_i} = w_i + 2\alpha * \varepsilon * f'(neta) * x_i$$

AND

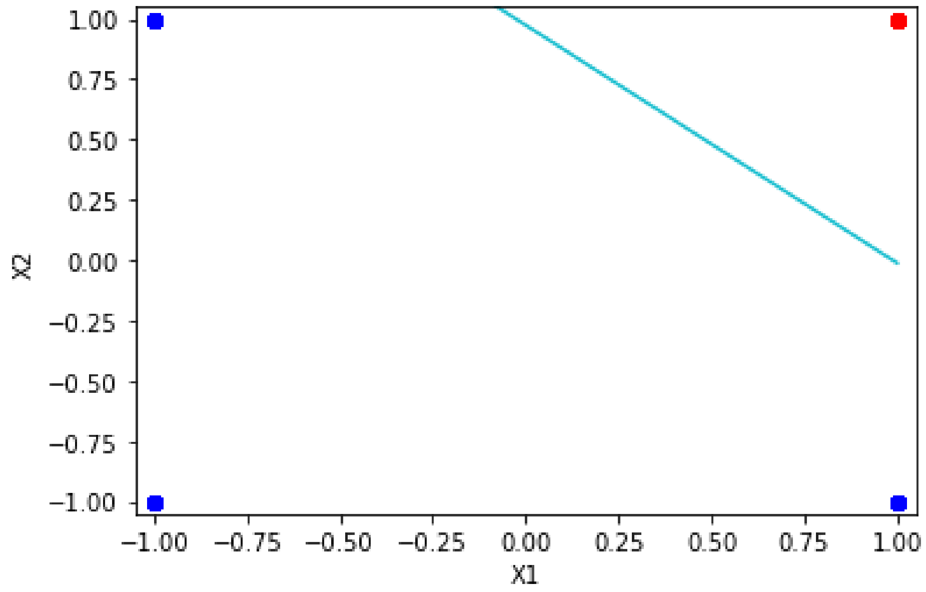
- Utilice una neurona no lineal con salida sigmoide para resolver el problema del AND

$$f(x) = \frac{1}{1+e^{-x}}$$

$$f'(x) = f(x) * (1 - f(x))$$

$$f(x) = \frac{2}{1+e^{-2x}} - 1$$

$$f'(x) = 1 - f(x) * f(x)$$



Modifique el algoritmo del PERCEPTRON hecho en clase y utilice una neuronal no lineal para resolver el problema del AND

```
import numpy as np
X = np.array([[0,0], [0,1],[1,0],[1,1]])
T = np.array([0,0,0,1])
W = np.random.uniform(-0.5, 0.5,size=2)
b = np.random.uniform(-0.5, 0.5)

alfa = 0.1
MAX_ITE = 5000
COTA = 10e-06
ite = 0
E_ant = 0
E = 1
while ((ite<MAX_ITE) and (np.abs(E_ant - E) > COTA)):
    E_ant=E
    sumaError = 0
    for p in range(len(X)):
        neta = np.dot(W,X[p,:])+b
        Y = 1/(1+np.exp(-neta))
        deriv = Y * (1-Y)
        Error = T[p]-Y
        W = W + alfa * Error * deriv * X[p,:]
        b = b + alfa * Error * deriv
        sumaError = sumaError + Error**2
    E = sumaError / len(X)
    ite = ite + 1
    print ("ite= %d   w0= %8.5f   w1=%8.5f   E=%.8f" % (ite,W[0],W[1],E))
```

Los pesos iniciales son aleatorios

```python
import numpy as np
X = np.array([[0,0], [0,1],[1,0],[1,1]])
T = np.array([0,0,0,1])
W = np.random.uniform(-0.5, 0.5,size=2)
b = np.random.uniform(-0.5, 0.5)
alfa = 0.1
MAX_ITE = 5000
COTA = 10e-06
ite = 0
E_ant = 0
E = 1
while ((ite<MAX_ITE) and (np.abs(E_ant - E) > COTA)):
    E_ant=E
    sumaError = 0
    for p in range(len(X)):
        neta = np.dot(W,X[p,:])+b
        Y = 1/(1+np.exp(-neta))
        deriv = Y * (1-Y)
        Error = T[p]-Y
        W = W + alfa * Error * deriv * X[p,:]
        b = b + alfa * Error * deriv
        sumaError = sumaError + Error**2
    E = sumaError / len(X)
    ite = ite + 1
    print ("ite= %d   w0= %8.5f   w1=%8.5f   E=%.8f" % (ite,W[0],W[1],E))
```

Parámetros del entrenamiento

```python
import numpy as np
X = np.array([[0,0], [0,1],[1,0],[1,1]])
T = np.array([0,0,0,1])

W = np.random.uniform(-0.5, 0.5,size=2)
b = np.random.uniform(-0.5, 0.5)

alfa = 0.1
MAX_ITE = 5000
COTA = 10e-06

ite = 0
E_ant = 0
E = 1
while ((ite<MAX_ITE) and (np.abs(E_ant - E) > COTA)):
    E_ant=E
    sumaError = 0

    for p in range(len(X)):
        neta = np.dot(W,X[p,:])+b

        Y = 1/(1+np.exp(-neta))
        deriv = Y * (1-Y)

        Error = T[p]-Y

        W = W + alfa * Error * deriv * X[p,:]
        b = b + alfa * Error * deriv

        sumaError = sumaError + Error**2

    E = sumaError / len(X)
    ite = ite + 1

    print ("ite= %d   w0= %8.5f   w1=%8.5f   E=%.8f" % (ite,W[0],W[1],E))
```

Termina o bien porque realizó la máxima cantidad de intentos o porque el valor absoluto de la diferencia entre dos valores consecutivos de la función es inferior a cierta cota

```
import numpy as np
X = np.array([[0,0], [0,1],[1,0],[1,1]])
T = np.array([0,0,0,1])

W = np.random.uniform(-0.5, 0.5,size=2)
b = np.random.uniform(-0.5, 0.5)

alfa = 0.1
MAX_ITE = 5000
COTA = 10e-06

ite = 0
E_ant = 0
E = 1
while ((ite<MAX_ITE) and (np.abs(E_ant - E) > COTA)):
    E_ant=E
    sumaError = 0

    for p in range(len(X)):
        neta = np.dot(W,X[p,:])+b

        Y = 1/(1+np.exp(-neta))
        deriv = Y * (1-Y)

        Error = T[p]-Y

        W = W + alfa * Error * deriv * X[p,:]
        b = b + alfa * Error * deriv

        sumaError = sumaError + Error**2

    E = sumaError / len(X)
    ite = ite + 1

    print ("ite= %d   w0= %8.5f   w1=%8.5f   E=%.8f" % (ite,W[0],W[1],E))
```

Calculamos la salida del combinador lineal y evaluamos el error cometido

```
import numpy as np
X = np.array([[0,0], [0,1],[1,0],[1,1]])
T = np.array([0,0,0,1])
W = np.random.uniform(-0.5, 0.5,size=2)
b = np.random.uniform(-0.5, 0.5)

alfa = 0.1
MAX_ITE = 5000
COTA = 10e-06
ite = 0
E_ant = 0
E = 1
while ((ite<MAX_ITE) and (np.abs(E_ant - E) > COTA)):
    E_ant=E
    sumaError = 0
    for p in range(len(X)):
        neta = np.dot(W,X[p,:])+b
        Y = 1/(1+np.exp(-neta))
        deriv = Y * (1-Y)
        Error = T[p]-Y
        W = W + alfa * Error * deriv * X[p,:]
        b = b + alfa * Error * deriv
        sumaError = sumaError + Error**2
    E = sumaError / len(X)
    ite = ite + 1
    print ("ite= %d   w0= %8.5f   w1=%8.5f   E=%.8f" % (ite,W[0],W[1],E))
```

Son parte del vector gradiente

```python
import numpy as np
X = np.array([[0,0], [0,1],[1,0],[1,1]])
T = np.array([0,0,0,1])
W = np.random.uniform(-0.5, 0.5,size=2)
b = np.random.uniform(-0.5, 0.5)

alfa = 0.1
MAX_ITE = 5000
COTA = 10e-06
ite = 0
E_ant = 0
E = 1
while ((ite<MAX_ITE) and (np.abs(E_ant - E) > COTA)):
    E_ant=E
    sumaError = 0
    for p in range(len(X)):
        neta = np.dot(W,X[p,:])+b
        Y = 1/(1+np.exp(-neta))
        deriv = Y * (1-Y)
        Error = T[p]-Y
        W = W + alfa * Error * deriv * X[p,:]
        b = b + alfa * Error * deriv
        sumaError = sumaError + Error**2
    E = sumaError / len(X)
    ite = ite + 1
    print ("ite= %d   w0= %8.5f   w1=%8.5f   E=%.8f" % (ite,W[0],W[1],E))
```

Actualizamos los pesos en la dirección del gradiente negativo

```python
import numpy as np
X = np.array([[0,0], [0,1],[1,0],[1,1]])
T = np.array([0,0,0,1])

W = np.random.uniform(-0.5, 0.5,size=2)
b = np.random.uniform(-0.5, 0.5)

alfa = 0.1
MAX_ITE = 5000
COTA = 10e-06

ite = 0
E_ant = 0
E = 1
while ((ite<MAX_ITE) and (np.abs(E_ant - E) > COTA)):
    E_ant=E
    sumaError = 0

    for p in range(len(X)):
        neta = np.dot(W,X[p,:])+b

        Y = 1/(1+np.exp(-neta))
        deriv = Y * (1-Y)

        Error = T[p]-Y

        W = W + alfa * Error * deriv * X[p,:]
        b = b + alfa * Error * deriv

        sumaError = sumaError + Error**2

    E = sumaError / len(X)
    ite = ite + 1

    print ("ite= %d   w0= %8.5f   w1=%8.5f   E=%.8f" % (ite,W[0],W[1],E))
```

Acumulamos el cuadrado de los errores cometidos

```python
import numpy as np
X = np.array([[0,0], [0,1],[1,0],[1,1]])
T = np.array([0,0,0,1])
W = np.random.uniform(-0.5, 0.5,size=2)
b = np.random.uniform(-0.5, 0.5)

alfa = 0.1
MAX_ITE = 5000
COTA = 10e-06
ite = 0
E_ant = 0
E = 1
while ((ite<MAX_ITE) and (np.abs(E_ant - E) > COTA)):
    E_ant=E
    sumaError = 0
    for p in range(len(X)):
        neta = np.dot(W,X[p,:])+b
        Y = 1/(1+np.exp(-neta))
        deriv = Y * (1-Y)
        Error = T[p]-Y
        W = W + alfa * Error * deriv * X[p,:]
        b = b + alfa * Error * deriv
        sumaError = sumaError + Error**2
    E = sumaError / len(X)
    ite = ite + 1
    print ("ite= %d   w0= %8.5f   w1=%8.5f   E=%.8f" % (ite,W[0],W[1],E))
```

Dividimos por la cantidad de ejemplos para obtener el ECM

NeuronaGral_AND.ipynb

```
import numpy as np
X = np.array([[0,0], [0,1],[1,0],[1,1]])
T = np.array([0,0,0,1])
W = np.random.uniform(-0.5, 0.5,size=2)
b = np.random.uniform(-0.5, 0.5)

alfa = 0.1
MAX_ITE = 5000
COTA = 10e-06
ite = 0
E_ant = 0
E = 1
while ((ite<MAX_ITE) and (np.abs(E_ant - E) > COTA)):
    E_ant=E
    sumaError = 0
    for p in range(len(X)):
        neta = np.dot(W,X[p,:])+b
        Y = 1/(1+np.exp(-neta))
        deriv = Y * (1-Y)
        Error = T[p]-Y
        W = W + alfa * Error * deriv * X[p,:]
        b = b + alfa * Error * deriv
        sumaError = sumaError + Error**2
    E = sumaError / len(X)
    ite = ite + 1
    print ("ite= %d   w0= %8.5f   w1=%8.5f   E=%.8f" % (ite,W[0],W[1],E))
```

ClassNeuronaGral.py

nn = NeuronaGradiente(alpha=0.01, n_iter=50, cotaE=10E-07, FUN=‘sigmoid’, random_state=None, draw=0, title=['X1','X2'])

- Parámetros de entrada
  - **alpha**: valor en el intervalo (0, 1] que representa la velocidad de aprendizaje.
  - **n_iter**: máxima cantidad de iteraciones a realizar.
  - **cotaE**: termina si la diferencia entre dos errores consecutivos es menor que este valor.
  - **FUN**: función de activación – ‘sigmoid’, ‘tanh’, ‘purelin’.
  - **random_state**: None si los pesos se inicializan en forma aleatoria, un valor entero para fijar la semilla
  - **draw**: valor distinto de 0 si se desea ver el gráfico y 0 si no. Sólo si es 2D.
  - **title**: lista con los nombres de los ejes para el gráfico. Se usa sólo si **draw** no es cero.

