

Aprendizaje Profundo

Marzo de 2026

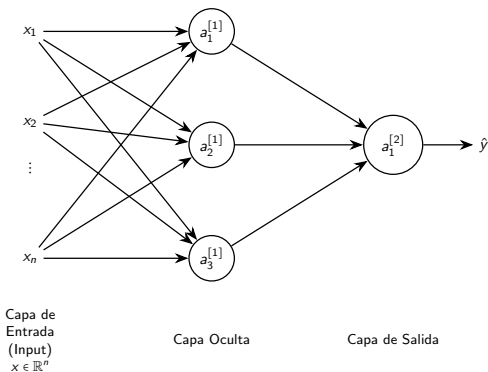
Contenido

- 1 Redes con dos capas
- 2 Redes Profundas
- 3 Funciones de Activación
- 4 Aproximación de Funciones
- 5 Salida Multiclase

Redes con dos capas

A Arquitectura dos capas

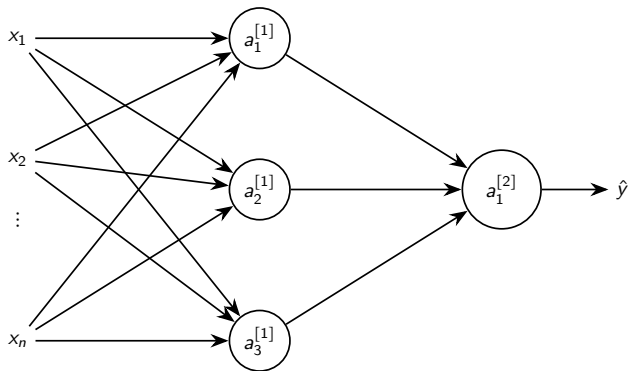
- ← Entrada
- ← Una capa oculta



Redes con dos capas

$$z^{(i)[l]} = W^{[l]T} x^{(i)} + b^{[l]}$$

$$a^{(i)[l]} = \sigma(z^{(i)[l]})$$

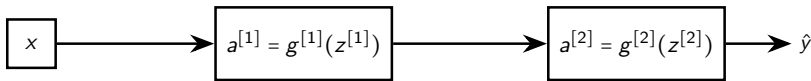


Capa de
Entrada
(Input)
 $x \in \mathbb{R}^n$

Capa Oculta

Capa de Salida

Forward Propagation



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

$$z^{[2]} = W^{[2]}a^{[1]} + b^{[2]}$$

Backpropagation

- El gradiente con respecto a $w^{[2]}, b^{[2]}$ es idéntico al caso anterior excepto que ahora $a^{[1]}$ juega el papel de x .
- Las siguientes diapositivas explican de forma esquemática el álgebra detrás del cálculo del gradiente. La forma eficiente de calcularlo es lo que se conoce como **backpropagation**.
- Este algoritmo se generaliza de forma natural al caso de redes profundas (i.e., redes con muchas capas ocultas).

Backpropagation

$$\begin{aligned}
 & \mathcal{L}(y_i, \hat{y}_i (w^{[1]}, b^{[1]}, w^{[2]}, b^{[2]})) \\
 &= y_i \ln(a^{[2]} (w^{[2]} a^{[1]} (w^{[1]} x + b^{[1]}) + b^{[2]})) \\
 &+ (1 - y_i) \ln(1 - a^{[2]} (z^{[2]}))
 \end{aligned}$$

$$\begin{aligned}
 &= y_i \ln(a^{[2]} (w^{[2]} a^{[1]} (z^{[1]}) + b^{[2]})) \\
 &+ (1 - y_i) \ln(1 - a^{[2]} (z^{[2]}))
 \end{aligned}$$

$$\frac{\partial \mathcal{L}}{\partial w_j^{[2]}} = (a^{[2]} (z^{[2]}) - y) a_j^{[1]}$$

$$\frac{\partial \mathcal{L}}{\partial b^{[2]}} = a^{[2]} (z^{[2]}) - y = \frac{\partial \mathcal{L}}{\partial a^{[2]}} \cdot \frac{\partial a^{[2]}}{\partial z^{[2]}} = d z^{[2]}$$

$$\frac{\partial z^{[2]}}{\partial w^{[1]}} = \underbrace{\frac{\partial z^{[2]}}{\partial a^{[1]}} \cdot \frac{\partial a^{[1]}}{\partial z^{[1]}}}_{d z^{[2]}} \cdot \underbrace{\frac{\partial z^{[1]}}{\partial w^{[1]}}}_{w^{[1]}} \cdot \underbrace{\frac{\partial a^{[1]}}{\partial z^{[1]}}}_{a^{[1]}} \cdot \underbrace{\frac{\partial z^{[1]}}{\partial w^{[1]}}}_{x^T}$$

$$\frac{\partial z^{[2]}}{\partial w^{[1]}} = d z^{[2]}$$

$$z_i^{[2]} = w_i^{[1]T} a^{[1]} + b_i^{[2]}$$

Multiplicar matriz por vector columna:

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} w_{11}x_1 + w_{12}x_2 \\ w_{21}x_1 + w_{22}x_2 \end{bmatrix}$$

$$z = w x = x_1 \begin{bmatrix} w_{11} \\ w_{21} \end{bmatrix} + x_2 \begin{bmatrix} w_{12} \\ w_{22} \end{bmatrix}$$

$$w_1 = \begin{bmatrix} w_{11} \\ w_{21} \end{bmatrix} \Rightarrow z_i = w_i^T x$$

$$w_2 = \begin{bmatrix} w_{12} \\ w_{22} \end{bmatrix}$$

Contenido

- 1 Redes con dos capas
- 2 Redes Profundas**
- 3 Funciones de Activación
- 4 Aproximación de Funciones
- 5 Salida Multiclase

Propagación Forward

- Caso general:

Propagación Forward

$$z^{[l]} = w^{[l]}x + b^{[l]}$$

$$a^{[l]} = g^{[l]}(z^{[l]})$$

$$z^{[l+1]} = w^{[l+1]}a^{[l]} + b^{[l+1]}$$

Ejemplo con Activación Sigmoid

Capa L

$$dz^{[L]} = a^{[L]} - y$$

$$dW^{[L]} = dz^{[L]} a^{[L-1]T}$$

$$db^{[L]} = dz^{[L]}$$

Ejemplo con Activación Sigmoid

Capa $l < L$

$$dz^{[l]} = W^{[l+1]T} dz^{[l+1]} * g^{[l]'}(z^{[l]})$$

$$dW^{[l]} = dz^{[l]} a^{[l]T}$$

$$db^{[l]} = dz^{[l]}$$

Contenido

- 1 Redes con dos capas
- 2 Redes Profundas
- 3 Funciones de Activación**
- 4 Aproximación de Funciones
- 5 Salida Multiclase

Funciones de Activación

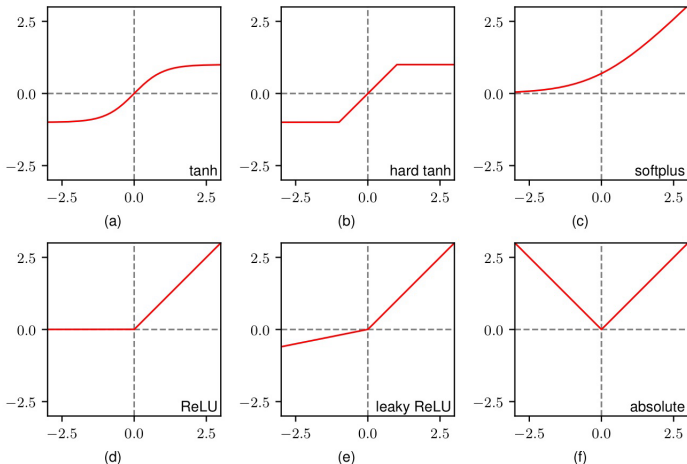
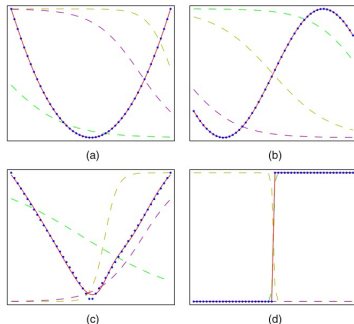


Figure 6.12 A variety of nonlinear activation functions.

Contenido

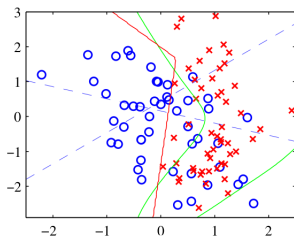
- 1 Redes con dos capas
- 2 Redes Profundas
- 3 Funciones de Activación
- 4 Aproximación de Funciones**
- 5 Salida Multiclase

Aproximación de Funciones



- Capacidad de aproximación de una red (e.g., cuadrática, seno, valor absoluto y Heavside). Los datos son los 50 puntos azules. Se entrena una red con dos capas, tres neuronas, función de activación tanh, salidas lineales. Las salidas de las tres neuronas ocultas se muestran con líneas punteadas.

Ejemplo Clasificación



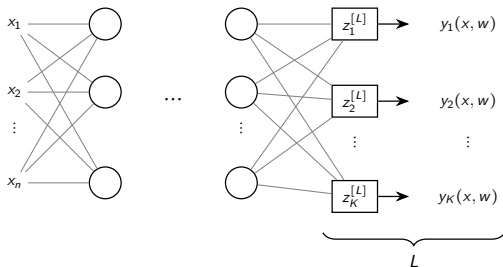
- Capacidad de aproximación de una red: las líneas punteadas son las salidas de cada una de las dos neuronas (hipersuperficies). Funciones de activación tanh y salida logística sigmoid.
- La línea verde es el clasificador Bayesiano. La roja el clasificador de la red.

Contenido

- 1 Redes con dos capas
- 2 Redes Profundas
- 3 Funciones de Activación
- 4 Aproximación de Funciones
- 5 Salida Multiclase**

(Softmax)

$$y_k(x, w) = p(y_k = 1 | x) = \frac{e^{z_k^{[L]}}}{\sum_{j=1}^K e^{z_j^{[L]}}}$$



- Obsérvese que la última capa es especial, las neuronas tienen funciones de activación distintas.