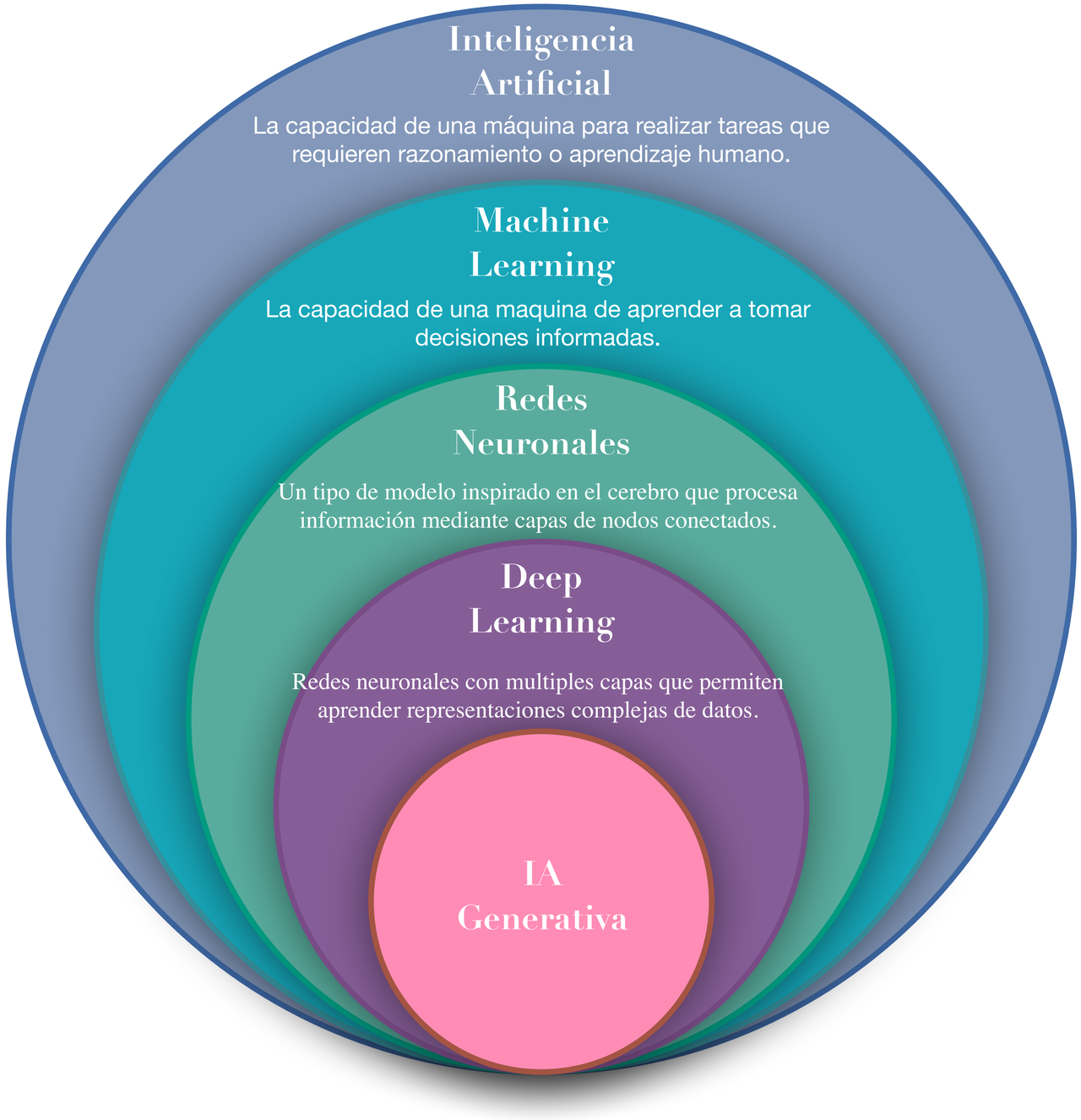
*FCEN, UBA*

# Introducción a las Redes Neuronales

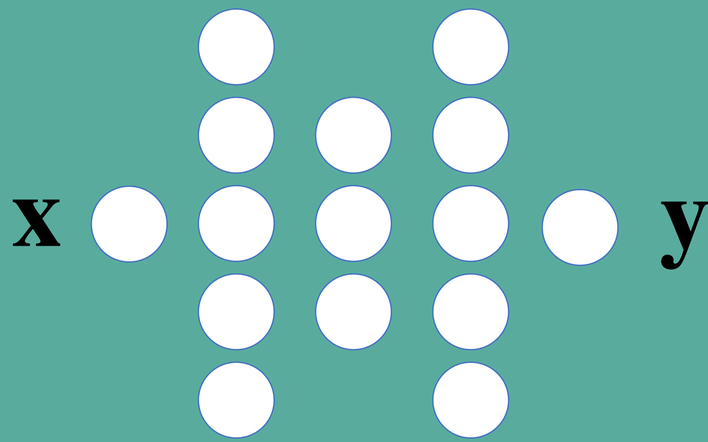
Cecilia Garraffo

# Outline

- Redes Neuronales Shallow
- Funciones de Activación
- Ajuste de Curvas con Redes Neuronales Shallow
- Funciones de Pérdida


Inteligencia Artificial
La capacidad de una máquina para realizar tareas que requieren razonamiento o aprendizaje humano.
Machine Learning
La capacidad de una maquina de aprender a tomar decisiones informadas.
Redes Neuronales
Un tipo de modelo inspirado en el cerebro que procesa información mediante capas de nodos conectados.
Deep Learning
Redes neuronales con multiples capas que permiten aprender representaciones complejas de datos.
IA Generativa

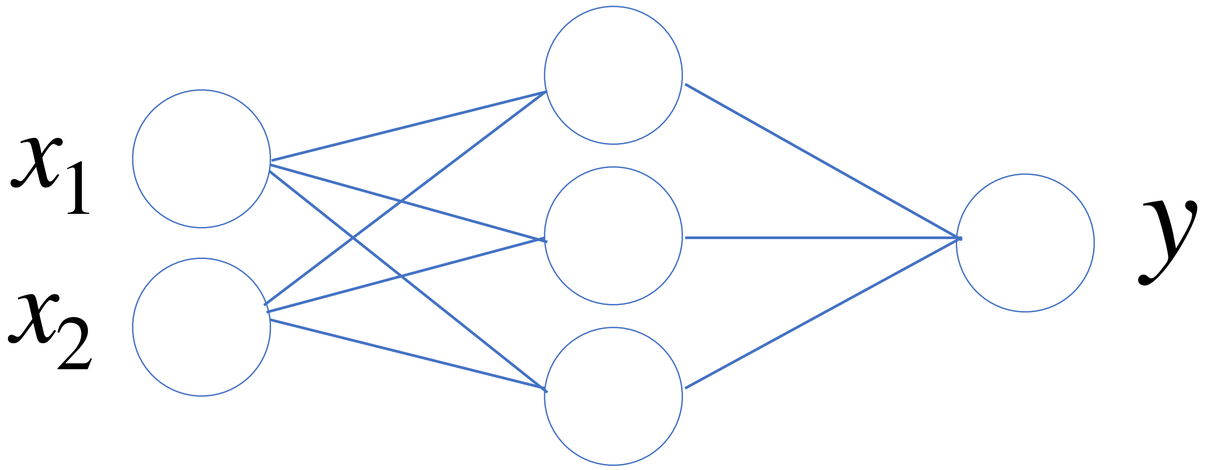
Redes
Neuronales




$$\mathbf{y} = \mathbf{F}[\mathbf{x}, \theta]$$

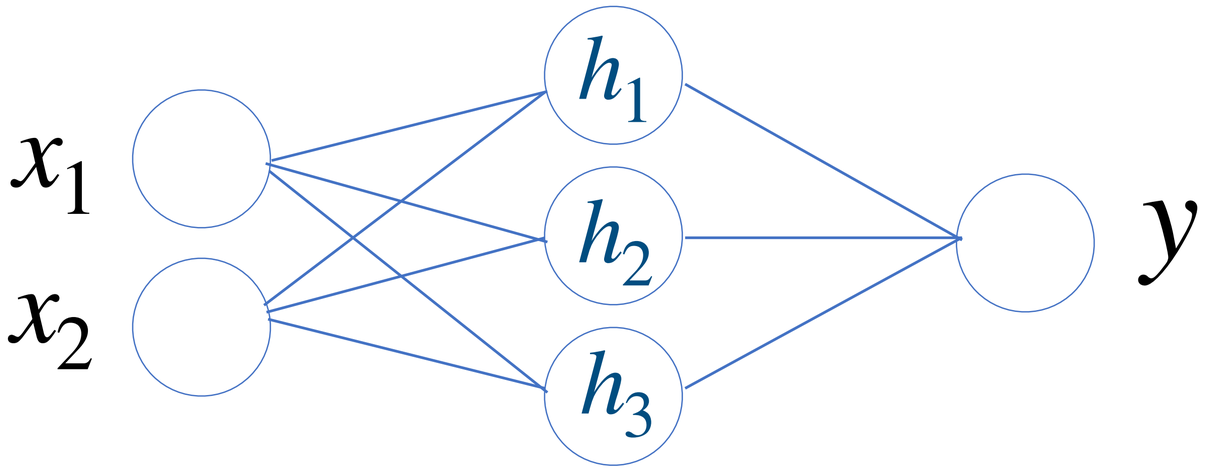
# Redes Neuronales: Shallow

$$\mathbf{y} = \mathbf{F}[\mathbf{x}, \theta]$$

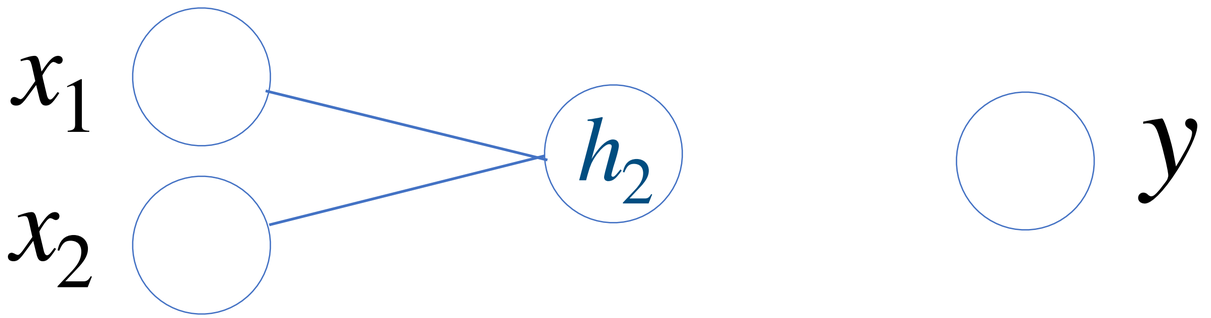


# Redes Neuronales: Shallow

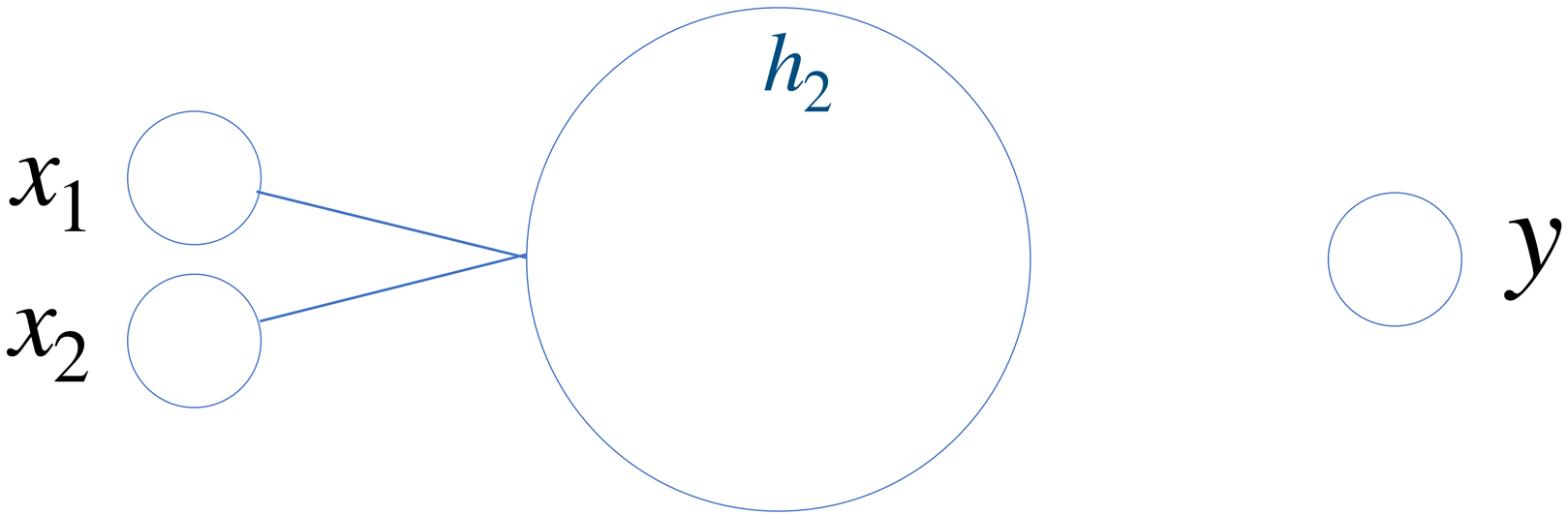
$$\mathbf{y} = \mathbf{F}[\mathbf{x}, \theta]$$