ClassNeuronaGral.py

nn = NeuronaGradiente(alpha=0.01, n_iter=50, cotaE=10E-07, FUN='sigmoid', random_state=None, draw=0, title=['X1','X2'])

nn.fit(X, T)

- Parámetros de entrada
  - **X** : arreglo de NxM donde N es la cantidad de ejemplos y M la cantidad de atributos.
  - **T** : arreglo de N elementos siendo N la cantidad de ejemplos
- Retorna
  - **w_** : arreglo de M elementos siendo M la cantidad de atributos de entrada
  - **b_** : valor numérico continuo correspondiente al bias.
  - **errors_**: errores cometidos en cada iteración.

ClassNeuronaGral.py

Y = nn.predict(X)

- Parámetros de entrada
  - **X** : arreglo de NxM donde N es la cantidad de ejemplos y M la cantidad de atributos.

- Retorna: un arreglo con el resultado de aplicar la neurona general entrenada previamente con fit() a la matriz de ejemplos X.
  - **Y** : arreglo de N elementos siendo N la cantidad de ejemplos

NeuronaGral_AND_RN.ipynb

```
import numpy as np
from ClassNeuronaGral import NeuronaGradiente

# Ejemplos de entrada de la función AND
X = np.array([[0,0], [0,1],[1,0],[1,1]])
X = 2*X-1
T = np.array([0,0,0,1])

ppn = NeuronaGradiente(alpha=0.1, n_iter=50, cotaE=10e-07, FUN='sigmoid',
                       random_state=None, draw=1, title=['x1', 'x2'])
ppn.fit(X,T)

#-- % de aciertos ---
Y = (ppn.predict(X)>0.5)*1
print("Y = ", Y)
print("T = ", T)
aciertos = sum(Y == T)
print("aciertos = %d   (%.2f%%)" % (aciertos, 100*aciertos/X.shape[0]))
```

Ejemplo

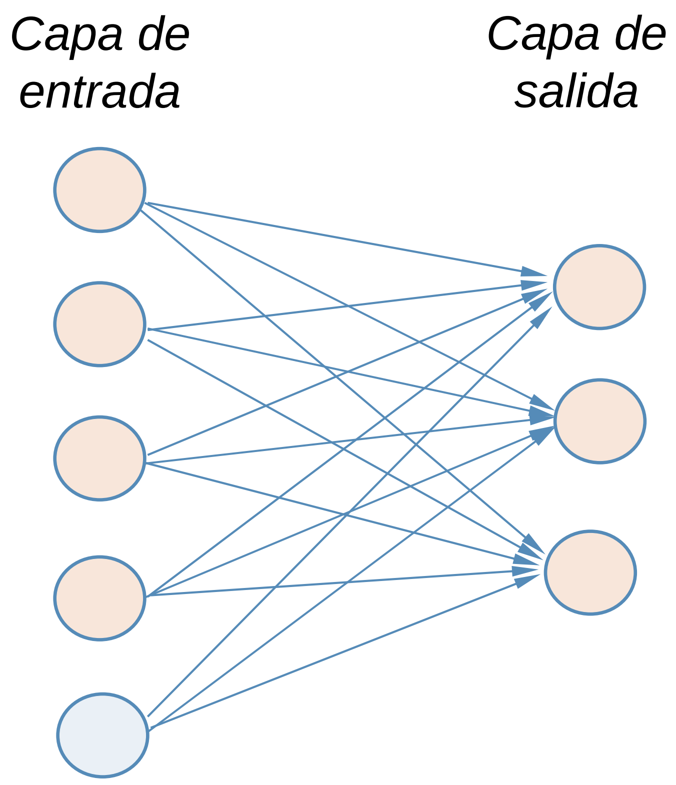
- Sobre una cinta transportadora circulan naranjas y melones. Se busca obtener un clasificador de frutas que facilite su almacenamiento. Para cada fruta se conoce su diámetro, en centímetros y su intensidad de color naranja, medida entre 0 y 255.

- Utilice la información del archivo **FrutasTrain.csv** para entrenar una **neurona no lineal** capaz de reconocer los dos tipos de fruta.

- Compare la manera de obtener la función discriminante de la neurona no lineal con respecto al perceptrón.
  - **NeuronaGral_FRUTAS_RN.ipynb**
  - **Perceptron_FRUTAS_RN.ipynb**

Clasificación con más de 2 clases

- Pueden utilizarse varias neuronas no lineales para resolver un problema de clasificación con más de 2 clases.
- Cada neurona de la capa de salida buscará responder por un valor de clase distinto.
- El error de la capa será la suma de los errores de las neuronas que la forman.



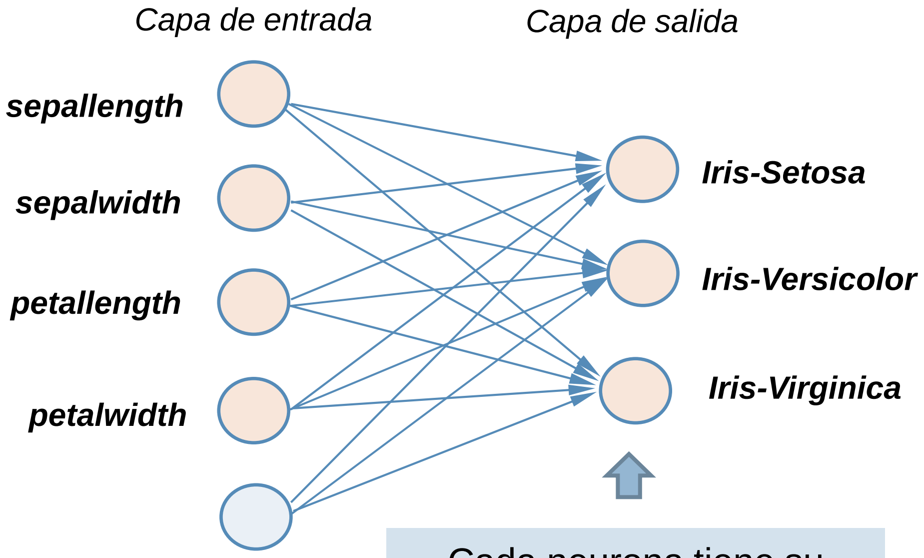
Clasificación de flores de Iris

X

```
[[-1.73,-0.05,-1.38,-1.31],
 [-0.37,-1.62, 0.22, 0.18],
 [ 1.11,-0.05, 0.93, 1.54],
 [-0.99, 0.39,-1.44,-1.31],
 [ 1.73, 1.29, 1.46, 1.81]]
```

Y

```
[[1,0,0],
 [0,1,0],
 [0,0,1],
 [1,0,0],
 [0,0,1]]
```



Cada neurona tiene su propio vector de pesos. Luego **W es una matriz** y **b es un vector**

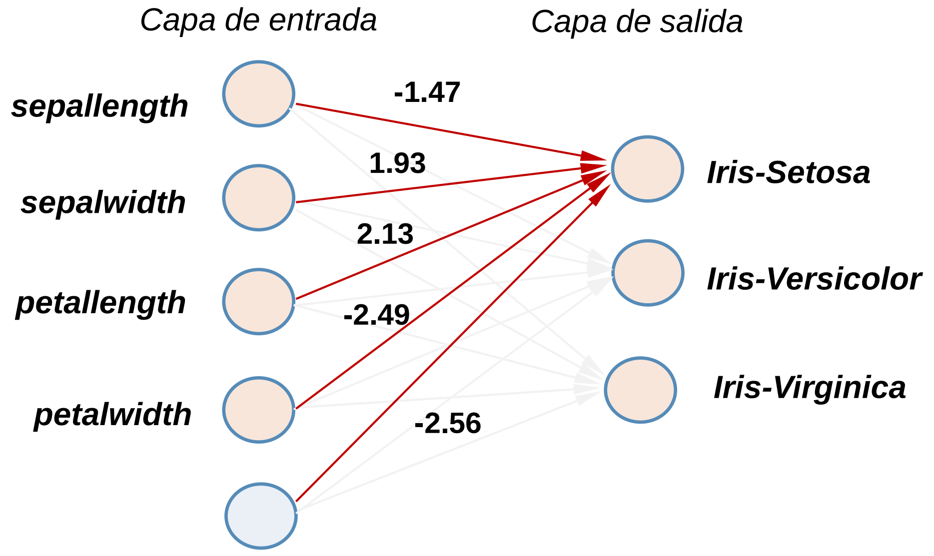
Clasificación de flores de Iris

X

```
[[-1.73,-0.05,-1.38,-1.31],
 [-0.37,-1.62, 0.22, 0.18],
 [ 1.11,-0.05, 0.93, 1.54],
 [-0.99, 0.39,-1.44,-1.31],
 [ 1.73, 1.29, 1.46, 1.81]]
```

Y

```
[[1,0,0],
 [0,1,0],
 [0,0,1],
 [1,0,0],
 [0,0,1]]
```



Inicialmente los pesos de la red **W** y **b** son aleatorios

```
[[-1.47, 1.93, 2.13,-2.49],
 [ 0.79,-1.38, 3.36,-3.57],
 [-1.57,-0.99,-6.17, 3.92]]
```

$\underbrace{\hspace{6em}}_{\mathbf{W}}$

```
[[-2.56],
 [-0.35],
 [-7.03]]
```

$\underbrace{\hspace{3em}}_{\mathbf{b}}$

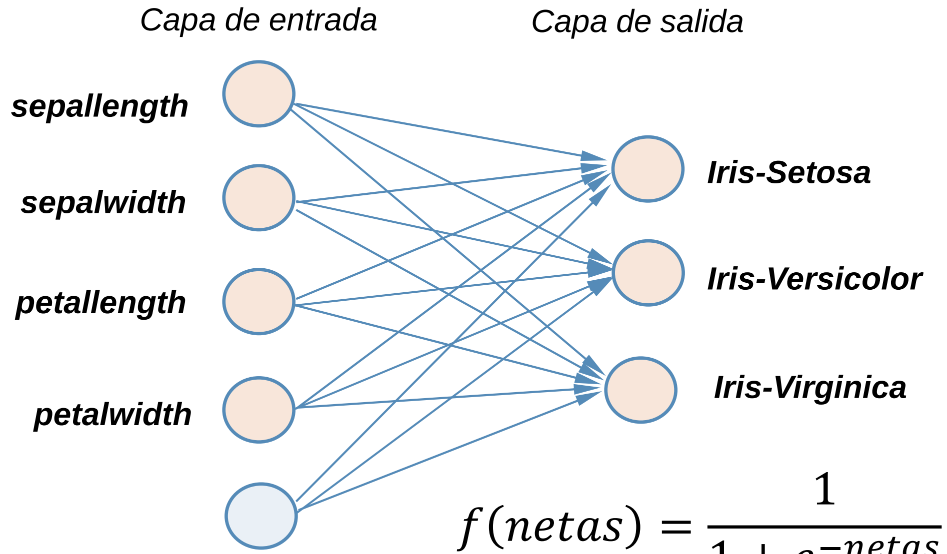
Clasificación de flores de Iris

X

```
[[-1.73,-0.05,-1.38,-1.31],
 [-0.37,-1.62, 0.22, 0.18],
 [ 1.11,-0.05, 0.93, 1.54],
 [-0.99, 0.39,-1.44,-1.31],
 [ 1.73, 1.29, 1.46, 1.81]]
```

Y

```
[[1,0,0],
 [0,1,0],
 [0,0,1],
 [1,0,0],
 [0,0,1]]
```



$$f(netas) = \frac{1}{1 + e^{-netas}}$$

Ingresar el primer ejemplo a la red y calcular su salida

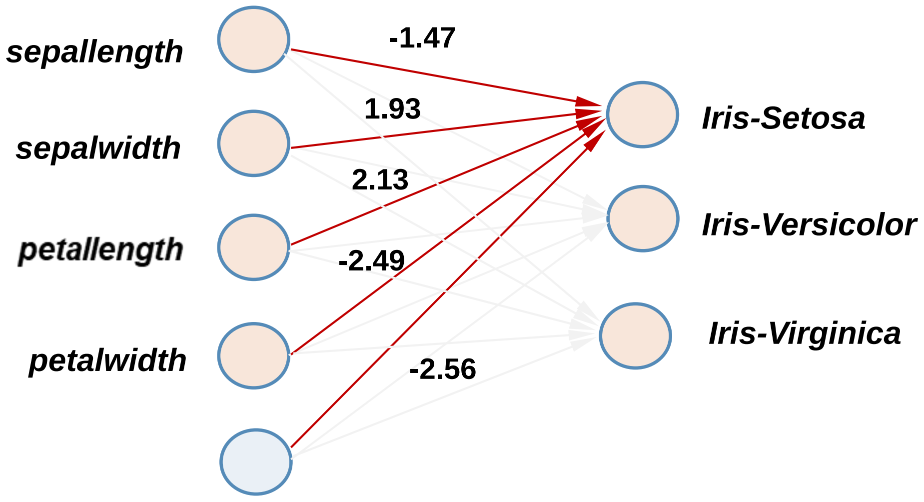
Clasificación de flores de Iris

X

```
[[-1.73,-0.05,-1.38,-1.31],
 [-0.37,-1.62, 0.22, 0.18],
 [ 1.11,-0.05, 0.93, 1.54],
 [-0.99, 0.39,-1.44,-1.31],
 [ 1.73, 1.29, 1.46, 1.81]]
```