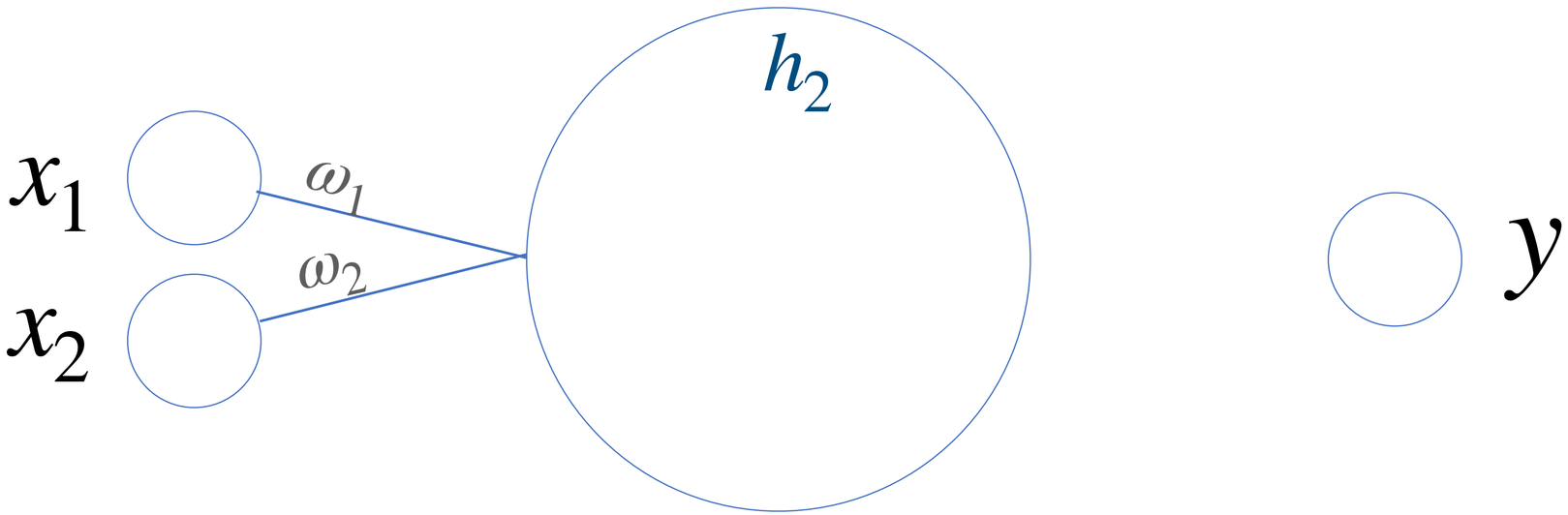
# Redes Neuronales: Shallow



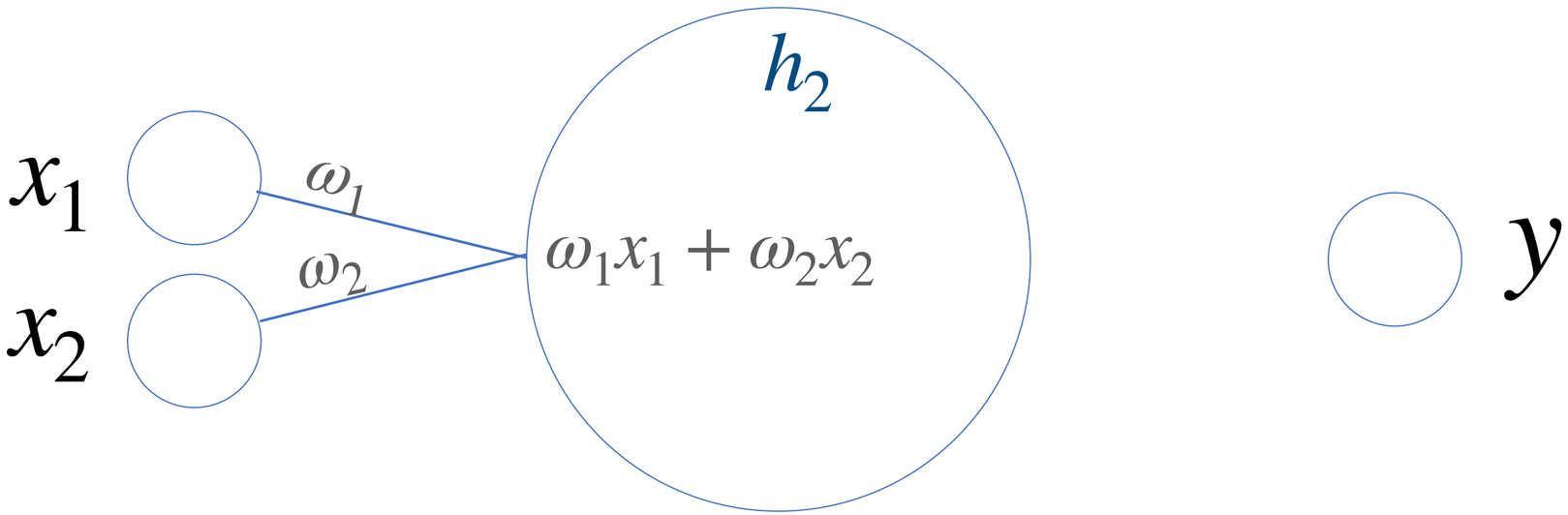
# Redes Neuronales: Shallow



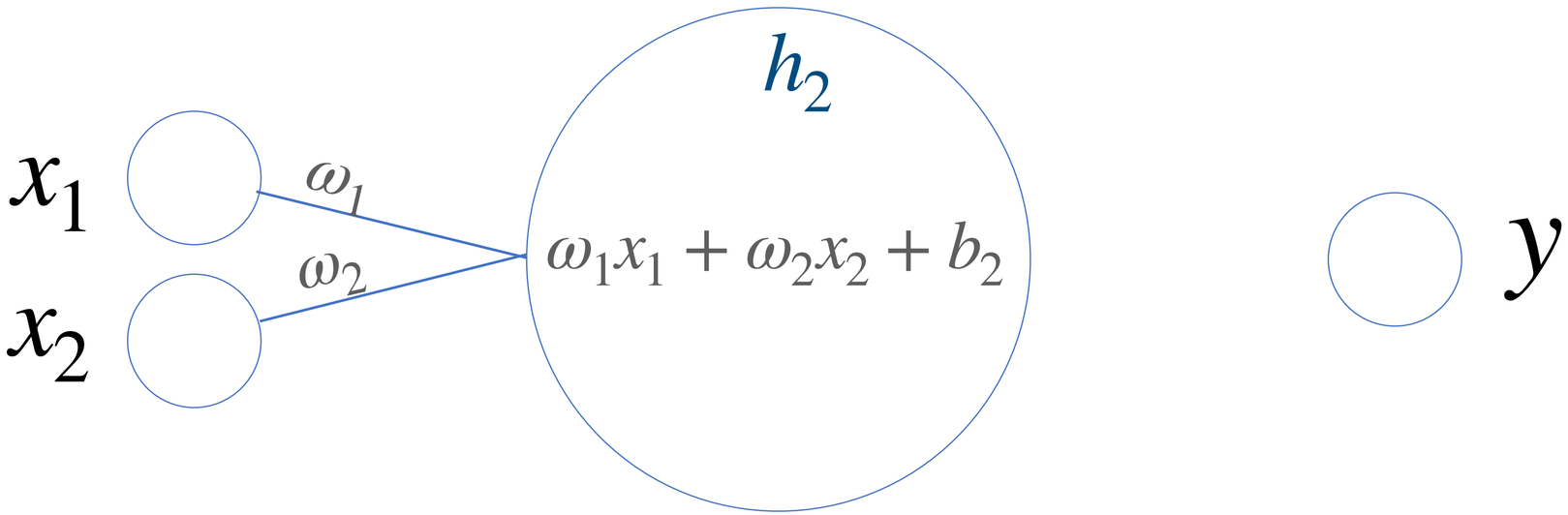
# Redes Neuronales: Shallow



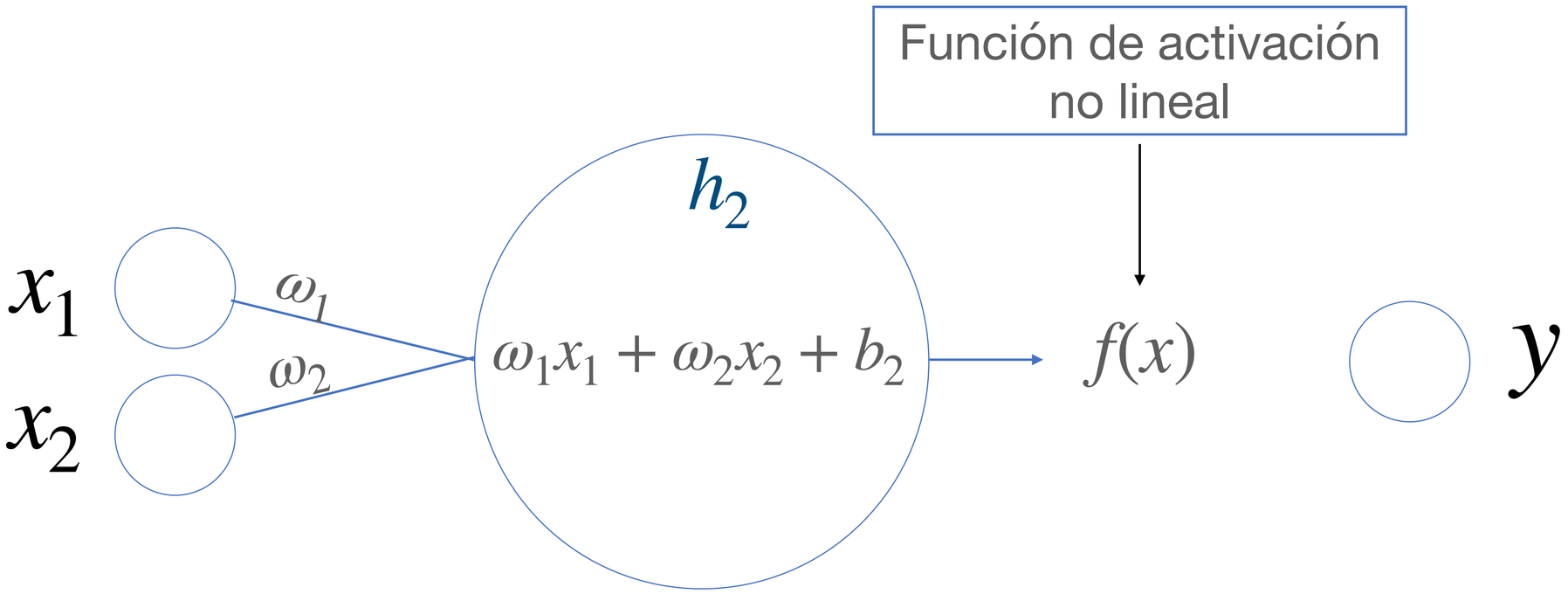
# Redes Neuronales: Shallow



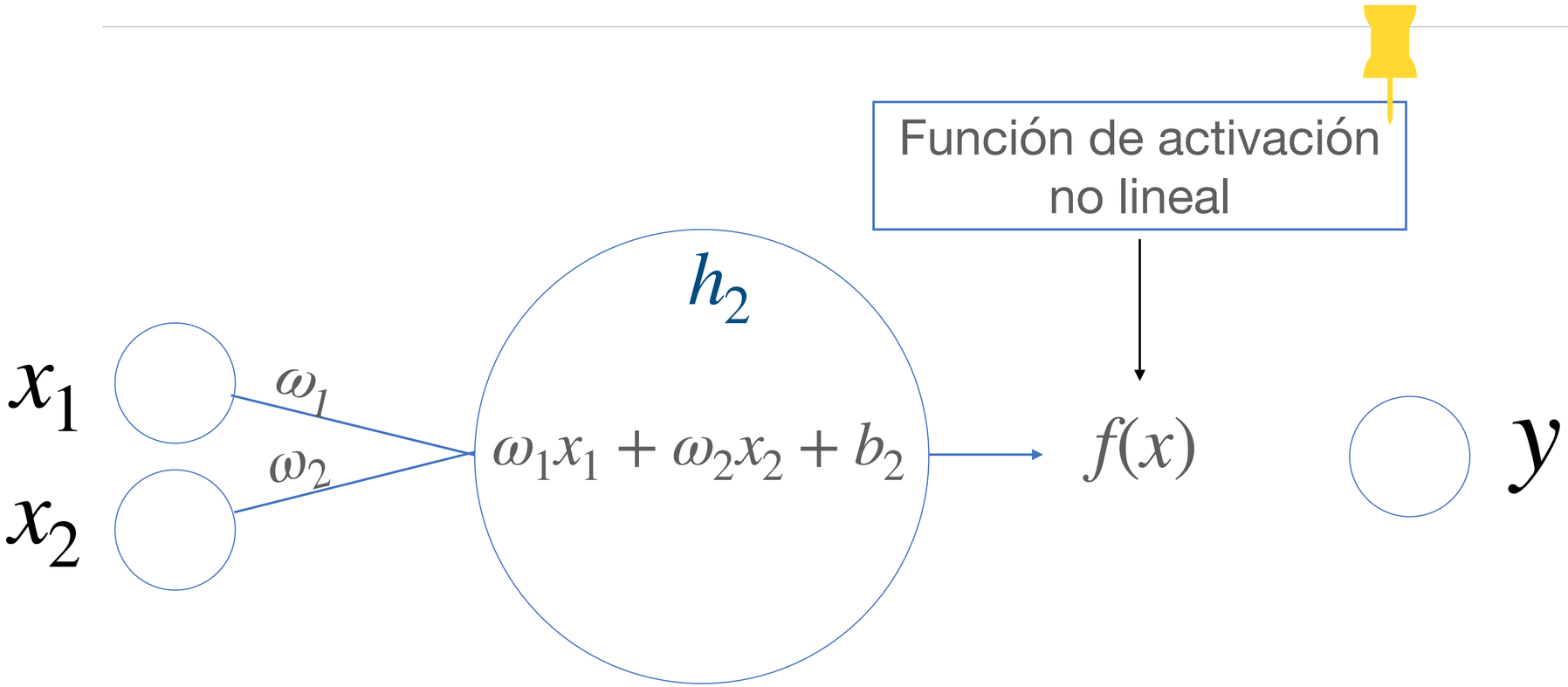
# Redes Neuronales: Shallow



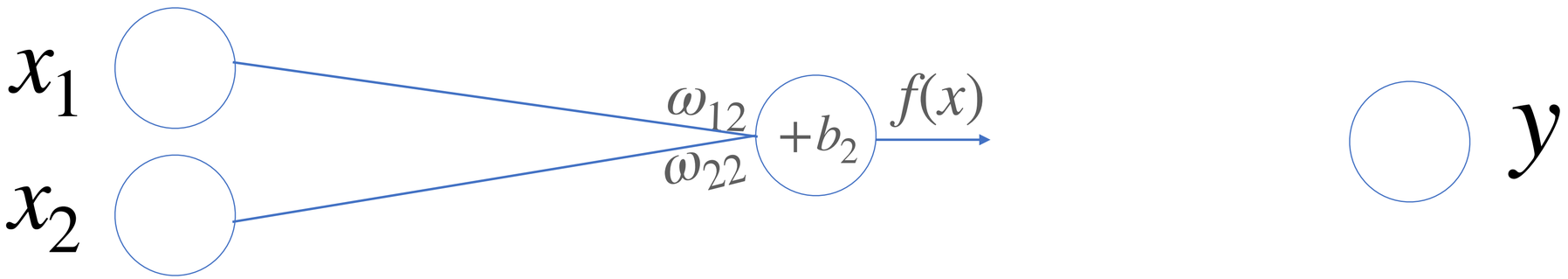
# Redes Neuronales: Shallow



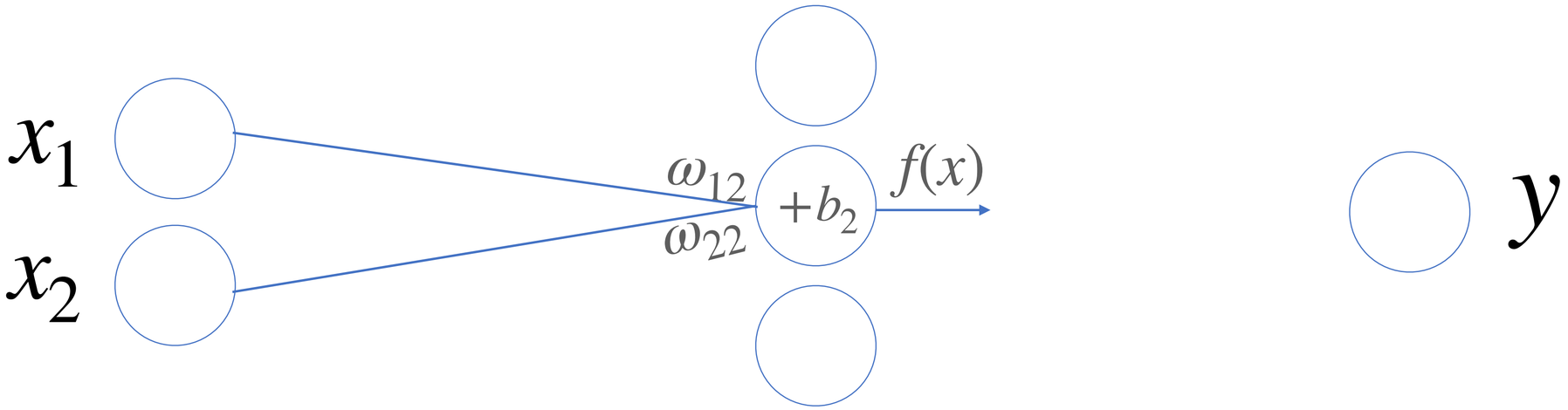
# Redes Neuronales: Shallow



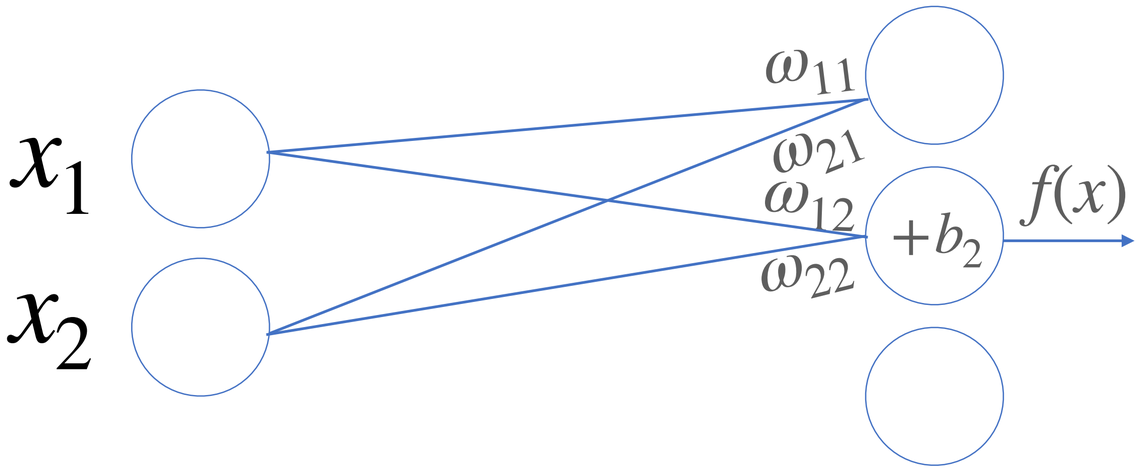
# Redes Neuronales: Shallow



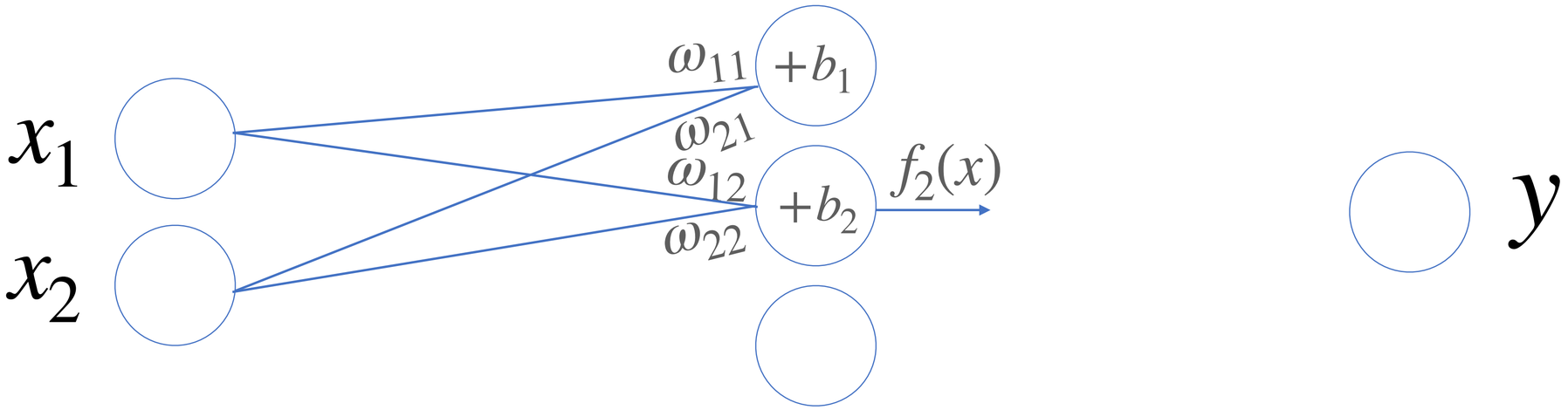
# Redes Neuronales: Shallow



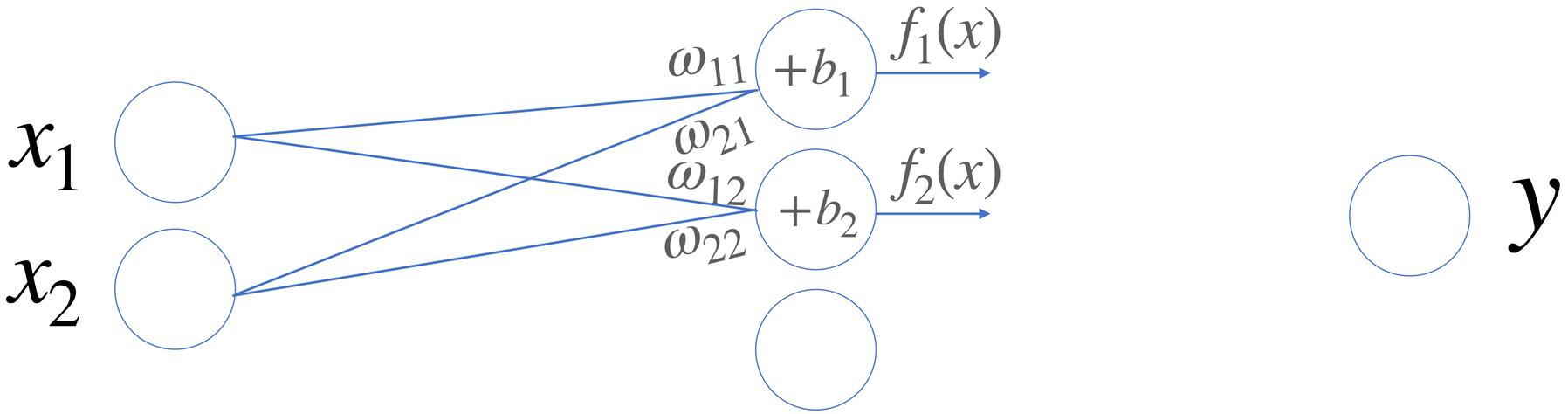
# Redes Neuronales: Shallow



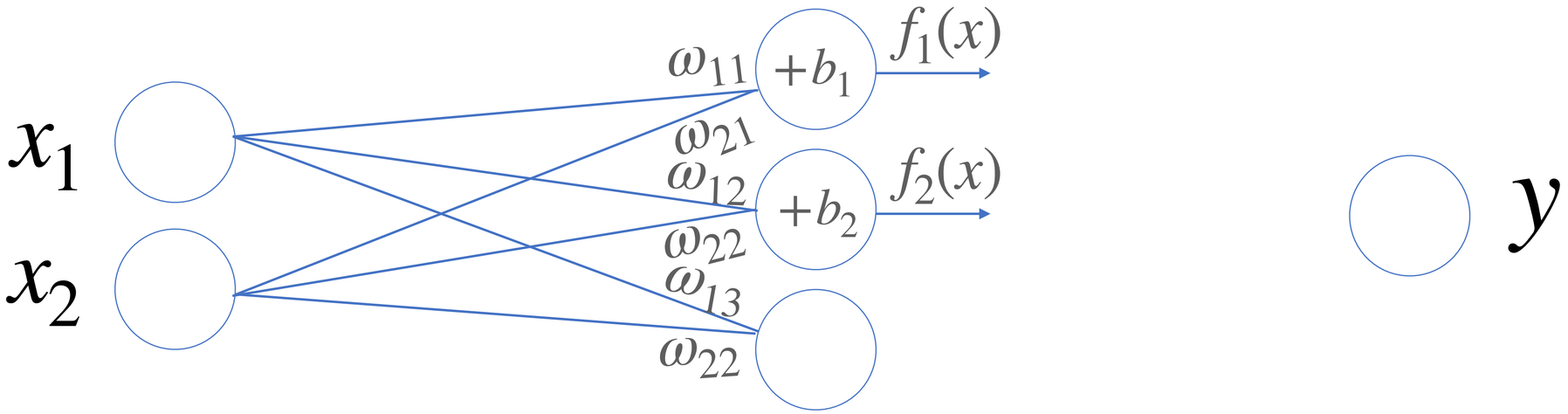
# Redes Neuronales: Shallow



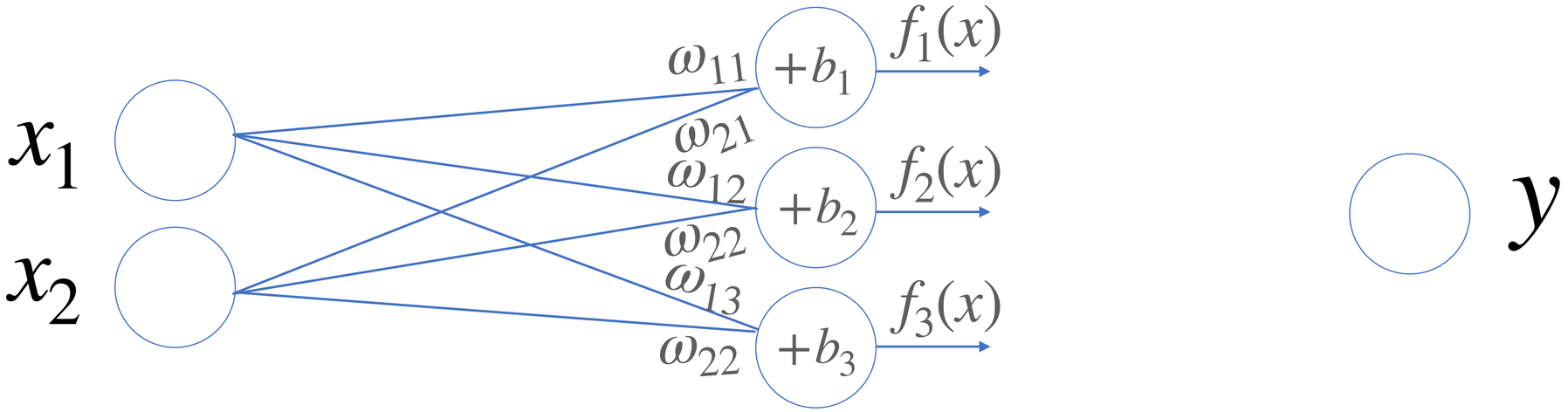
# Redes Neuronales: Shallow



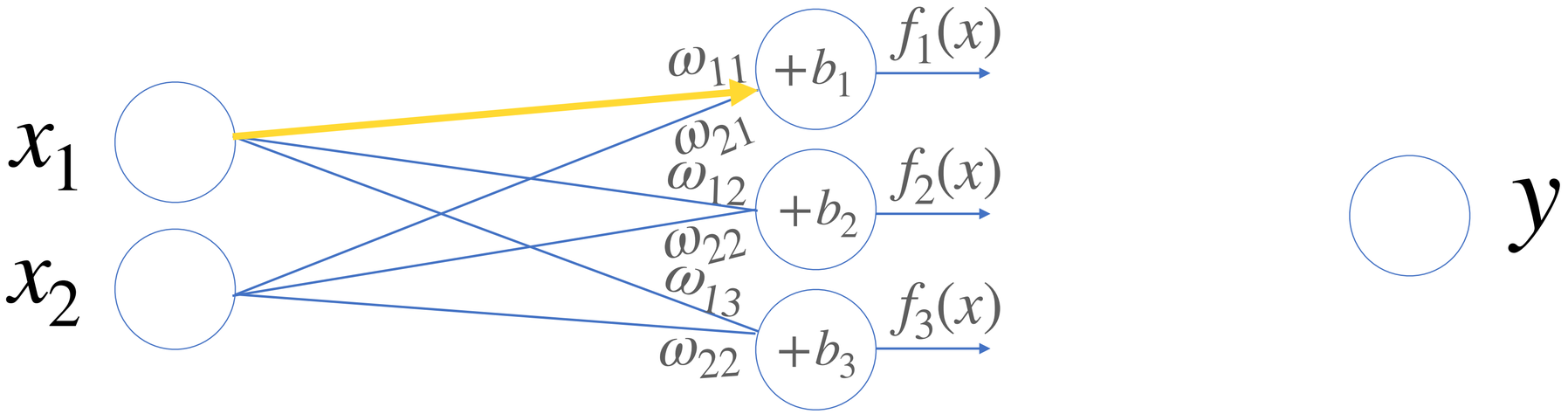
# Redes Neuronales: Shallow



# Redes Neuronales: Shallow

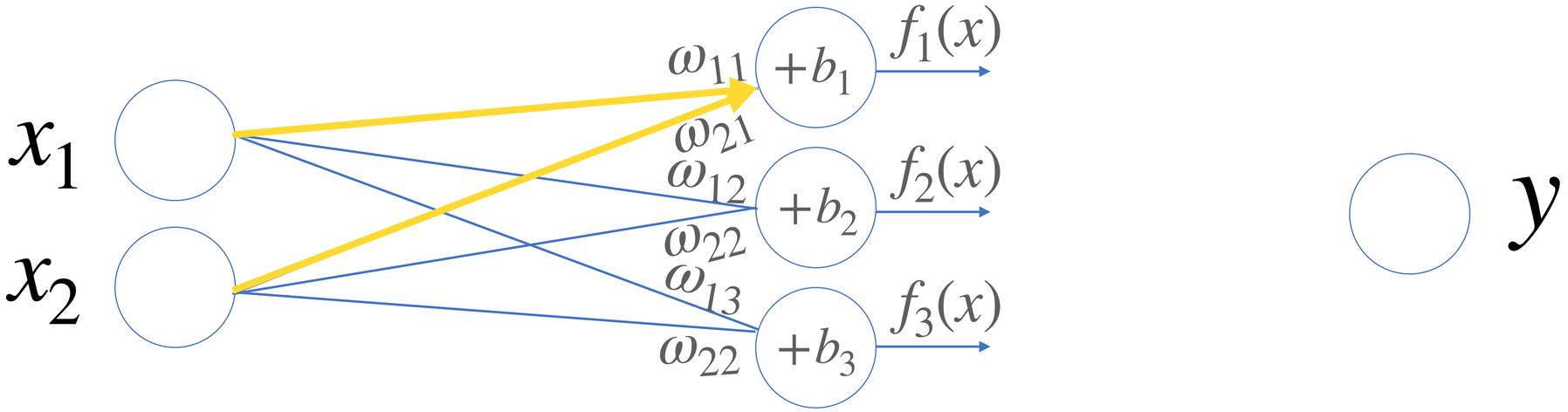


# Redes Neuronales: Shallow



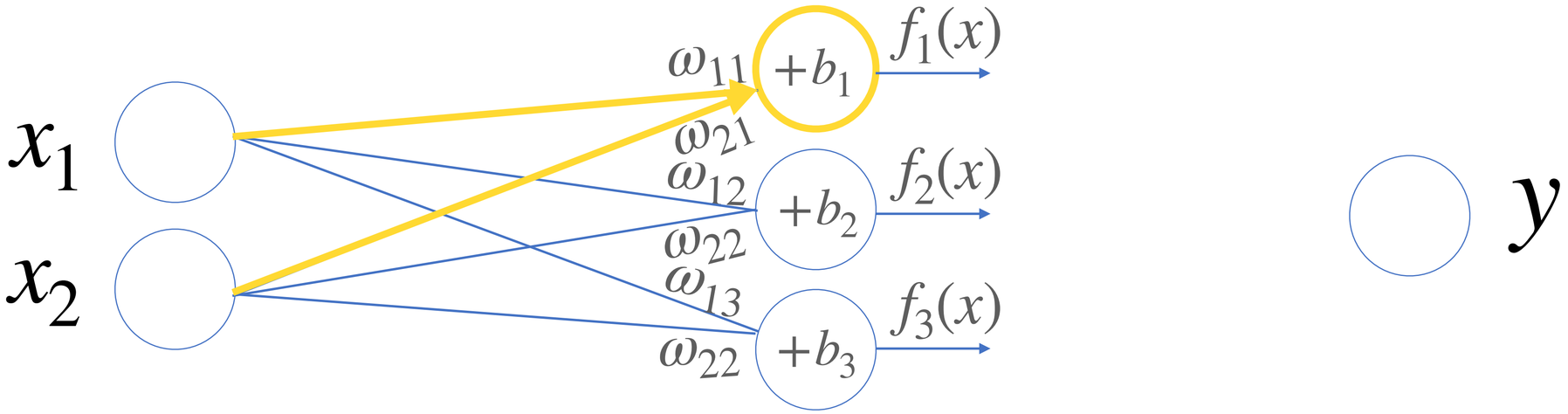

$x_1\omega_{11}$

# Redes Neuronales: Shallow



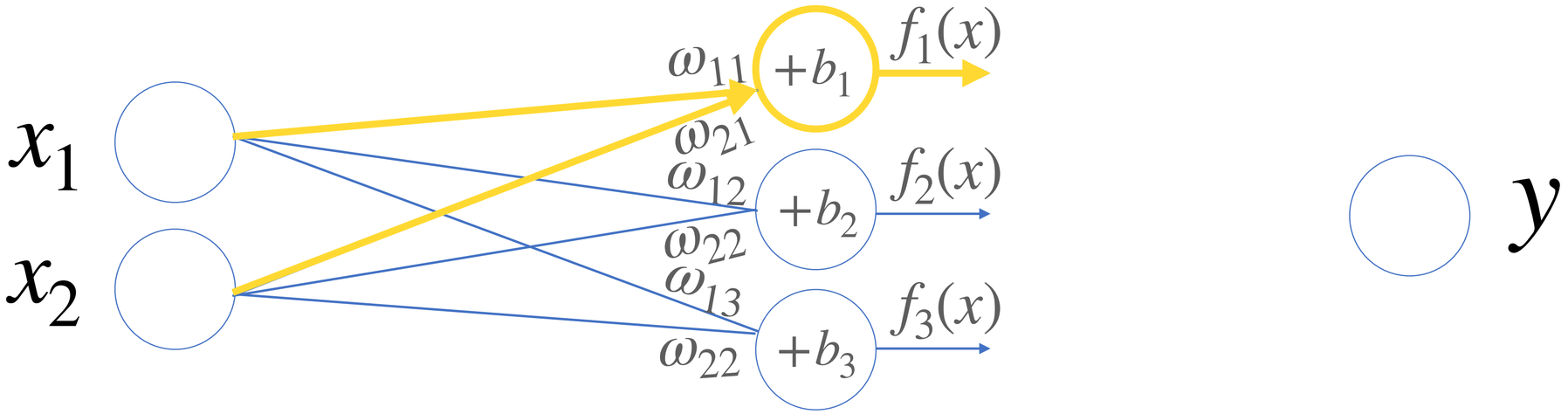

$$x_1\omega_{11} + x_2\omega_{21}$$

# Redes Neuronales: Shallow



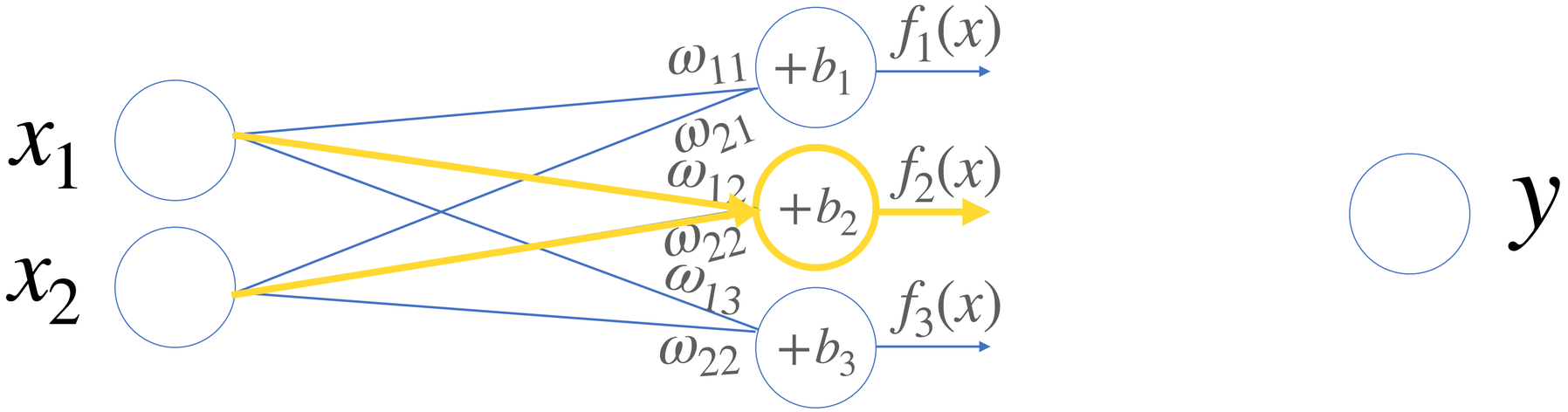

$$x_1\omega_{11} + x_2\omega_{21} + b_1$$