Y

```
[[1,0,0],
 [0,1,0],
 [0,0,1],
 [1,0,0],
 [0,0,1]]
```



W

```
[[-1.47, 1.93, 2.13,-2.49],
 [ 0.79,-1.38, 3.36,-3.57],
 [-1.57,-0.99,-6.17, 3.92]]
```

*

x^T

```
[[-1.73],
 [-0.05],
 [-1.38],
 [-1.31]]
```

+

b

```
[[-2.56],
 [-0.35],
 [-7.03]]
```

=

Clasificación de flores de Iris

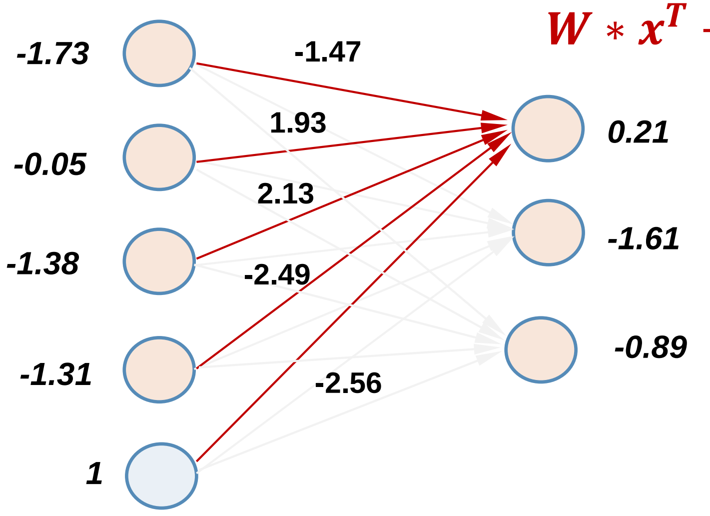
X

```
[[-1.73,-0.05,-1.38,-1.31],
 [-0.37,-1.62, 0.22, 0.18],
 [ 1.11,-0.05, 0.93, 1.54],
 [-0.99, 0.39,-1.44,-1.31],
 [ 1.73, 1.29, 1.46, 1.81]]
```

Y

```
[[1,0,0],
 [0,1,0],
 [0,0,1],
 [1,0,0],
 [0,0,1]]
```

$W * x^T + b$



W

```
[[-1.47, 1.93,-2.13,-2.49],
 [ 0.79,-1.38, 3.36,-3.57],
 [-1.57,-0.99, 6.17, 3.92]]
```

*

x^T

```
[[-1.73],
 [-0.05],
 [-1.38],
 [-1.31]]
```

+

b

```
[[-2.56],
 [-0.35],
 [-7.03]]
```

=

$netas = W * x^T + b$

```
[[  0.21],
 [ -1.61],
 [ -0.89]]
```

Clasificación de flores de Iris

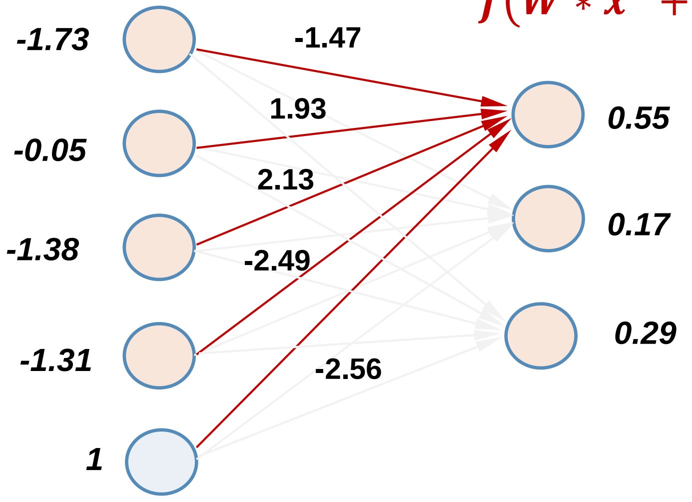
X

```
[[-1.73,-0.05,-1.38,-1.31],
 [-0.37,-1.62, 0.22, 0.18],
 [ 1.11,-0.05, 0.93, 1.54],
 [-0.99, 0.39,-1.44,-1.31],
 [ 1.73, 1.29, 1.46, 1.81]]
```

Y

```
[[1,0,0],
 [0,1,0],
 [0,0,1],
 [1,0,0],
 [0,0,1]]
```

$f(W * x^T + b)$



W

```
[[-1.47, 1.93, 2.13,-2.49],
 [ 0.79,-1.38, 3.36,-3.57],
 [-1.57,-0.99,-6.17, 3.92]]
```

*

x^T

```
[[-1.73],
 [-0.05],
 [-1.38],
 [-1.31]]
```

\+

b

```
[[-2.56],
 [-0.35],
 [-7.03]]
```

=

$netas = W * x^T + b$

```
[[  0.21],
 [ -1.61],
 [ -0.89]]
```

salidas = $f(netas)$ =

```
[[ 1/(1+exp(-0.21))],
 [ 1/(1+exp(1.61))],
 [ 1/(1+exp(0.89))]]
```

=

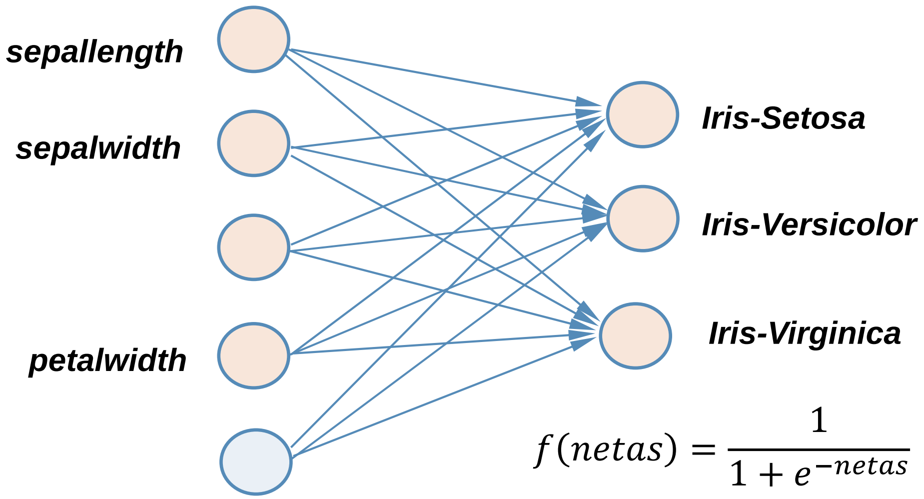
```
[[0.5521]
 [0.1669]
 [0.2921]]
```

Calculemos el error cometido en cada neurona

Clasificación de flores de Iris

X **Y**

```
[[-1.73,-0.05,-1.38,-1.31],    [[1,0,0],
 [-0.37,-1.62, 0.22, 0.18],     [0,1,0],
 [ 1.11,-0.05, 0.93, 1.54],     [0,0,1],
 [-0.99, 0.39,-1.44,-1.31],     [1,0,0],
 [ 1.73, 1.29, 1.46, 1.81]]     [0,0,1]]
```



$$f(netas) = \frac{1}{1 + e^{-netas}}$$

- Error en la respuesta de la red para este ejemplo

```
ErrorSalida =  y.T - salidas

                [[1.0]      [[0.5521]      [[ 0.4479]
ErrorSalida =    [0.0]   -   [0.1669]   =   [-0.1669]
                 [0.0]]      [0.2921]]      [-0.2921]]

                            salidas
```

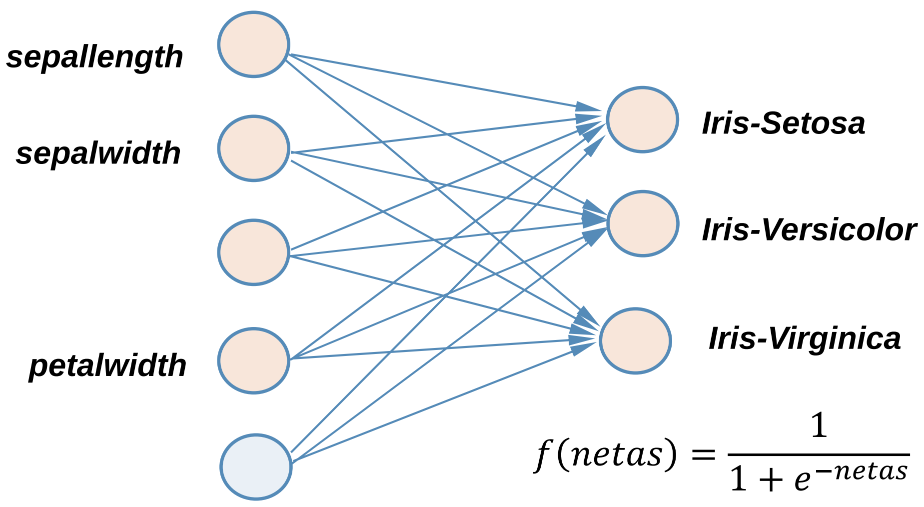
Clasificación de flores de Iris

X

```
[[-1.73,-0.05,-1.38,-1.31],
 [-0.37,-1.62, 0.22, 0.18],
 [ 1.11,-0.05, 0.93, 1.54],
 [-0.99, 0.39,-1.44,-1.31],
 [ 1.73, 1.29, 1.46, 1.81]]
```

Y

```
[[1,0,0],
 [0,1,0],
 [0,0,1],
 [1,0,0],
 [0,0,1]]
```



$$f(netas) = \frac{1}{1 + e^{-netas}}$$

- **Función de Costo: ECM**

  Factores para corregir W y b

```
salidas*(1-salidas)
```

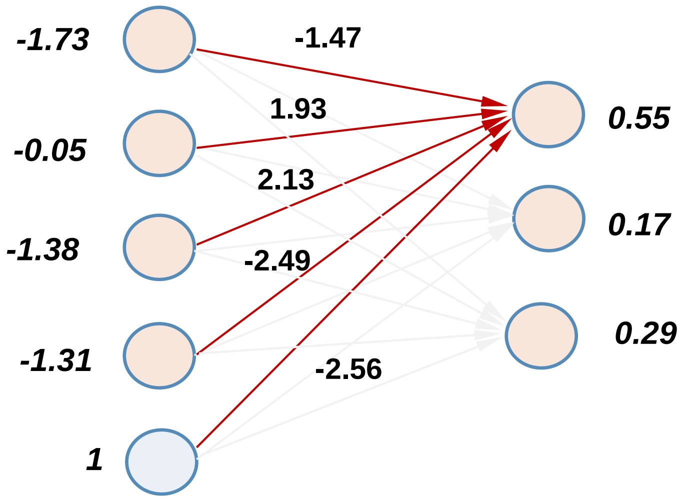
```
delta =   ErrorSalida   .*  derivada_Fun
```

```
           [[ 0.4479]         [[0.2473]         [[ 0.1108]
delta =     [-0.1669]    .*    [0.1390]    =     [-0.0232]
            [-0.2921]]         [0.2068]]         [-0.0604]]
```

Clasificación de flores de Iris

X **Y**

```
[[-1.73,-0.05,-1.38,-1.31],    [[1,0,0],
 [-0.37,-1.62, 0.22, 0.18],     [0,1,0],
 [ 1.11,-0.05, 0.93, 1.54],     [0,0,1],
 [-0.99, 0.39,-1.44,-1.31],     [1,0,0],
 [ 1.73, 1.29, 1.46, 1.81]]     [0,0,1]]
```



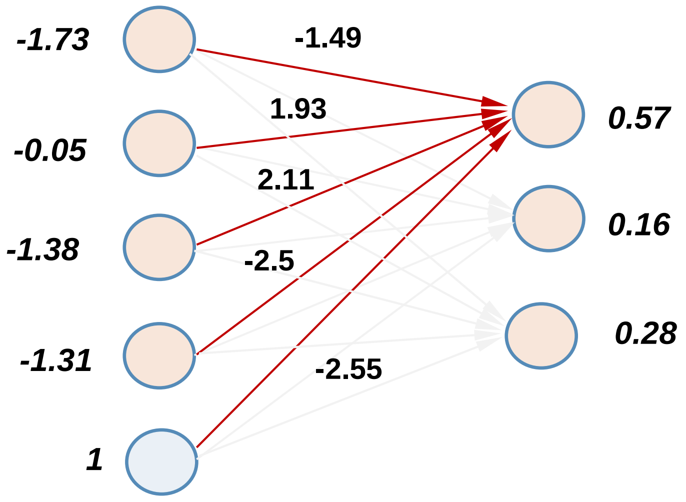
- **Modificación de W y b**

```
W = W + alfa * delta * X

     [[-1.47, 1.93, 2.13,-2.49],         [[ 0.1108]                                 [[-1.49  1.93  2.11 -2.5 ]
W =  [ 0.79,-1.38, 3.36,-3.57], + alfa * [-0.0232]  * [[-1.73,-0.05,-1.38,-1.31]]= [ 0.79 -1.38  3.36 -3.57]
     [-1.57,-0.99,-6.17, 3.92]]           [-0.0604]]                                 [-1.56 -0.99 -6.16  3.93]]

                        [[-2.56],            [[ 0.1108]      [[-2.55]
b = b + alfa * delta =   [-0.35],   + alfa *  [-0.0232]   =   [-0.35]
                         [-7.03]]             [-0.0604]]      [-7.04]]
```

NeuronaNoLineal-Multiclase-IRIS.ipynb

Clasificación de flores de Iris

X	Y
[[-1.73,-0.05,-1.38,-1.31],	[[1,0,0],
[-0.37,-1.62, 0.22, 0.18],	[0,1,0],
[1.11,-0.05, 0.93, 1.54],	[0,0,1],
[-0.99, 0.39,-1.44,-1.31],	[1,0,0],
[1.73, 1.29, 1.46, 1.81]]	[0,0,1]]



- **Modificación de W y b**

```
W = W + alfa * delta * X
```

```
     [[-1.47, 1.93, 2.13,-2.49],          [[ 0.1108]                                   [[-1.49  1.93  2.11 -2.5 ]
W =  [ 0.79,-1.38, 3.36,-3.57], + alfa * [-0.0232]  * [[-1.73,-0.05,-1.38,-1.31]]=  [ 0.79 -1.38  3.36 -3.57]
     [-1.57,-0.99,-6.17, 3.92]]           [-0.0604]]                                   [-1.56 -0.99 -6.16  3.93]]
```

```
                          [[-2.56],            [[ 0.1108]      [[-2.55]
b = b + alfa * delta =     [-0.35],  + alfa *   [-0.0232]   =   [-0.35]
                           [-7.03]]             [-0.0604]]      [-7.04]]
```

*¿Cuánto vale **delta** si la función de Costo es Entropía Cruzada Binaria?*

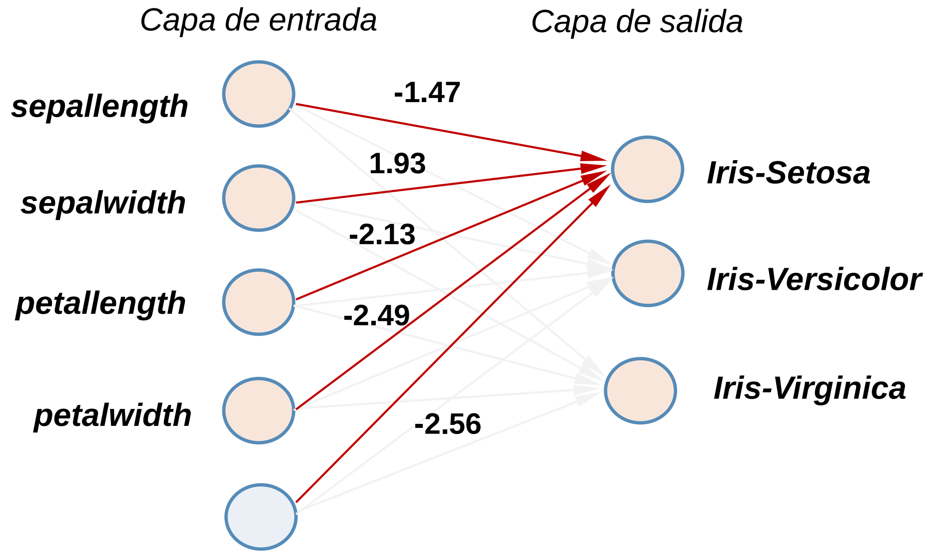
Clasificación de flores de Iris

X

```
[[-1.73,-0.05,-1.38,-1.31],
 [-0.37,-1.62, 0.22, 0.18],
 [ 1.11,-0.05, 0.93, 1.54],
 [-0.99, 0.39,-1.44,-1.31],
 [ 1.73, 1.29, 1.46, 1.81]]
```

Y

```
[[1,0,0],
 [0,1,0],
 [0,0,1],
 [1,0,0],
 [0,0,1]]
```



Para obtener el resultado de la red debe calcularse

$$f(W * x^T + b)$$

siendo f la función de activación

```
[[-1.47, 1.93,-2.13,-2.49],
 [ 0.79,-1.38, 3.36,-3.57],
 [-1.57,-0.99, 6.17, 3.92]]
```

W

```
[[-2.56],
 [-0.35],
 [-7.03]]
```

b

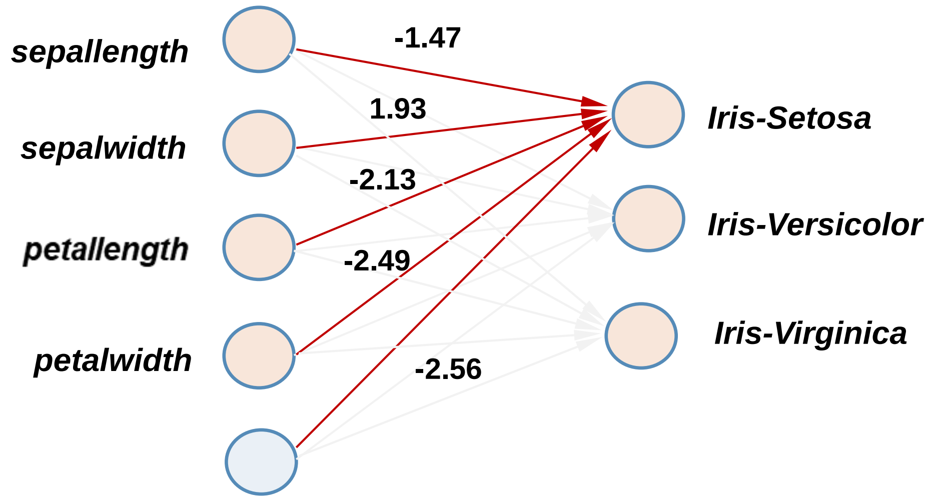
Clasificación de flores de Iris

X

```
[[-1.73,-0.05,-1.38,-1.31],
 [-0.37,-1.62, 0.22, 0.18],
 [ 1.11,-0.05, 0.93, 1.54],
 [-0.99, 0.39,-1.44,-1.31],
 [ 1.73, 1.29, 1.46, 1.81]]
```

Y

```
[[1,0,0],
 [0,1,0],
 [0,0,1],
 [1,0,0],
 [0,0,1]]
```



W $\quad$ x^T $\quad$ **b**

```
[[-1.47, 1.93,-2.13,-2.49],     [[-1.73],      [[-2.56],
 [ 0.79,-1.38, 3.36,-3.57],  *   [-0.05],   +   [-0.35],   =
 [-1.57,-0.99, 6.17, 3.92]]      [-1.38],       [-7.03]]
                                 [-1.31]]
```

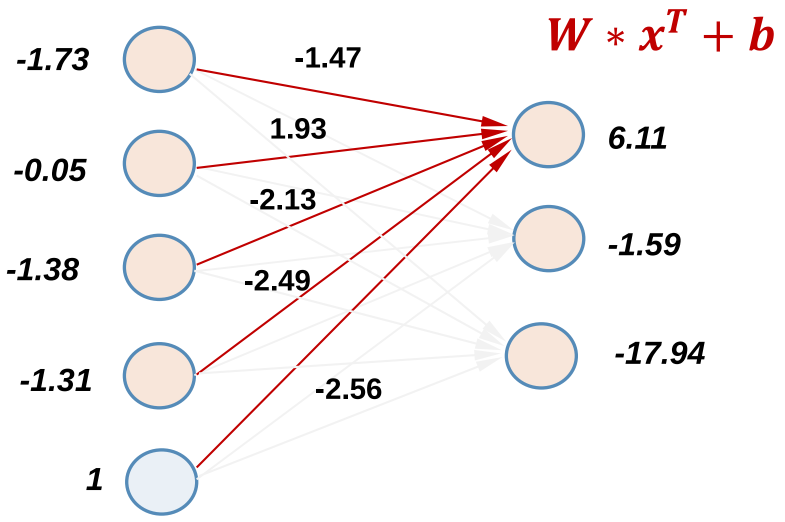
Clasificación de flores de Iris

X

```
[[-1.73,-0.05,-1.38,-1.31],
 [-0.37,-1.62, 0.22, 0.18],
 [ 1.11,-0.05, 0.93, 1.54],
 [-0.99, 0.39,-1.44,-1.31],
 [ 1.73, 1.29, 1.46, 1.81]]
```

Y

```
[[1,0,0],
 [0,1,0],
 [0,0,1],
 [1,0,0],
 [0,0,1]]
```



W

```
[[-1.47, 1.93,-2.13,-2.49],
 [ 0.79,-1.38, 3.36,-3.57],
 [-1.57,-0.99, 6.17, 3.92]]
```

*

x^T

```
[[-1.73],
 [-0.05],
 [-1.38],
 [-1.31]]
```

\+

b

```
[[-2.56],
 [-0.35],
 [-7.03]]
```

=

$W * x^T + b$

```
[[  6.11],
 [ -1.59],
 [-17.94]]
```

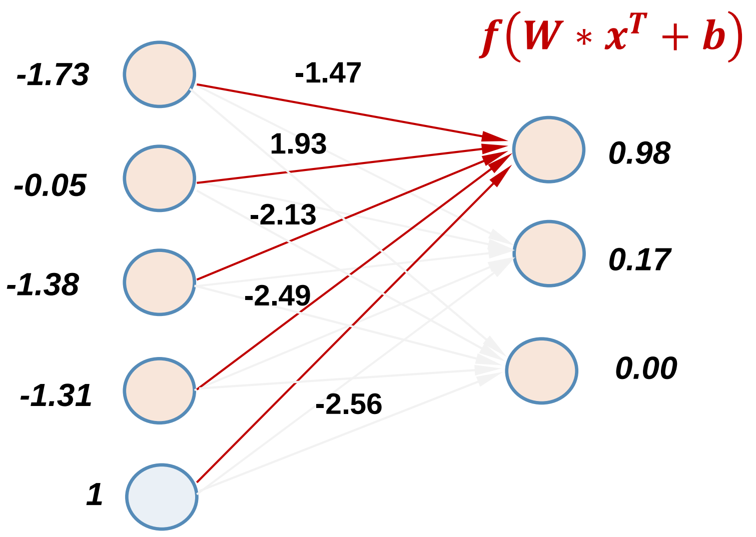
Clasificación de flores de Iris

X

```
[[-1.73,-0.05,-1.38,-1.31],
 [-0.37,-1.62, 0.22, 0.18],
 [ 1.11,-0.05, 0.93, 1.54],
 [-0.99, 0.39,-1.44,-1.31],
 [ 1.73, 1.29, 1.46, 1.81]]
```

Y

```
[[1,0,0],
 [0,1,0],
 [0,0,1],
 [1,0,0],
 [0,0,1]]
```



W $*$ x^T $+$ **b** $=$ $W * x^T + b$

W:

```
[[-1.47, 1.93,-2.13,-2.49],
 [ 0.79,-1.38, 3.36,-3.57],
 [-1.57,-0.99, 6.17, 3.92]]
```

x^T:

```
[[-1.73],
 [-0.05],
 [-1.38],
 [-1.31]]
```

b:

```
[[-2.56],
 [-0.35],
 [-7.03]]
```

$W * x^T + b$:

```
[[  6.11],
 [ -1.59],
 [-17.94]]
```

$f\left(W * x^T + b\right) =$

```
[[ 1/(1+exp( -6.11))],
 [ 1/(1+exp(  1.59))],
 [ 1/(1+exp( 17.94))]]
```