# Redes Neuronales: Shallow



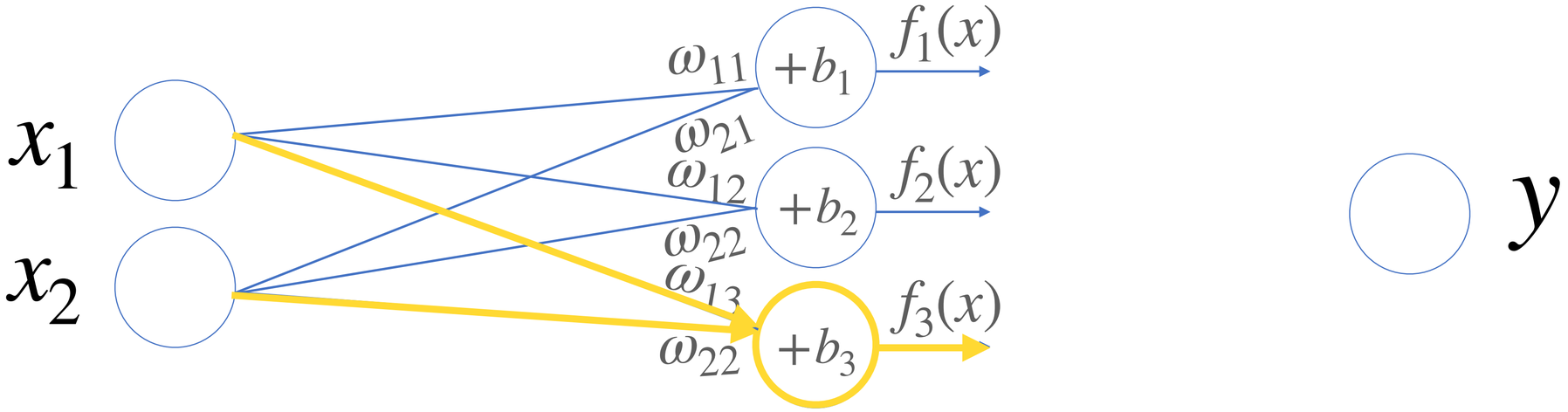

$$f_1(x_1\omega_{11} + x_2\omega_{21} + b_1)$$

# Redes Neuronales: Shallow



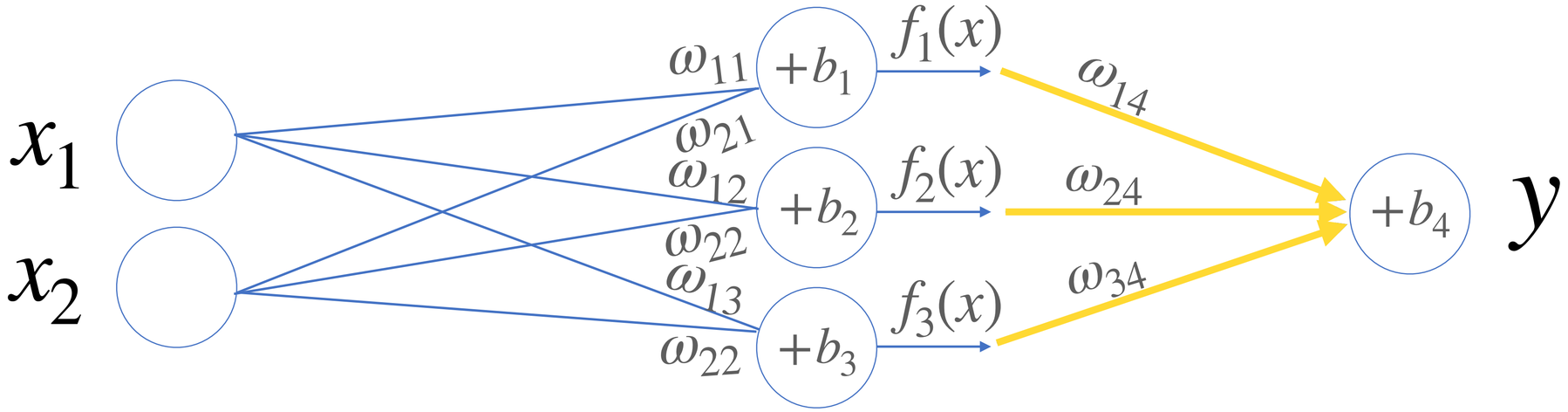

$f_1(x_1\omega_{11} + x_2\omega_{21} + b_1)$ $f_2(x_1\omega_{12} + x_2\omega_{22} + b_2)$

# Redes Neuronales: Shallow



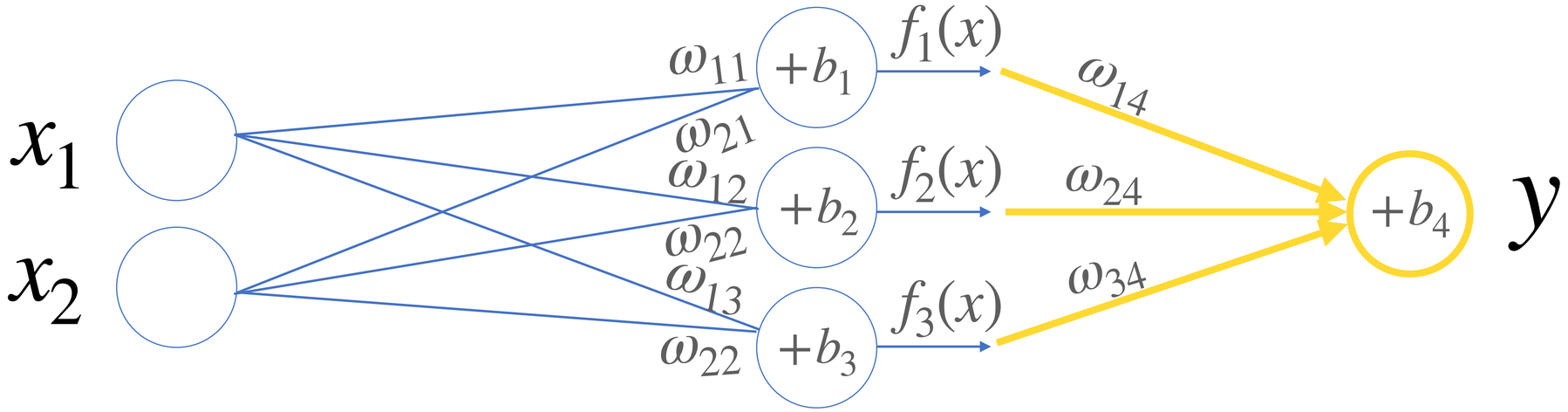

$f_1(x_1\omega_{11} + x_2\omega_{21} + b_1)$ $f_2(x_1\omega_{12} + x_2\omega_{22} + b_2)$ $f_3(x_1\omega_{13} + x_2\omega_{23} + b_3)$

# Redes Neuronales: Shallow



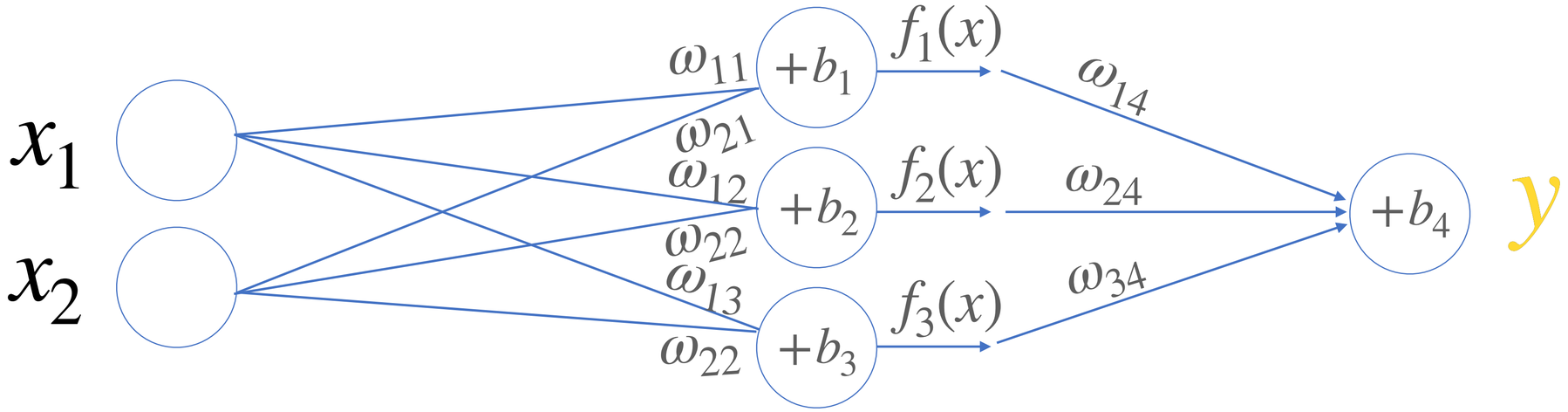

$$\omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3)$$

# Redes Neuronales: Shallow




$$\omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