$=$

```
[[ 0.98],
 [ 0.17],
 [ 0.00]]
```

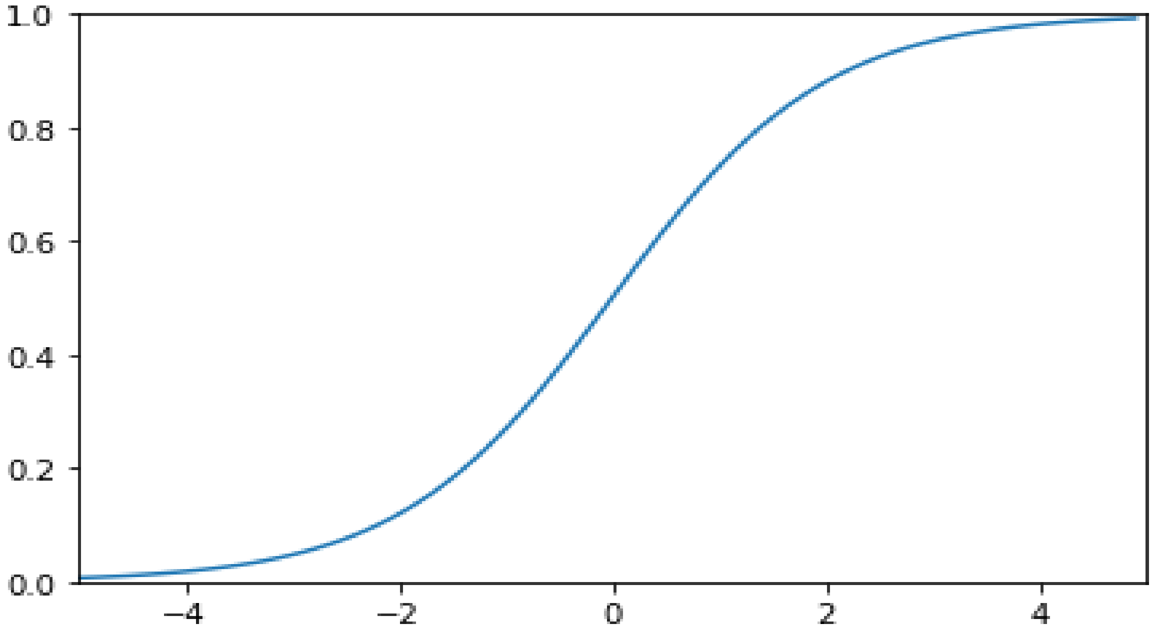
Se interpreta como

```
[[ 1 ],
 [ 0 ],
 [ 0 ]]
```

Función sigmoid()

X	$f(X)$
-5.00	0.01
-4.00	0.02
-3.00	0.05
-2.00	0.12
-1.39	0.20
-1.00	0.27
0.00	0.50
1.00	0.73
1.39	0.80
2.00	0.88
3.00	0.95
4.00	0.98
5.00	0.99

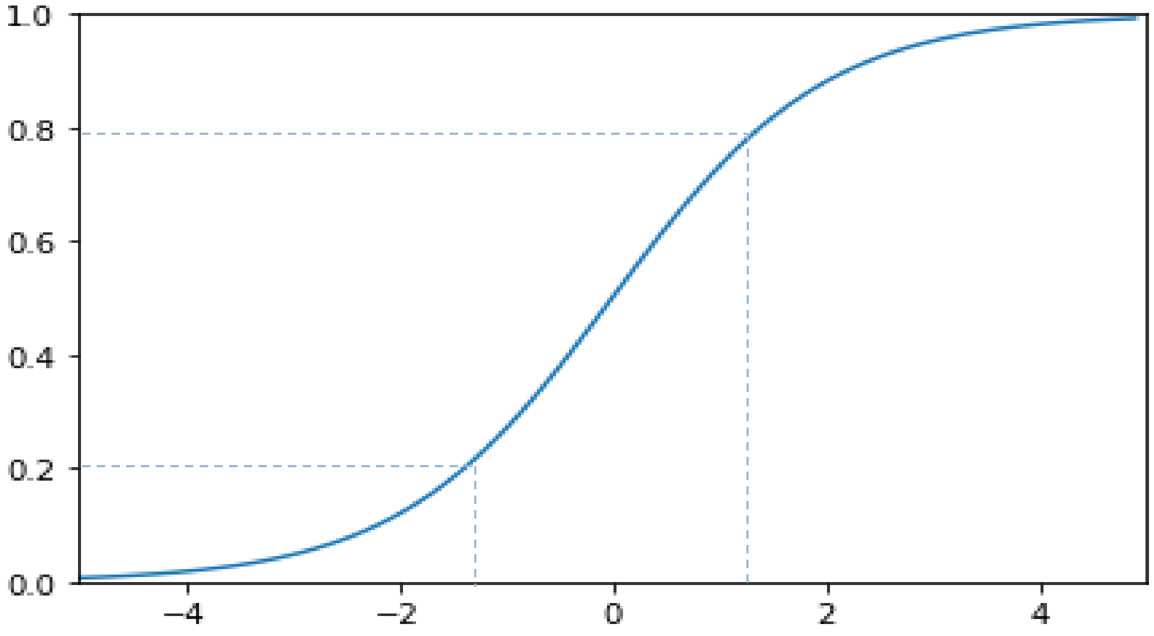
$$f(x) = \frac{1}{1 + \exp(-x)}$$



Función sigmoid()

X	$f(X)$
-5.00	0.01
-4.00	0.02
-3.00	0.05
-2.00	0.12
-1.39	**0.20**
-1.00	0.27
0.00	0.50
1.00	0.73
1.39	**0.80**
2.00	0.88
3.00	0.95
4.00	0.98
5.00	0.99

$$f(x) = \frac{1}{1 + \exp(-x)}$$



Funciones de costo

- Error cuadrático medio

$$C = \frac{1}{n}\sum_{n}(t - y)^2 = \frac{1}{n}\sum_{n}(t - f(neta))^2$$

- Entropía cruzada binaria

$$C = -\frac{1}{n}\sum_{n}[t \ln y + (1 - t)\ln(1 - y)]$$

Entropía cruzada binaria

- Es una función de costo que puede usarse con neuronas con función de activación sigmoide entre 0 y 1

$$C = -\frac{1}{n}\sum_{n}[\ t \ln y + (1-t)\ \ln(1-y)\ ]$$

donde

- t es el valor binario esperado
- $y = 1/(1 + e^{-\sum x_i w_i})$ es la salida de la neurona

- Ver que es una función de costo
  - C > 0
  - C tiende a 0 (cero) a medida que la neurona aprende la salida deseada.

Derivada de la entropía cruzada binaria

$$C = -\frac{1}{n}\sum_n [t \ln y + (1-t)\ \ln(1-y)]$$

$$\frac{\partial C}{\partial w_j} = -\frac{1}{n}\sum_n \left(\frac{t}{y} - \frac{1-t}{1-y}\right)\frac{\partial\, y}{\partial w_j}$$

$f(neta)$

$$\frac{\partial C}{\partial w_j} = -\frac{1}{n}\sum_n \left(\frac{t}{f(neta)} - \frac{1-t}{1-f(neta)}\right)\frac{\partial f(neta)}{\partial w_j}$$

Derivada de la entropía cruzada binaria

$$C = -\frac{1}{n}\sum_{n}[t \ln y + (1-t)\ \ln(1-y)]$$

$$\frac{\partial C}{\partial w_j} = -\frac{1}{n}\sum_{n}\left(\frac{t}{f(neta)} - \frac{1-t}{1-f(neta)}\right)\frac{\partial f(neta)}{\partial w_j}$$

$$\frac{t-f(neta)}{f(neta)(1-f(neta))}$$

$$f'(neta)x_j$$

Derivada de la entropía cruzada binaria

$$C = -\frac{1}{n}\sum_{n}[t \ln y + (1-t)\ \ln(1-y)]$$

$$\frac{\partial C}{\partial w_j} = -\frac{1}{n}\sum_{n}\left(\frac{t - f(neta)}{f(neta)(1-f(neta))}\right) f'(neta)\, x_j$$

Si $f(neta) = \frac{1}{1+e^{-neta}}$, $f'(neta) = f(neta)\big(1 - f(neta)\big)$

Derivada de la entropía cruzada binaria

$$C = -\frac{1}{n}\sum_{n}[t \ln \mathrm{y} + (1-t)\ \ln(1-y)]$$

$$\frac{\partial C}{\partial w_j} = -\frac{1}{n}\sum_{n}\left(\frac{t - f(neta)}{f(neta)(1 - f(neta))}\right) f'(neta)\, x_j$$

$$\frac{\partial C}{\partial w_j} = -\frac{1}{n}\sum_{n}(t - f(neta))\, x_j$$

Derivada de la entropía cruzada binaria

$$C = -\frac{1}{n}\sum_{n}[t \ln \mathrm{y} + (1-t)\ \ln(1-y)]$$

$$\frac{\partial C}{\partial w_j} = -\frac{1}{n}\sum_{n}\left(\frac{t - f(neta)}{f(neta)(1 - f(neta))}\right) f'(neta)\, x_j$$

$$\frac{\partial C}{\partial w_j} = -\frac{1}{n}\sum_{n}(t - y)\, x_j$$

Derivada de la entropía cruzada binaria

$$C = -\frac{1}{n}\sum_{n}[t \ln \mathrm{y} + (1-t)\ \ln(1-y)]$$

$$\frac{\partial C}{\partial w_j} = -\frac{1}{n}\sum_{n}(t-y)\, x_j$$

- Si utilizamos descenso de gradiente estocástico sólo se mide el error cometido en el ejemplo k

$$\frac{\partial C_k}{\partial w_j} = -(t_k - y_k)\, x_j$$

Gradiente de la función de costo ECM calculado sobre el k-ésimo ejemplo

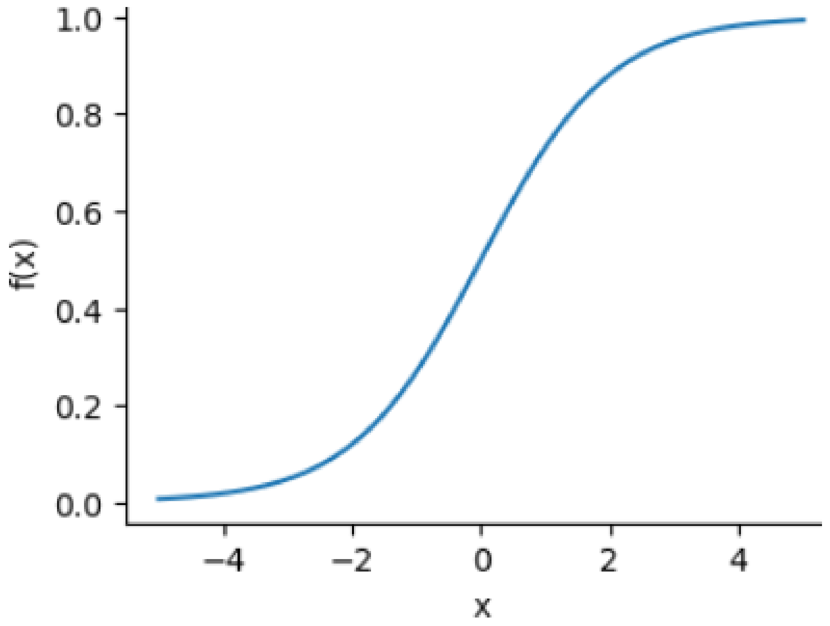
$$\frac{\partial ECM_k}{\partial w_j} = -(t_k - y_k)\boxed{f'(neta)}x_j$$

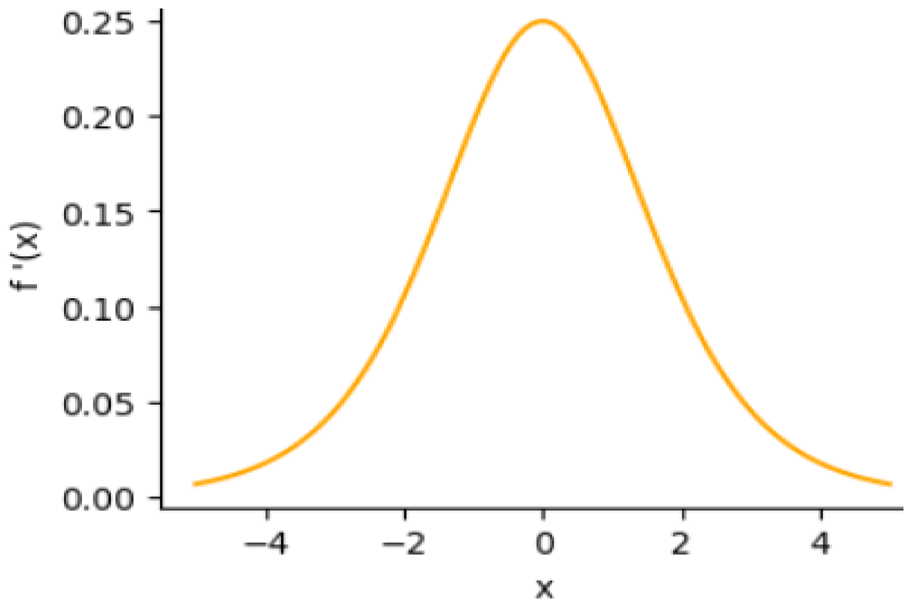
¿Qué valores toma f'(neta) cuando se trata de la función sigmoide entre 0 y 1?