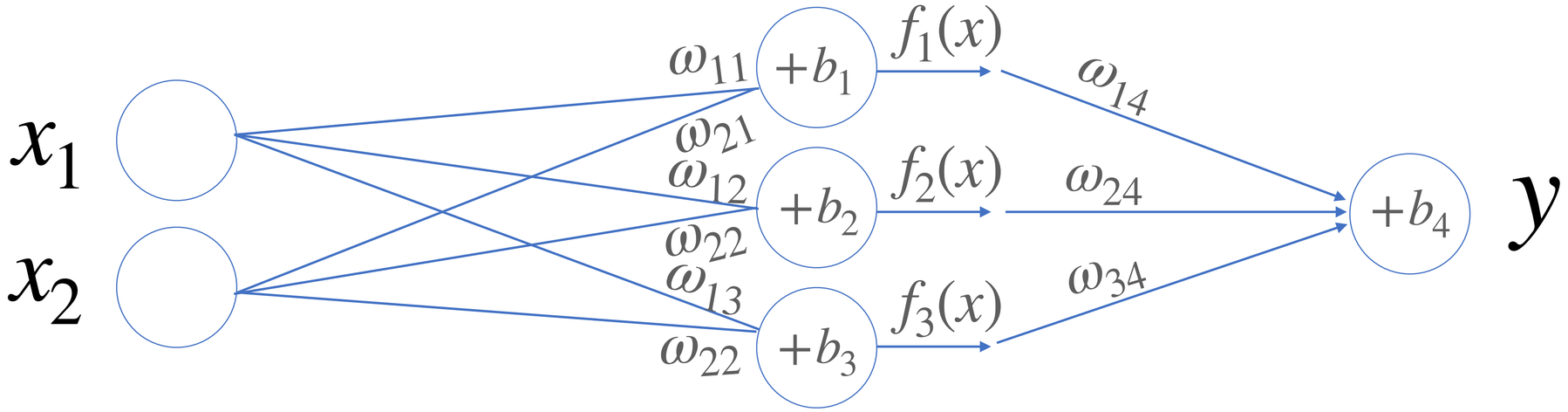
# Redes Neuronales: Shallow




$$y = \omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

# Redes Neuronales: Quiz!

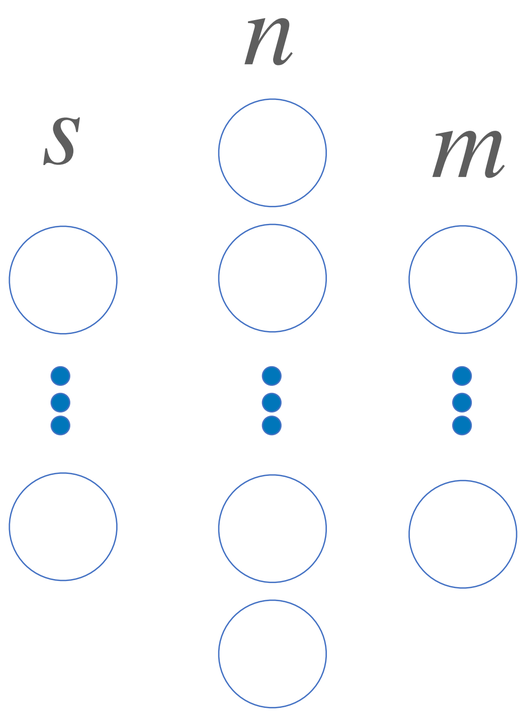
Cuantos parámetros tiene esta red neuronal?




$$y = \omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

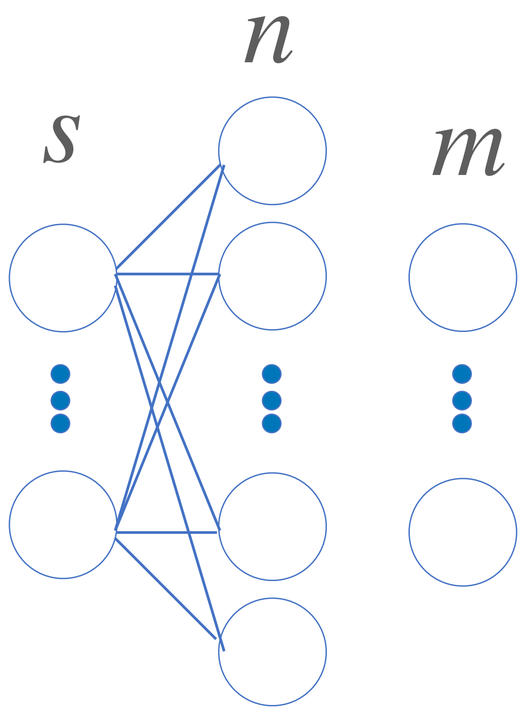
# Redes Neuronales: Quiz!

Cuantos parámetros tiene  una red neuronal de shallow con s inputs, m outputs y n neuronas?



# Redes Neuronales: Quiz!

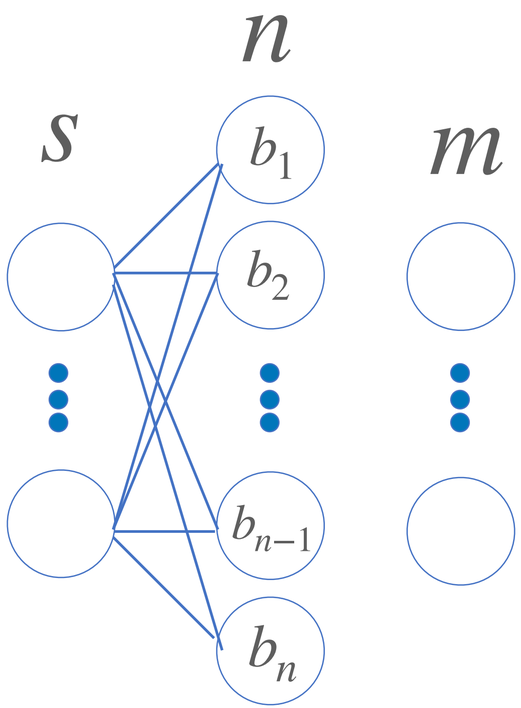
Cuantos parámetros tiene una red neuronal de shallow con s inputs, m outputs y n neuronas?