Función sigmoide

$$f(x) = \frac{1}{1 + \exp(-x)}$$

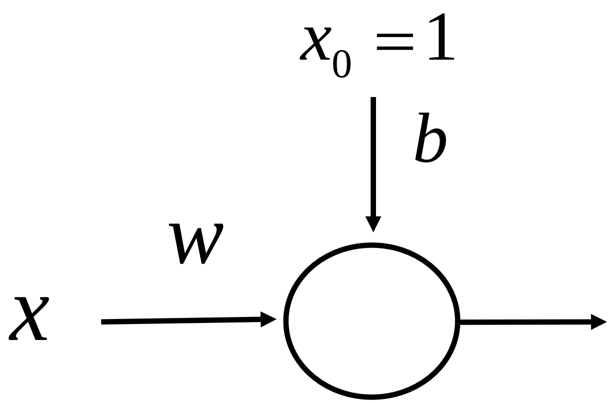
$$f'(x) = f(x) * (1 - f(x))$$





Ejemplo: Entrene una neurona con función de activación sigmoide entre 0 y 1 para que reciba un 1 y responda 0

- Usando como Función de Costo el **Error Cuadrático Medio** (ECM)



$$y = \frac{1}{1+e^{-(wx+b)}}$$

$x = 1$ (entrada)

t $= 0$ (salida esperada)

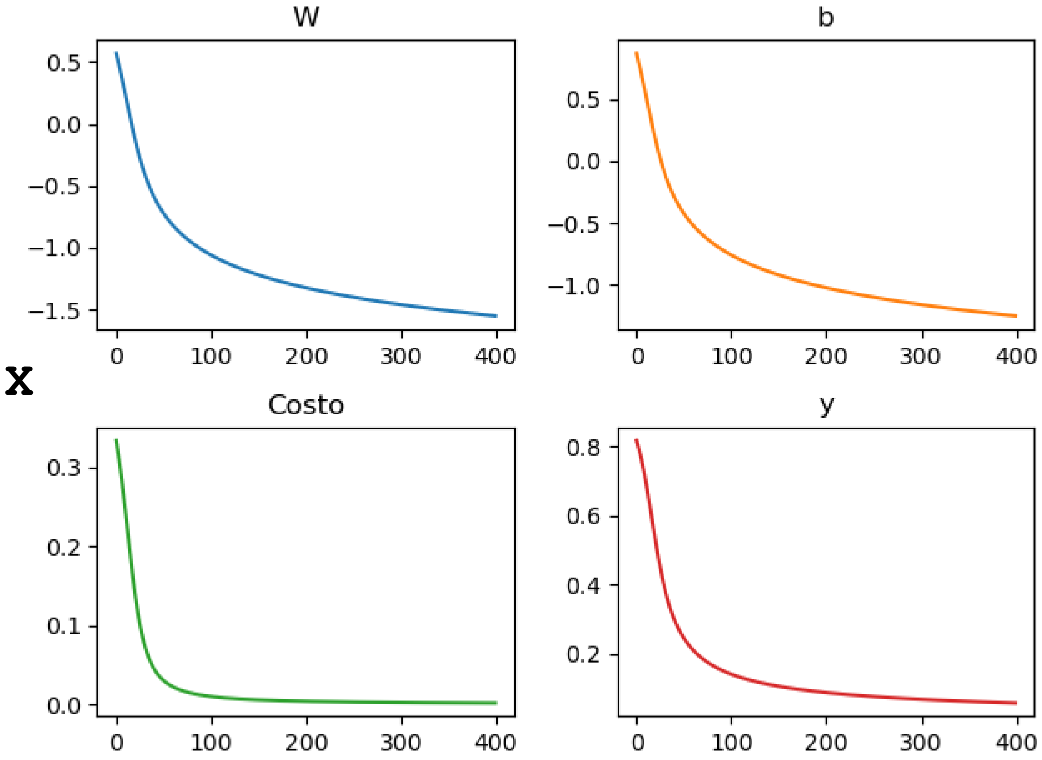
Función de costo (para 1 ejemplo)

$$C = \frac{(t-y)^2}{2}$$

$$\frac{\partial C}{\partial w} = -(t-y)\,\underbrace{[y\,(1-y)]}_{f'(neta)}\;x$$

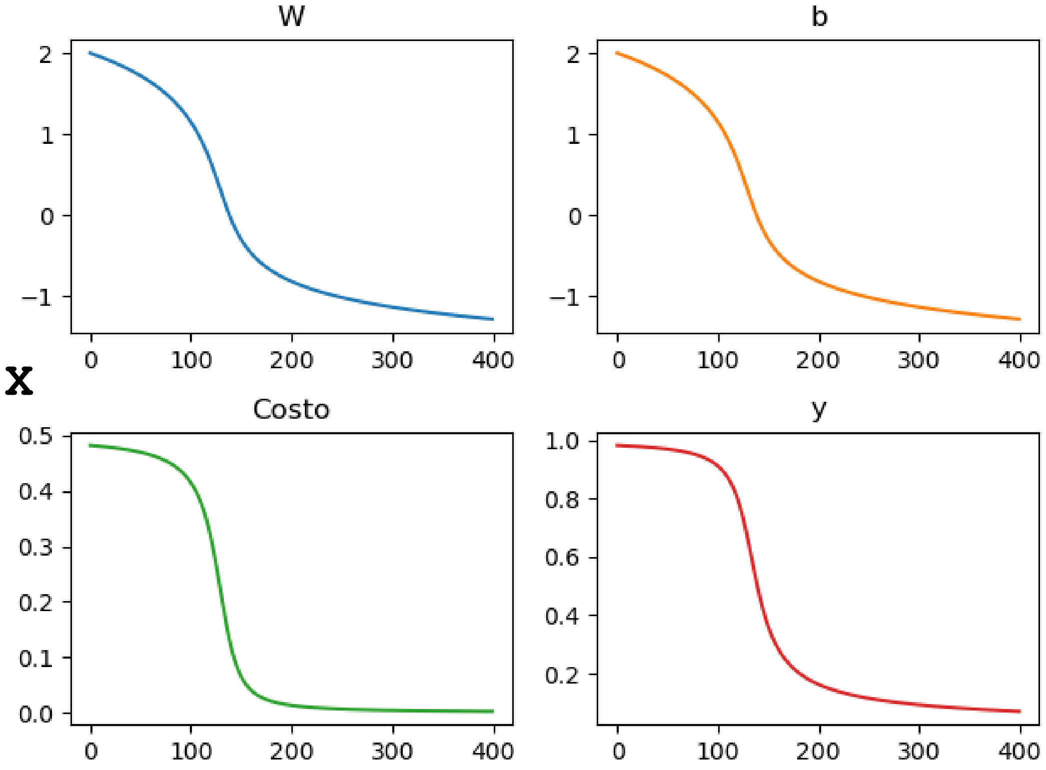
NeuronaGral_1Ej.ipynb

```
X = 1
T = 0
W = 0.6
b = 0.9
MAX_ITE, alfa = 2000, 0.25
ite = 0
C_ant, Costo = 0, 1
while (ite<MAX_ITE) and (np.abs(C_ant-Costo)>10e-05):
    C_ant = Costo

    neta = W * X + b
    y = 1.0/(1+np.exp(-neta))

    Error = T - y

    Costo = (Error**2)/2

    gradiente_W = - Error * ( y * (1-y)) * X
    gradiente_b = - Error * ( y * (1-y))

    W = W - alfa * gradiente_W
    b = b - alfa * gradiente_b

    ite = ite + 1
```



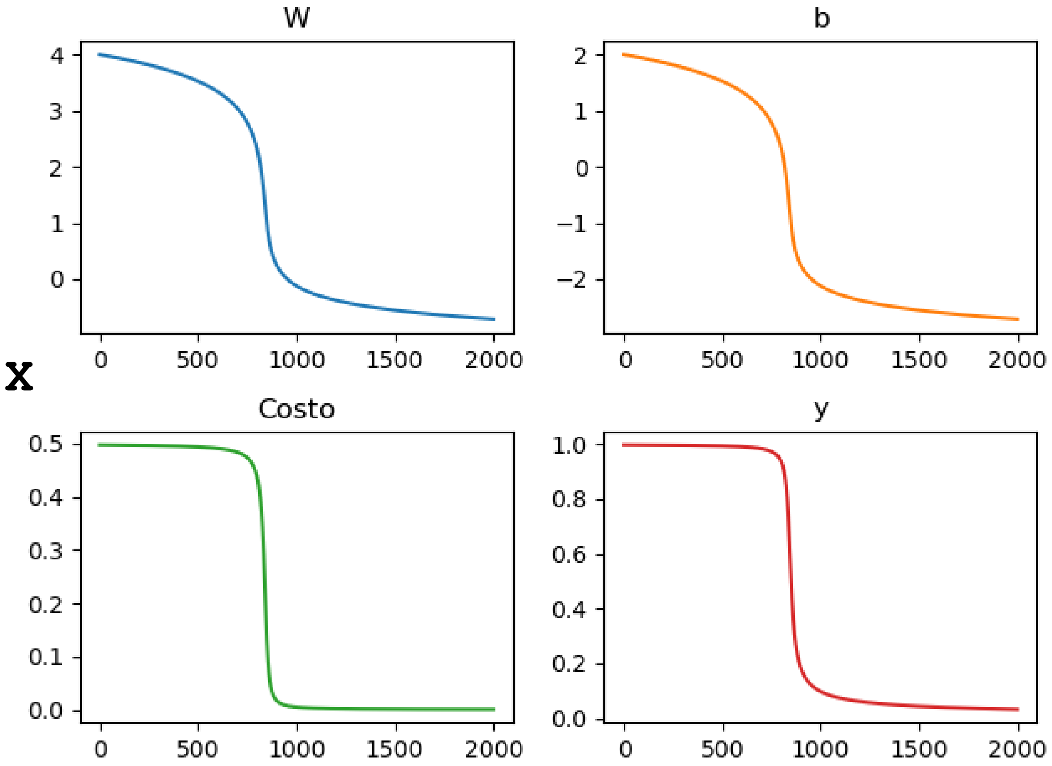
NeuronaGral_1Ej.ipynb

```
X = 1
T = 0
W = 2
b = 2
MAX_ITE, alfa = 2000, 0.25
ite = 0
C_ant, Costo = 0, 1
while (ite<MAX_ITE) and (np.abs(C_ant-Costo)>10e-05):
    C_ant = Costo

    neta = W * X + b
    y = 1.0/(1+np.exp(-neta))

    Error = T - y

    Costo = (Error**2)/2

    gradiente_W = - Error * ( y * (1-y)) * X
    gradiente_b = - Error * ( y * (1-y))

    W = W - alfa * gradiente_W
    b = b - alfa * gradiente_b

    ite = ite + 1
```



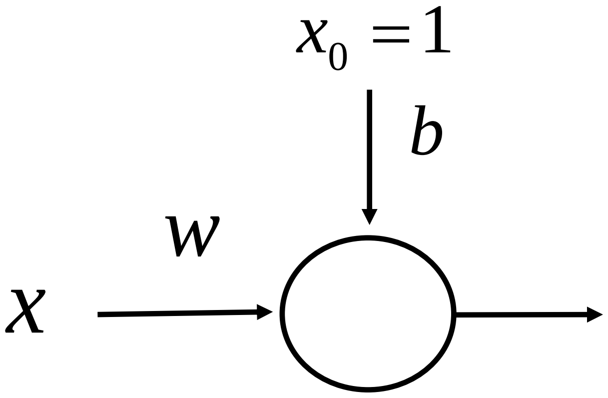
NeuronaGral_1Ej.ipynb

```
X = 1
T = 0
W = 4
b = 2
MAX_ITE, alfa = 2000, 0.25
ite = 0
C_ant, Costo = 0, 1
while (ite<MAX_ITE) and (np.abs(C_ant-Costo)>10e-05):
    C_ant = Costo
    neta = W * X + b
    y = 1.0/(1+np.exp(-neta))
    Error = T - y
    Costo = (Error**2)/2
    gradiente_W = - Error * ( y * (1-y)) * X
    gradiente_b = - Error * ( y * (1-y))
    W = W - alfa * gradiente_W
    b = b - alfa * gradiente_b
    ite = ite + 1
```



Ejemplo: Entrene una neurona con función de activación sigmoide entre 0 y 1 para que reciba un 1 y responda 0

- Función de Costo: **Entropía Cruzada Binaria** (EC_binaria)



$$y = \frac{1}{1 + e^{-(wx+b)}}$$

$x = 1$ (entrada)

t $= 0$ (salida esperada)

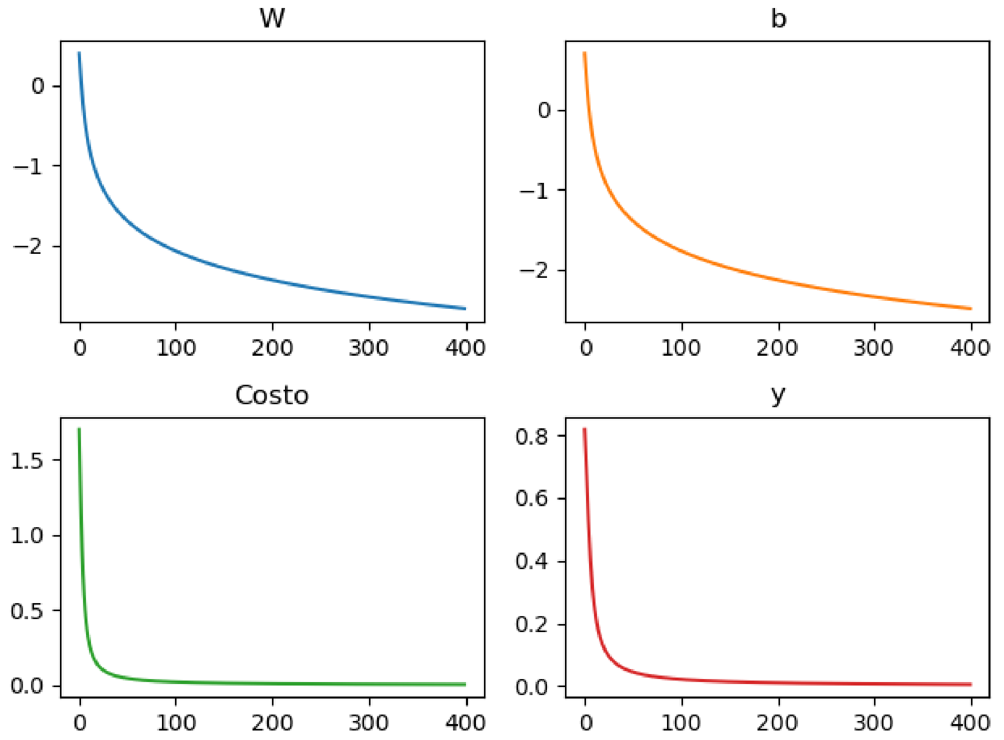
Función de costo (para 1 ejemplo)

$$C = -(t \ln y + (1-t)\ln(1-y))$$

$$\frac{\partial C}{\partial w} = -(t-y)\ x$$

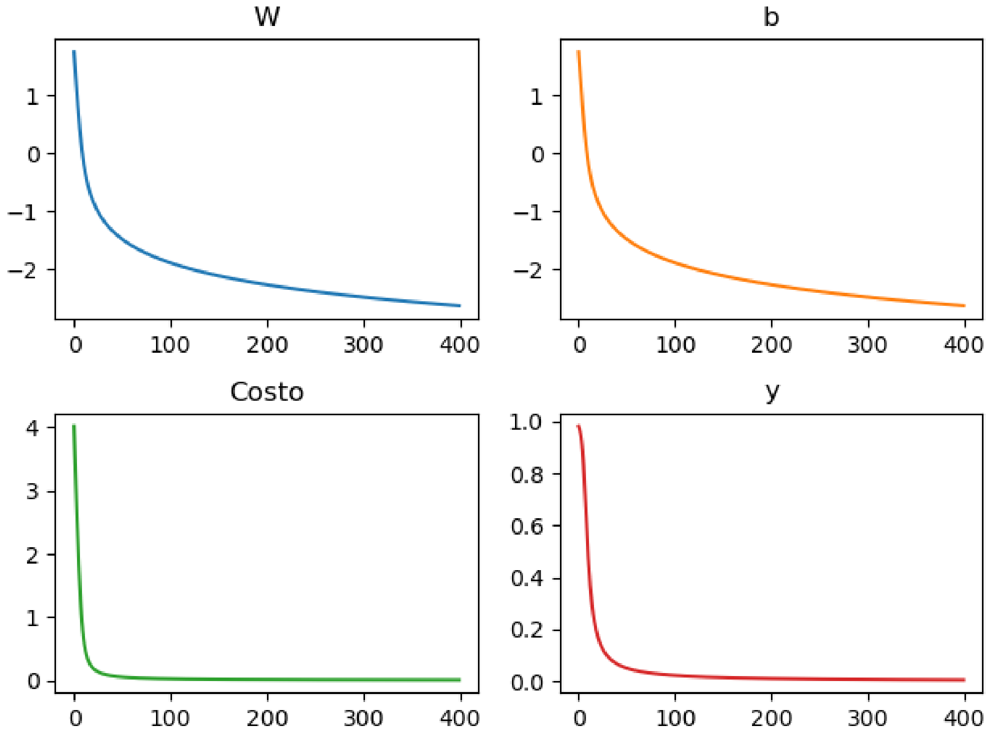
NeuronaGral_1Ej_EC.ipynb

```
X = 1
T = 0
W = 0.6
b = 0.9
MAX_ITE, alfa = 2000, 0.25
ite = 0
C_ant, Costo = 0, 1
while (ite<MAX_ITE) and (np.abs(C_ant-Costo)>10e-05):
    C_ant = Costo

    neta = W * X + b
    y = 1.0/(1+np.exp(-neta))
    Error = T - y

    Costo = -(T*np.log(y)+(1-T)*np.log(1-y))

    gradiente_W = - Error * X
    gradiente_b = - Error

    W = W - alfa * gradiente_W
    b = b - alfa * gradiente_b

    ite = ite + 1
```



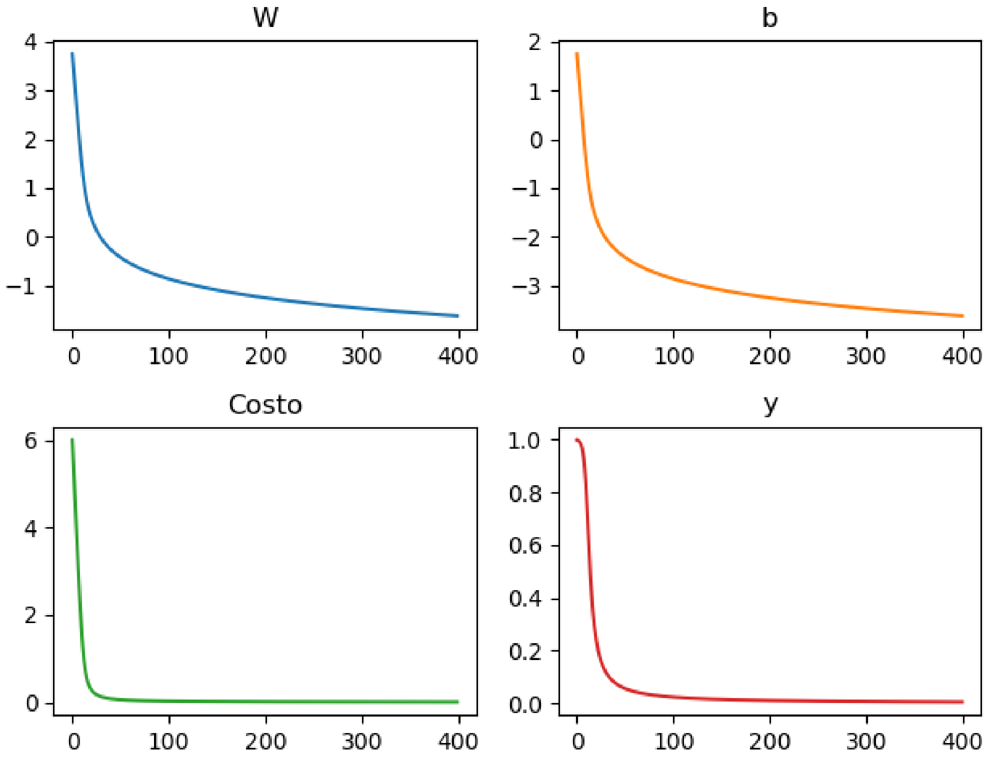
NeuronaGral_1Ej_EC.ipynb

```
X = 1
T = 0
W = 2
b = 2
MAX_ITE, alfa = 2000, 0.25
ite = 0
C_ant, Costo = 0, 1
while (ite<MAX_ITE) and (np.abs(C_ant-Costo)>10e-05):
    C_ant = Costo

    neta = W * X + b
    y = 1.0/(1+np.exp(-neta))
    Error = T - y

    Costo = -(T*np.log(y)+(1-T)*np.log(1-y))

    gradiente_W = - Error * X
    gradiente_b = - Error

    W = W - alfa * gradiente_W
    b = b - alfa * gradiente_b

    ite = ite + 1
```



NeuronaGral_1Ej_EC.ipynb

```
X = 1
T = 0
W = 4
b = 2
MAX_ITE, alfa = 2000, 0.25
ite = 0
C_ant, Costo = 0, 1
while (ite<MAX_ITE) and (np.abs(C_ant-Costo)>10e-05):
    C_ant = Costo

    neta = W * X + b
    y = 1.0/(1+np.exp(-neta))
    Error = T - y

    Costo = -(T*np.log(y)+(1-T)*np.log(1-y))

    gradiente_W = - Error * X
    gradiente_b = - Error

    W = W - alfa * gradiente_W
    b = b - alfa * gradiente_b

    ite = ite + 1
```



Funciones de costo

Entropía Cruzada Binaria

- **Mejor ajuste** para problemas de clasificación binaria.
- Produce **gradientes más grandes** cuando las predicciones están muy alejadas de las etiquetas verdaderas, acelerando el entrenamiento.

*La **entropía cruzada** es más adecuada para clasificación binaria, ya que aprovecha la naturaleza probabilística de la sigmoide.*

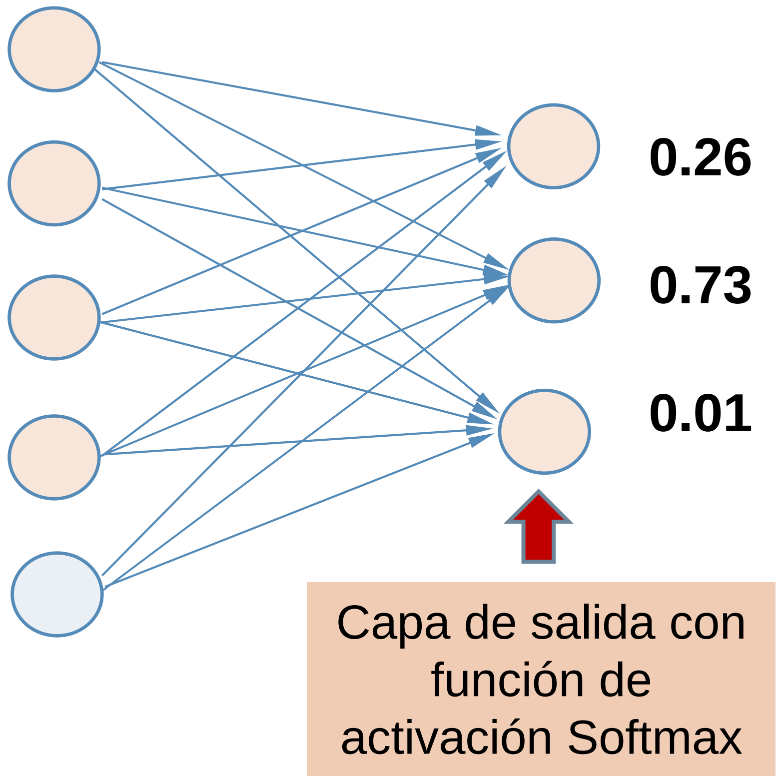
Error Cuadrático Medio

- Usualmente empleado en problemas de **regresión**.
- Gradientes más pequeños cuando la salida está cerca de la sigmoide, lo que puede provocar **aprendizaje más lento**.

*El **ECM**, aunque puede usarse en clasificación, no genera gradientes tan eficientes, especialmente cuando las predicciones son extremas (cercanas a 0 o 1).*

Función Softmax

- Se utiliza como función de activación en la última capa para normalizar la salida de la red a una distribución de probabilidad.