$s \times n$

# Redes Neuronales: Quiz!

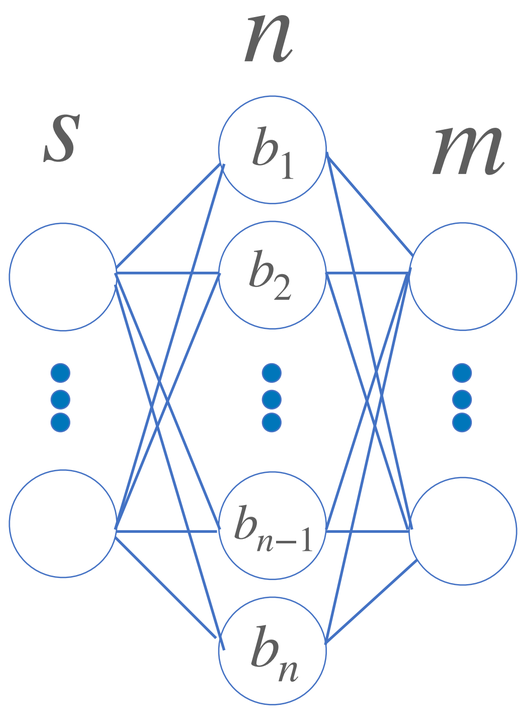
Cuantos parámetros tiene una red neuronal de shallow con s inputs, m outputs y n neuronas?




$$s \times n + n$$

# Redes Neuronales: Quiz!

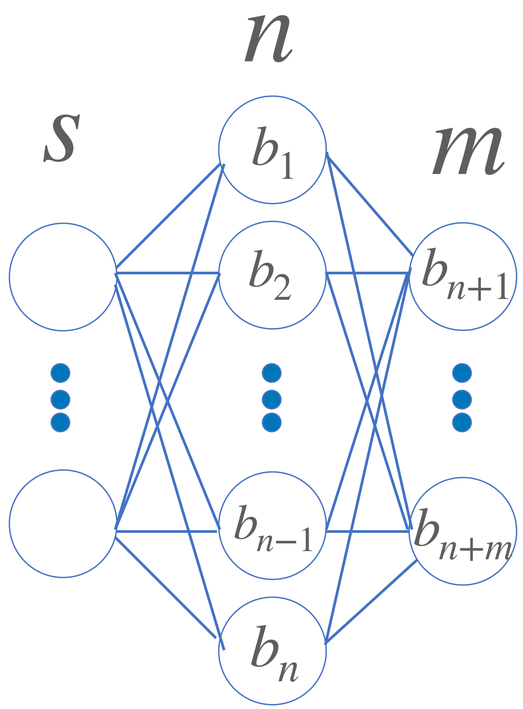
Cuantos parámetros tiene una red neuronal de shallow con s inputs, m outputs y n neuronas?




$$s \times n + n + n \times m$$

# Redes Neuronales: Quiz!

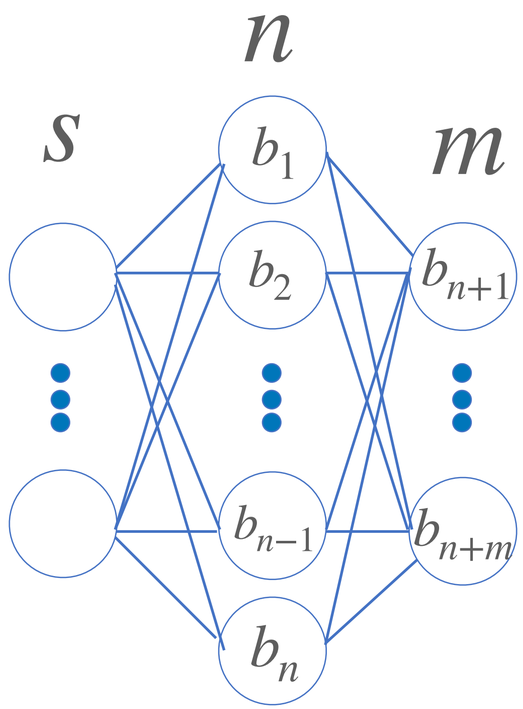
Cuantos parámetros tiene una red neuronal de shallow con s inputs, m outputs y n neuronas?




$$s \times n + n + n \times m + m$$

# Redes Neuronales: Quiz!

Cuantos parámetros tiene una red neuronal de shallow con s inputs, m outputs y n neuronas?

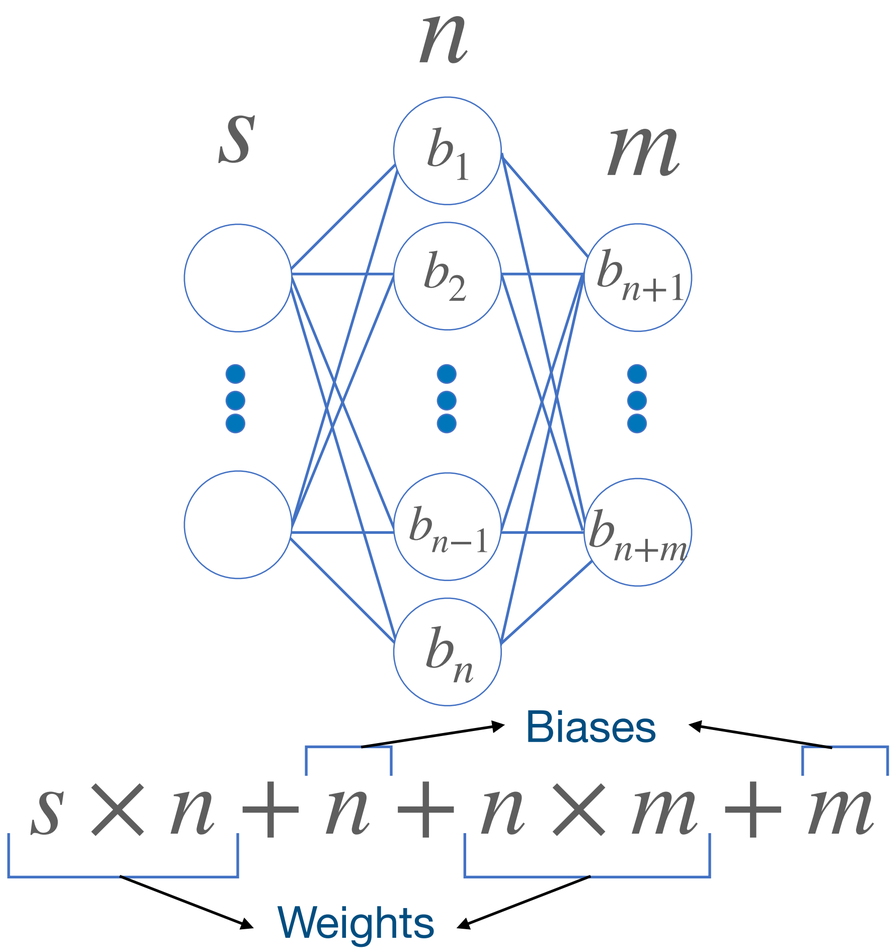



$$s \times n + n + n \times m + m$$

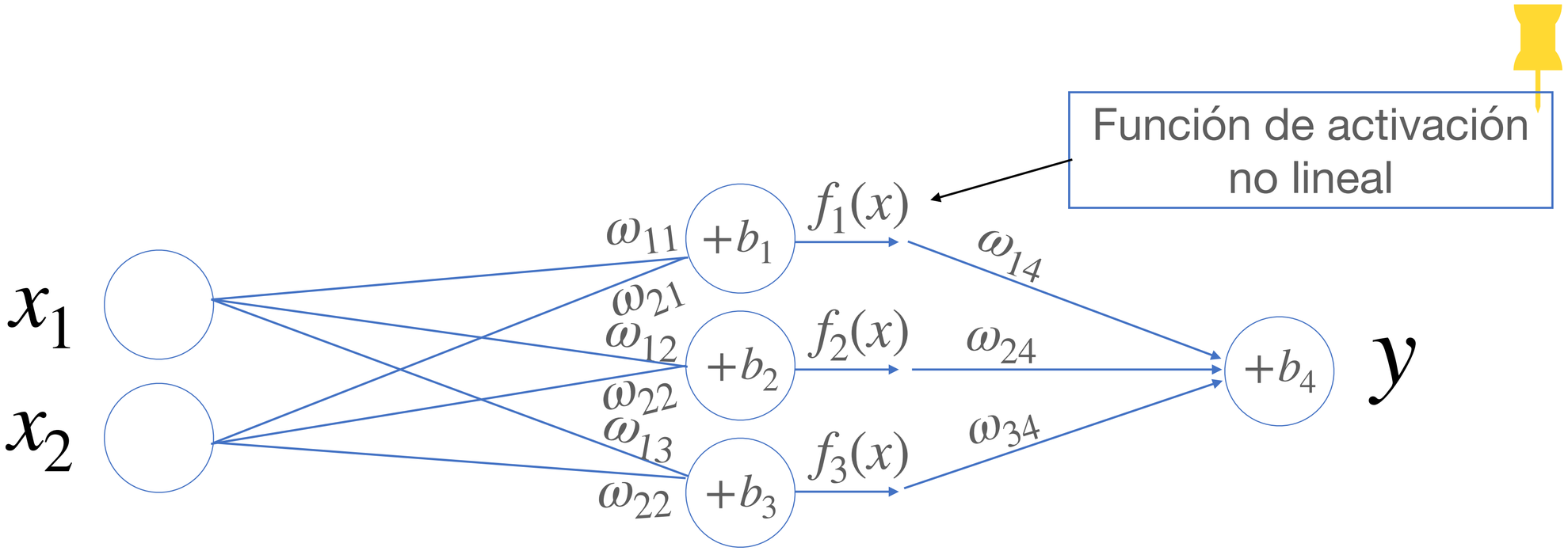
Weights

# Redes Neuronales: Quiz!

Cuantos parámetros tiene una red neuronal de shallow con s inputs, m outputs y n neuronas?



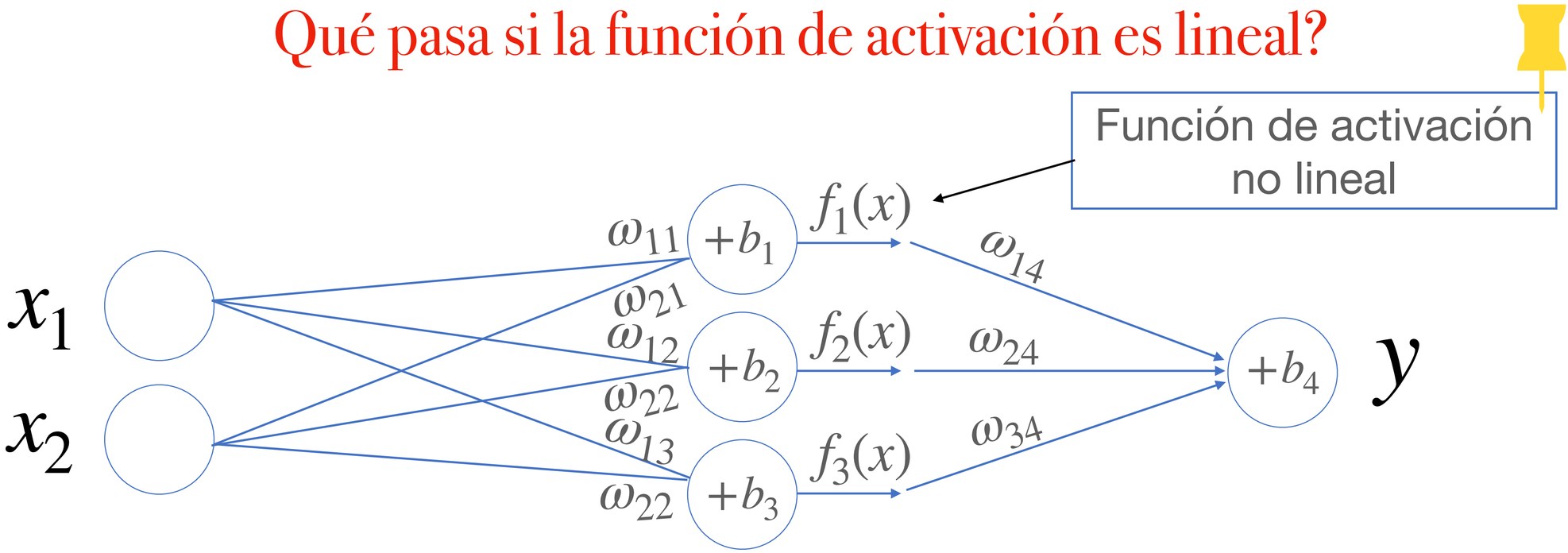
# Redes Neuronales: Shallow




$$y = \omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

# Redes Neuronales: Shallow

Qué pasa si la función de activación es lineal?




$$y = \omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

# Redes Neuronales: Shallow

$$f_1(x) = f_2(x) = f_3(x) = ax + b$$

# Redes Neuronales: Shallow

$$f_1(x) = f_2(x) = f_3(x) = ax + b$$

$$y = \omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

# Redes Neuronales: Shallow

$$f_1(x) = f_2(x) = f_3(x) = ax + b$$

$$y = \omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

$$y = \omega_{14} a\,(x_1\omega_{11} + x_2\omega_{21} + b_1) + b$$

# Redes Neuronales: Shallow

$$f_1(x) = f_2(x) = f_3(x) = ax + b$$

$$y = \omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

$$y = \omega_{14} a\,(x_1\omega_{11} + x_2\omega_{21} + b_1) + b$$
$$+ \omega_{24}\, a\,(x_1\omega_{12} + x_2\omega_{22} + b_2) + b$$