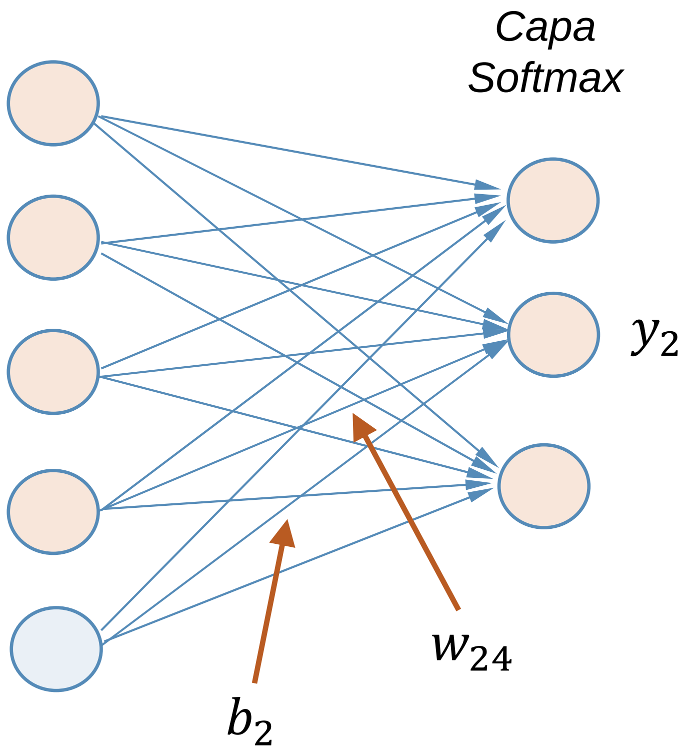


Capa softmax

$$neta_j = \sum_i w_{ji}\, x_i + b_j$$

$$y_j = \frac{e^{neta_j}}{\sum_k e^{neta_k}}$$

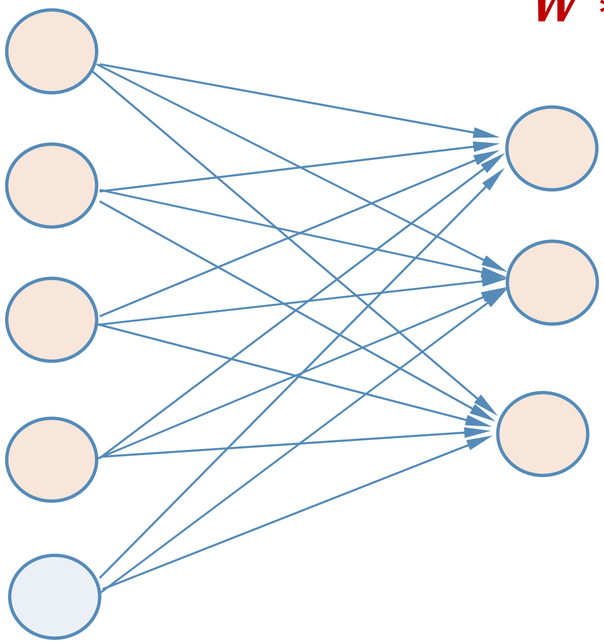
- La salida de la capa es una distribución de probabilidad
  - $y_j > 0 \quad j = 1..k$
  - $\sum_j y_j = 1$



Ver que el incremento en algún y_j producirá disminuciones en el resto

Función Softmax

- Ejemplo



$W * x^T + b$

5

4

2

Función sigmoide

$f(W * x^T + b)$

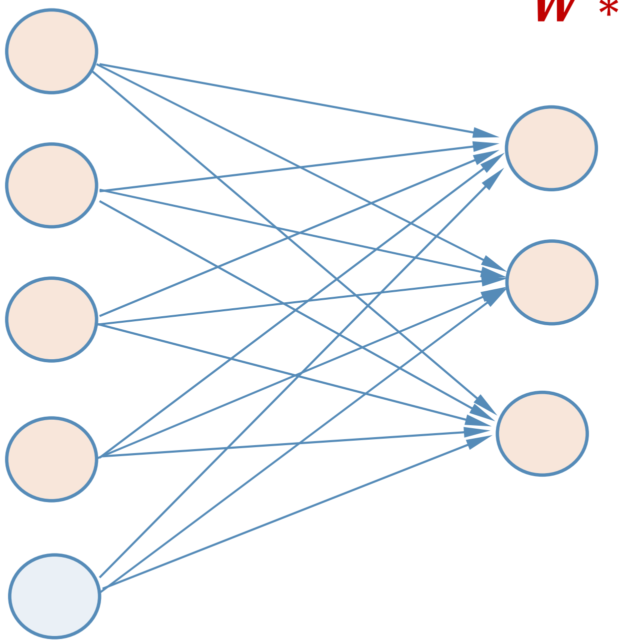
0.99

0.98

0.88

Función Softmax

- Ejemplo



$W * x^T + b$

$f(W * x^T + b)$

$W * x^T + b$		$f(W * x^T + b)$
5	***Función softmax***	**0.71**
4	$\longrightarrow$	**0.26**
2		**0.04**

Función softmax

Entrada

Propagación hacia adelante →

Salidas de la propagación hacia adelante

cat	dog	horse
5	4	2
4	2	8
4	4	1

Función sigmoide →

Salidas de la capa con función sigmoide

cat	dog	horse
0.99	0.98	0.88
0.98	0.88	0.99
0.98	0.98	0.73

Función softmax

Neta	sigmoid	exp(neta)	softmax
5	0,99	148,41	0,71
4	0,98	54,60	0,26
2	0,88	7,39	0,04

Entrada

Propagación hacia adelante →

Salidas de la propagación hacia adelante

cat	dog	horse
5	4	2
4	2	8
4	4	1

Función Softmax →

Salidas de la capa Softmax

cat	dog	horse
0.71	0.26	0.04
0.02	0.00	0.98
0.49	0.49	0.02

Capa Softmax

$$y_j = \frac{e^{neta_j}}{\sum_k e^{neta_k}}$$

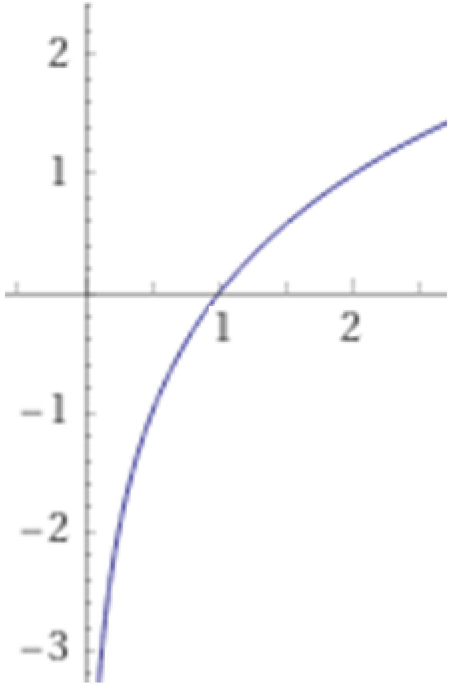
- **Función de costo: Entropía cruzada categórica**

$$C = -\sum_k t_k \ln y_k$$

donde $\boldsymbol{t}$ es un vector binario que vale 1 sólo en la posición correspondiente al valor de clase esperado. Luego

$$C = -\ln y_s$$

s es la neurona correspondiente al valor de clase esperado



Capa Softmax

$$y_j = \frac{e^{neta_j}}{\sum_k e^{neta_k}}$$

- **Función de costo: Entropía cruzada categórica**

$$C = -\ln y_s$$

S es la neurona correspondiente al valor de clase esperado

- Derivada de la función de costo

$$\frac{\partial C}{\partial w_{jk}} = -(t_j - y_j)\, x_k \qquad \frac{\partial C}{\partial b_j} = -(t_j - y_j)$$

Coincide con la derivada de la entropía cruzada binaria

Capa softmax

_RNMulticlase_IRIS_Softmax_EC_RN.ipynb_

- **Función de costo: Entropía cruzada categórica**

Entrada | **Salida Softmax** | **Loss, L(a)**

cat	dog	horse
0.71	0.26	0.04
0.02	0.00	0.98
0.49	0.49	0.02

$-\log(y_s)$ *en la clase correcta* →

NLL
0.34
0.02
0.71

La clase correcta está pintada de rojo

Total = 1.07

- Sólo se evalúa en la neurona correspondiente a la salida esperada.
- Cuando más cerca está de 1 menor será el error.
- A menor valor de la neurona softmax correspondiente a la clase correcta, mayor error.

ClassRNMulticlase.py

nn = RNMulticlase (alpha=0.01, n_iter=50, cotaE=10E-07, FUN=‘sigmoid’, COSTO='ECM', random_state=None)

- Parámetros de entrada
  - **alpha**: valor en el intervalo (0, 1] que representa la velocidad de aprendizaje.
  - **n_iter**: máxima cantidad de iteraciones a realizar.
  - **cotaE**: termina si la diferencia entre dos errores consecutivos es menor que este valor.
  - **FUN:** función de activación – ‘sigmoid’, ‘tanh’, ‘softmax’.
  - **COSTO**: función de costo – ‘ECM’, ‘EC_binaria’, ‘EC’
  - **random_state:** None si los pesos se inicializan en forma aleatoria, un valor entero para fijar la semilla

ClassRNMulticlase.py

nn = RNMulticlase(alpha=0.01, n_iter=50, cotaE=10E-07, FUN='sigmoid',
COSTO='ECM', random_state=None)

nn.fit(X, T)

- Parámetros de entrada
  - **X** : arreglo de NxM donde N es la cantidad de ejemplos y M la cantidad de atributos.
  - **T** : arreglo de NxK donde N es la cantidad de ejemplos y K es la cantidad de neuronas de salida
- Retorna
  - **w_** : arreglo de KxM siendo K el tamaño de la capa de salida y M la cantidad de atributos de entrada.
  - **b_** : arreglo de K elementos formado por los bias de cada una de las K neuronas de la capa de salida.
  - **errors_**: errores cometidos en cada iteración.

ClassRNMulticlase.py

Y = nn.predict_nOut(X)

- Parámetros de entrada
  - **X** : arreglo de NxM donde N es la cantidad de ejemplos y M la cantidad de atributos.

- Retorna: un arreglo con el resultado de aplicar la neurona general entrenada previamente con fit() a la matriz de ejemplos X.
  - **Y** : arreglo de NxK donde N es la cantidad de ejemplos y K es la cantidad de neuronas de salida con valores continuos.

ClassRNMulticlase.py

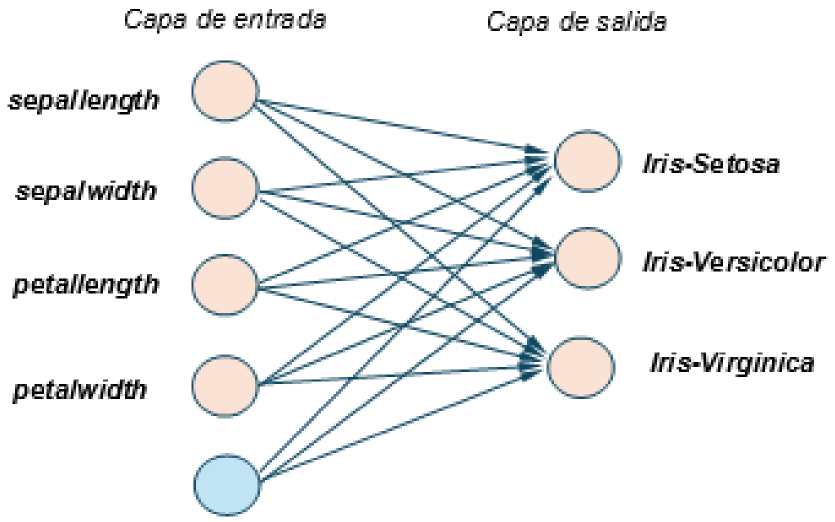
Y = nn.predict(X)

- Parámetros de entrada
  - **X** : arreglo de NxM donde N es la cantidad de ejemplos y M la cantidad de atributos.

- Retorna: un arreglo con el resultado de aplicar la neurona general entrenada previamente con fit() a la matriz de ejemplos X.
  - **Y** : arreglo de N elementos indicando para cada ejemplo el número de la clase predicha.

RNMulticlase_IRIS_RN.ipynb

Sigmoide vs Softmax



- Si la función de activación es **sigmoide**
  - Las neuronas de salida aprenden en forma independiente.
  - La salida no tiene que sumar 1.
  - Permite interpretar cada neurona como una probabilidad independiente.
  - Puede ser útil si las clases no son mutuamente excluyentes.
  - Es conveniente utilizar como función de costo la Entropía cruzada binaria (gradientes más grandes).

- Si la función de activación es **Softmax**
  - Las neuronas de salida aprenden en forma conjunta.
  - La suma de las salidas siempre es 1, lo que permite interpretar las salidas como una distribución de probabilidad.
  - Muy útil cuando las clases son mutuamente excluyentes.
  - Utiliza como función de costo la Entropía Cruzada Categórica.