# Redes Neuronales: Shallow

$$f_1(x) = f_2(x) = f_3(x) = ax + b$$

$$y = \omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

$$y = \omega_{14} a\,(x_1\omega_{11} + x_2\omega_{21} + b_1) + b$$
$$+ \omega_{24}\, a\,(x_1\omega_{12} + x_2\omega_{22} + b_2) + b$$
$$+ \omega_{34} a\,(x_1\omega_{13} + x_2\omega_{23} + b_3) + b$$

# Redes Neuronales: Shallow

$$f_1(x) = f_2(x) = f_3(x) = ax + b$$

$$y = \omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

$$\begin{aligned} y = {} & \omega_{14}\, a\,(x_1\omega_{11} + x_2\omega_{21} + b_1) + b \\ & + \omega_{24}\, a\,(x_1\omega_{12} + x_2\omega_{22} + b_2) + b \\ & + \omega_{34} a\,(x_1\omega_{13} + x_2\omega_{23} + b_3) + b + b_4 \end{aligned}$$

# Redes Neuronales: Shallow

$$f_1(x) = f_2(x) = f_3(x) = ax + b$$

$$y = \omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

$$y = \omega_{14} a\,(x_1\omega_{11} + x_2\omega_{21} + b_1) + b$$
$$+ \omega_{24}\, a\,(x_1\omega_{12} + x_2\omega_{22} + b_2) + b$$
$$+ \omega_{34} a\,(x_1\omega_{13} + x_2\omega_{23} + b_3) + b + b_4$$

$$y = a\,x_1\omega_{11}\omega_{14} + a\,x_2\omega_{21}\omega_{14} + a\,b_1\omega_{14}$$

# Redes Neuronales: Shallow

$$f_1(x) = f_2(x) = f_3(x) = ax + b$$

$$y = \omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

$$y = \omega_{14} a\,(x_1\omega_{11} + x_2\omega_{21} + b_1) + b$$
$$+ \omega_{24}\, a\,(x_1\omega_{12} + x_2\omega_{22} + b_2) + b$$
$$+ \omega_{34} a\,(x_1\omega_{13} + x_2\omega_{23} + b_3) + b + b_4$$

$$y = a\,x_1\omega_{11}\omega_{14} + a\,x_2\omega_{21}\omega_{14} + a\,b_1\omega_{14}$$
$$+ a\,x_1\omega_{12}\omega_{24} + a\,x_2\omega_{22}\omega_{24} + a\,b_2\omega_{24}$$

# Redes Neuronales: Shallow

$$f_1(x) = f_2(x) = f_3(x) = ax + b$$

$$y = \omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

$$y = \omega_{14} a\,(x_1\omega_{11} + x_2\omega_{21} + b_1) + b$$
$$+ \omega_{24}\, a\,(x_1\omega_{12} + x_2\omega_{22} + b_2) + b$$
$$+ \omega_{34} a\,(x_1\omega_{13} + x_2\omega_{23} + b_3) + b + b_4$$

$$y = a\,x_1\omega_{11}\omega_{14} + a\,x_2\omega_{21}\omega_{14} + a\,b_1\omega_{14}$$
$$+ a\,x_1\omega_{12}\omega_{24} + a\,x_2\omega_{22}\omega_{24} + a\,b_2\omega_{24}$$
$$a\,x_1\omega_{13}\omega_{34} + a\,x_2\omega_{23}\omega_{34} + a\,b_3\omega_{34} + 3b + b_4$$

# Redes Neuronales: Shallow

$$f_1(x) = f_2(x) = f_3(x) = ax + b$$

$$y = \omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

$$y = \omega_{14} a\,(x_1\omega_{11} + x_2\omega_{21} + b_1) + b$$
$$+ \omega_{24}\, a\,(x_1\omega_{12} + x_2\omega_{22} + b_2) + b$$
$$+ \omega_{34} a\,(x_1\omega_{13} + x_2\omega_{23} + b_3) + b + b_4$$

$$y = a\,x_1\omega_{11}\omega_{14} + a\,x_2\omega_{21}\omega_{14} + a\,b_1\omega_{14}$$
$$+ a\,x_1\omega_{12}\omega_{24} + a\,x_2\omega_{22}\omega_{24} + a\,b_2\omega_{24}$$
$$a\,x_1\omega_{13}\omega_{34} + a\,x_2\omega_{23}\omega_{34} + a\,b_3\omega_{34} + 3b + b_4$$

$$y = a\,(\omega_{11}\omega_{14} + \omega_{12}\omega_{24} + \omega_{13}\omega_{34})\,x_1 + a\,(\omega_{21}\omega_{14} + \omega_{22}\omega_{24} + \omega_{23}\omega_{34})\,x_2$$
$$+ a\,(b_1\omega_{14} + b_2\omega_{24} + b_3\omega_{34}) + 3b + b_4$$

# Redes Neuronales: Shallow

$$f_1(x) = f_2(x) = f_3(x) = ax + b$$

$$y = \omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

$$y = \omega_{14}\, a\,(x_1\omega_{11} + x_2\omega_{21} + b_1) + b$$
$$+ \omega_{24}\, a\,(x_1\omega_{12} + x_2\omega_{22} + b_2) + b$$
$$+ \omega_{34} a\,(x_1\omega_{13} + x_2\omega_{23} + b_3) + b + b_4$$

$$y = a\, x_1\omega_{11}\omega_{14} + a\, x_2\omega_{21}\omega_{14} + a\, b_1\omega_{14}$$
$$+ a\, x_1\omega_{12}\omega_{24} + a\, x_2\omega_{22}\omega_{24} + a\, b_2\omega_{24}$$
$$a\, x_1\omega_{13}\omega_{34} + a\, x_2\omega_{23}\omega_{34} + a\, b_3\omega_{34} + 3b + b_4$$

$$y = \underbrace{a\,(\omega_{11}\omega_{14} + \omega_{12}\omega_{24} + \omega_{13}\omega_{34})}_{c_1} x_1 + \underbrace{a\,(\omega_{21}\omega_{14} + \omega_{22}\omega_{24} + \omega_{23}\omega_{34})}_{c_2} x_2 + \underbrace{a\,(b_1\omega_{14} + b_2\omega_{24} + b_3\omega_{34}) + 3b + b_4}_{c_3}$$

# Redes Neuronales: Shallow

$$f_1(x) = f_2(x) = f_3(x) = ax + b$$

$$y = \omega_{14} f_1(x_1\omega_{11} + x_2\omega_{21} + b_1) + \omega_{24} f_2(x_1\omega_{12} + x_2\omega_{22} + b_2) + \omega_{34} f_3(x_1\omega_{13} + x_2\omega_{23} + b_3) + b_4$$

$$y = \omega_{14}\, a\,(x_1\omega_{11} + x_2\omega_{21} + b_1) + b$$
$$+ \omega_{24}\, a\,(x_1\omega_{12} + x_2\omega_{22} + b_2) + b$$
$$+ \omega_{34} a\,(x_1\omega_{13} + x_2\omega_{23} + b_3) + b + b_4$$

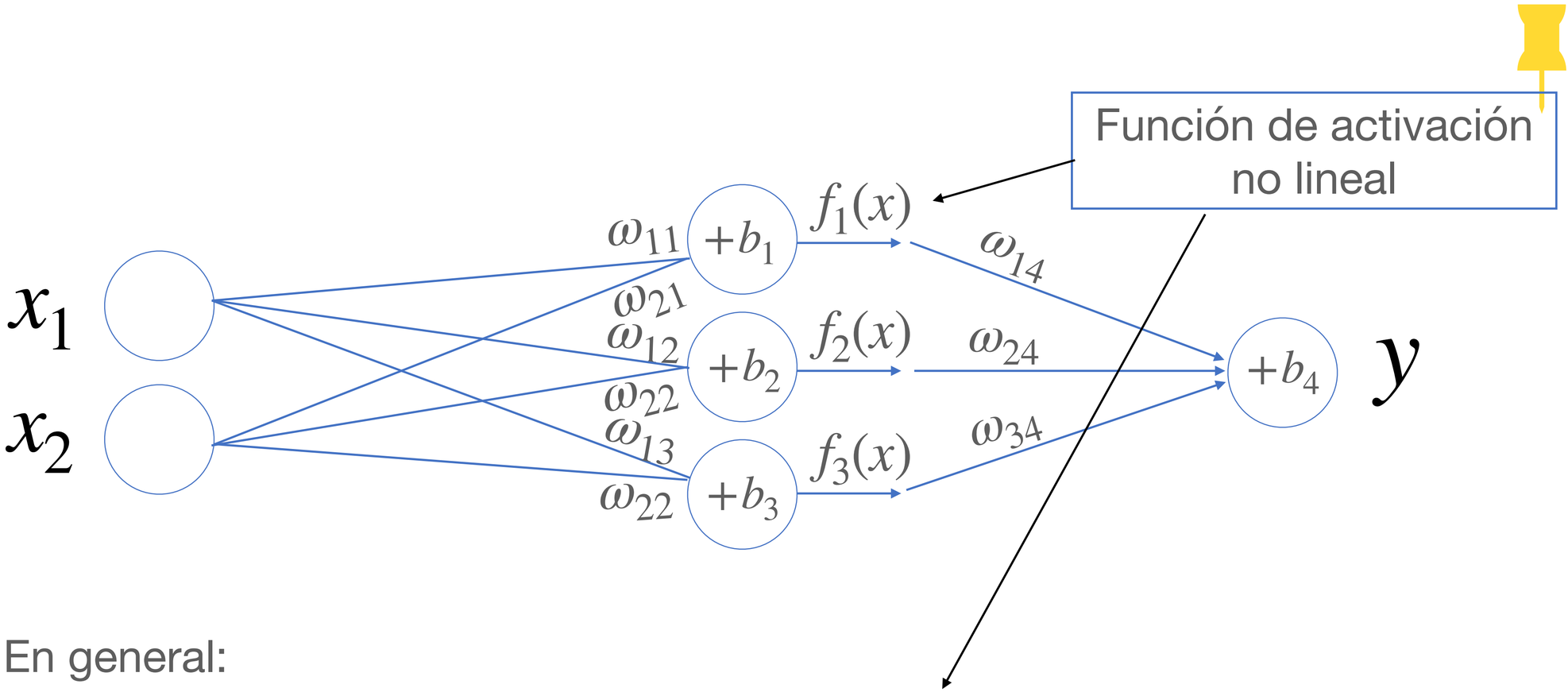
$$y = a\, x_1\omega_{11}\omega_{14} + a\, x_2\omega_{21}\omega_{14} + a\, b_1\omega_{14}$$
$$+ a\, x_1\omega_{12}\omega_{24} + a\, x_2\omega_{22}\omega_{24} + a\, b_2\omega_{24}$$
$$a\, x_1\omega_{13}\omega_{34} + a\, x_2\omega_{23}\omega_{34} + a\, b_3\omega_{34} + 3b + b_4$$

$c_1$ $c_2$ $c_3$

$$y = a\,(\omega_{11}\omega_{14} + \omega_{12}\omega_{24} + \omega_{13}\omega_{34})x_1 + a\,(\omega_{21}\omega_{14} + \omega_{22}\omega_{24} + \omega_{23}\omega_{34})x_2$$
$$+ a\,(b_1\omega_{14} + b_2\omega_{24} + b_3\omega_{34}) + 3b + b_4$$

$$y = c_1\, x_1 + c_2\, x_2 + c_3$$

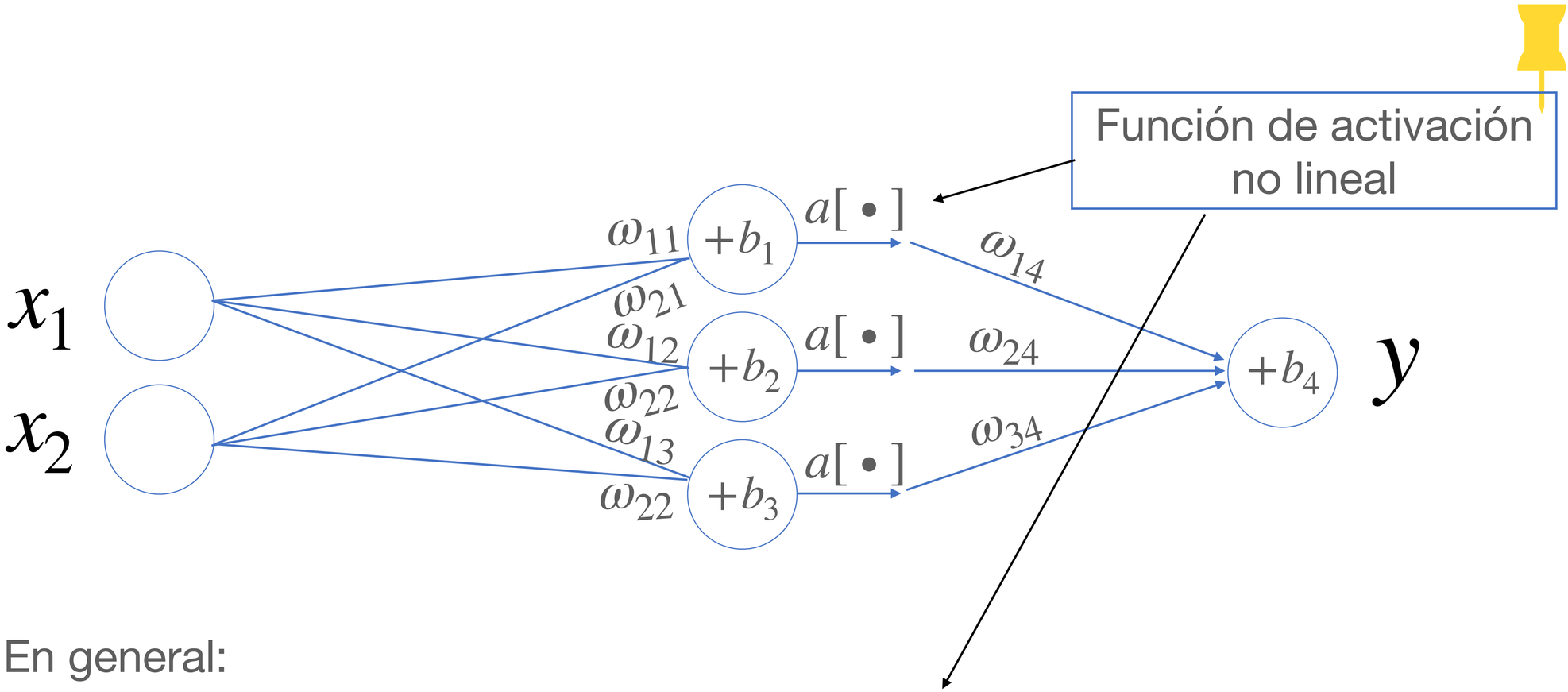
# Redes Neuronales: Función de Activación




En general:

$$f_1(x) = f_2(x) = f_3(x) \longrightarrow a[\bullet]$$

# Redes Neuronales: Función de Activación



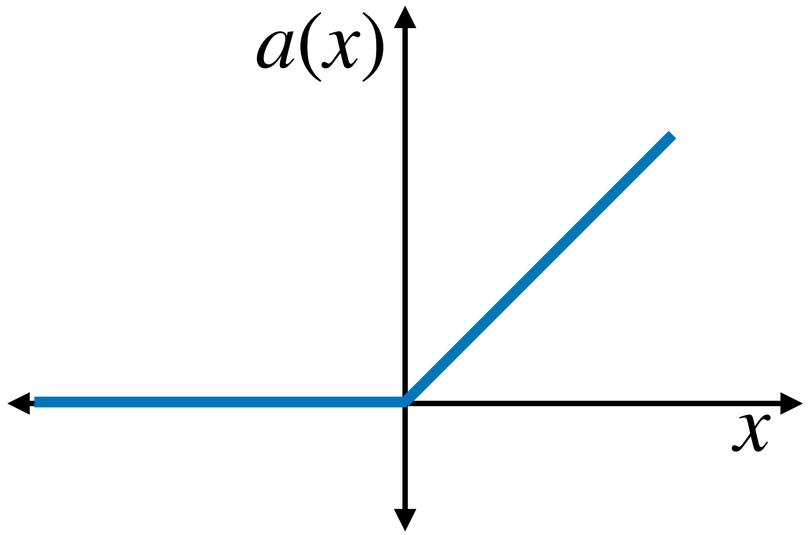

En general:

$$f_1(x) = f_2(x) = f_3(x) \longrightarrow a[\bullet]$$

# Redes Neuronales: Función de Activación

Rectified Linear Unit (ReLU)

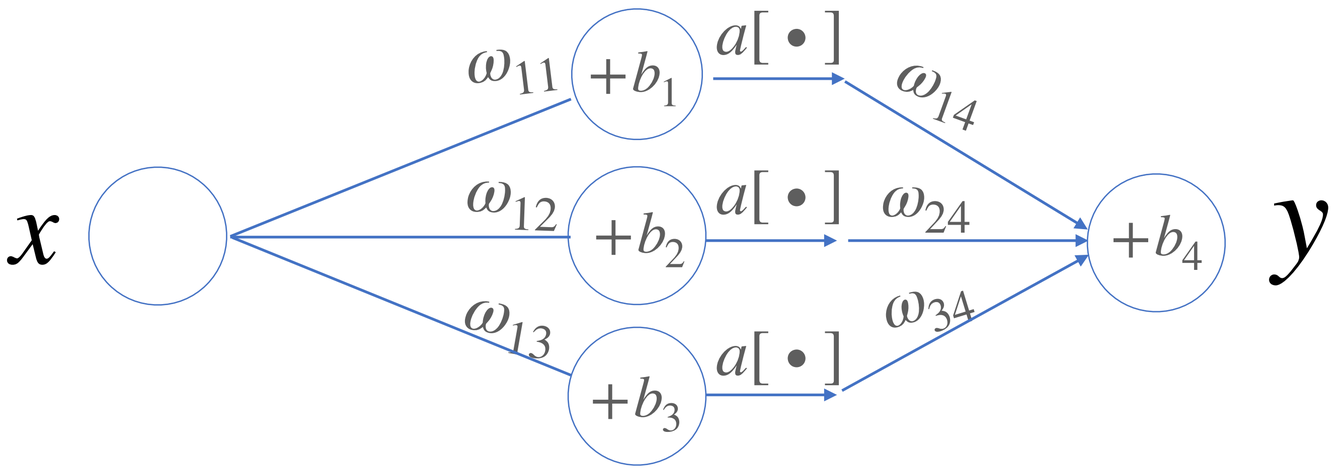
$$a(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$$

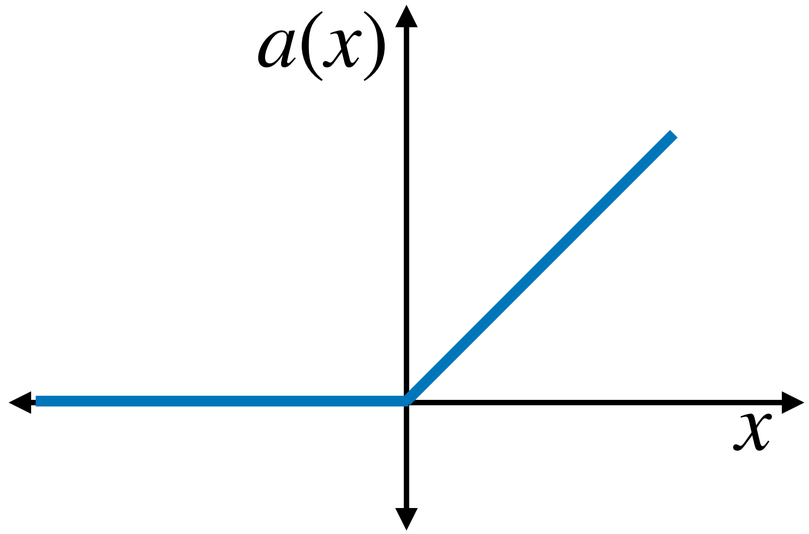


# Redes Neuronales: Función de Activación

Rectified Linear Unit (ReLU)

$$a(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$$

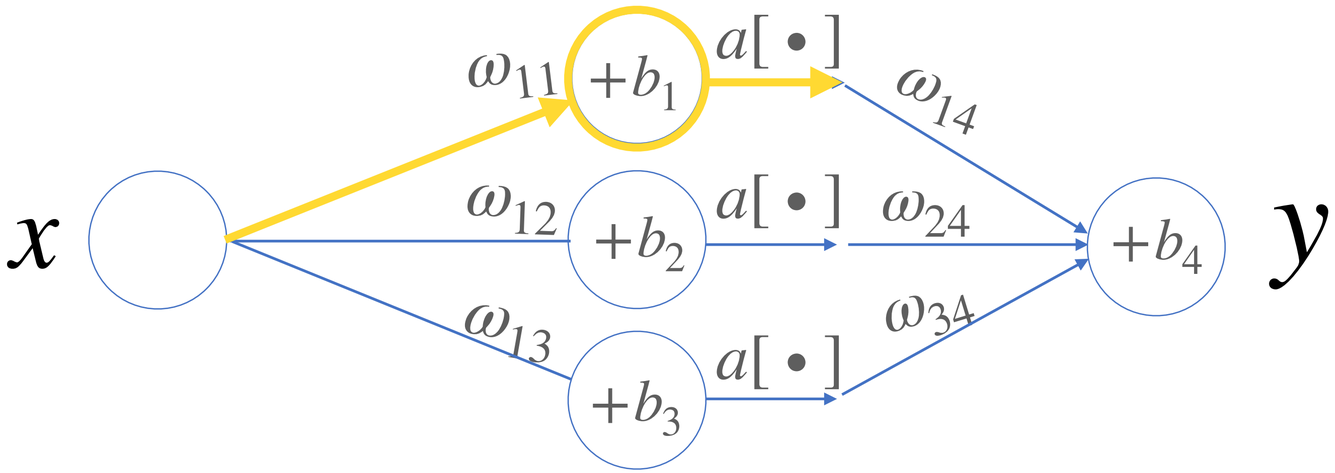


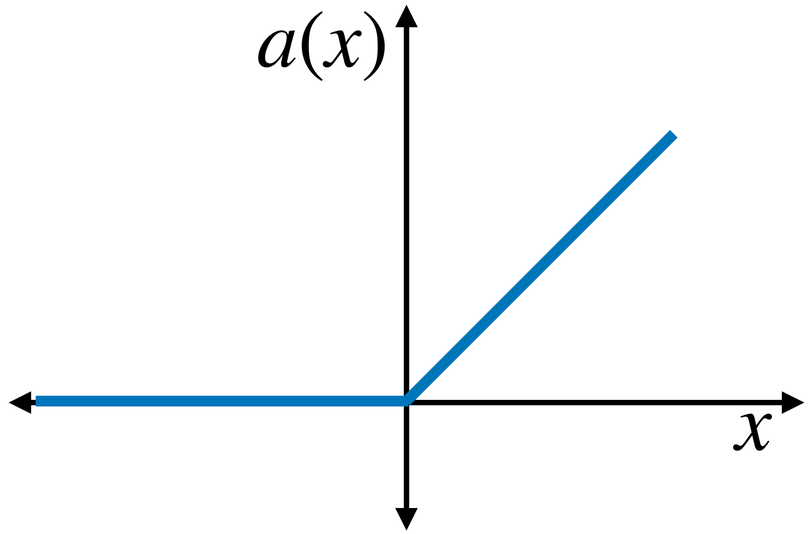
# Redes Neuronales: Función de Activación

Rectified Linear Unit (ReLU)

$$a(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$$



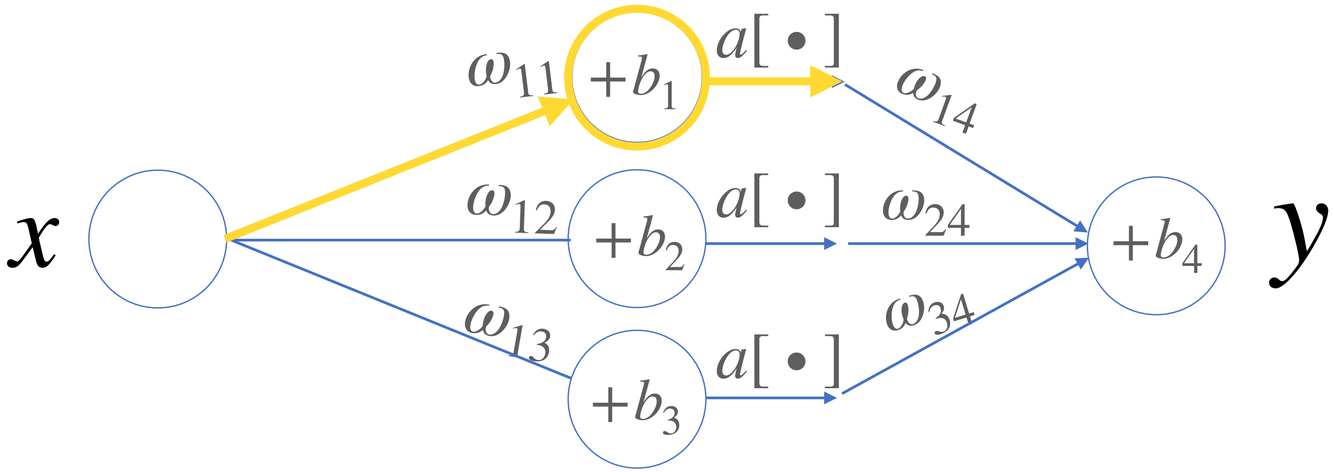

$\omega_{11} = 1.5 \quad b_1 = 3$

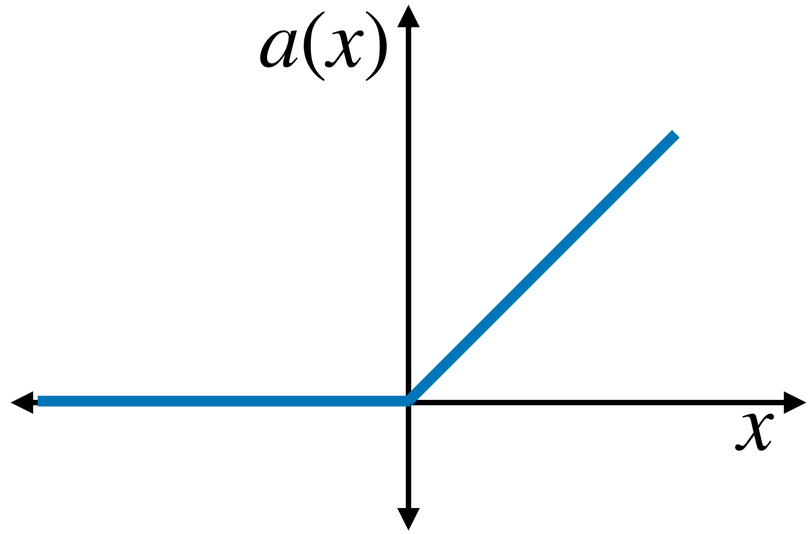


# Redes Neuronales: Función de Activación

Rectified Linear Unit (ReLU)

$$a(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$$



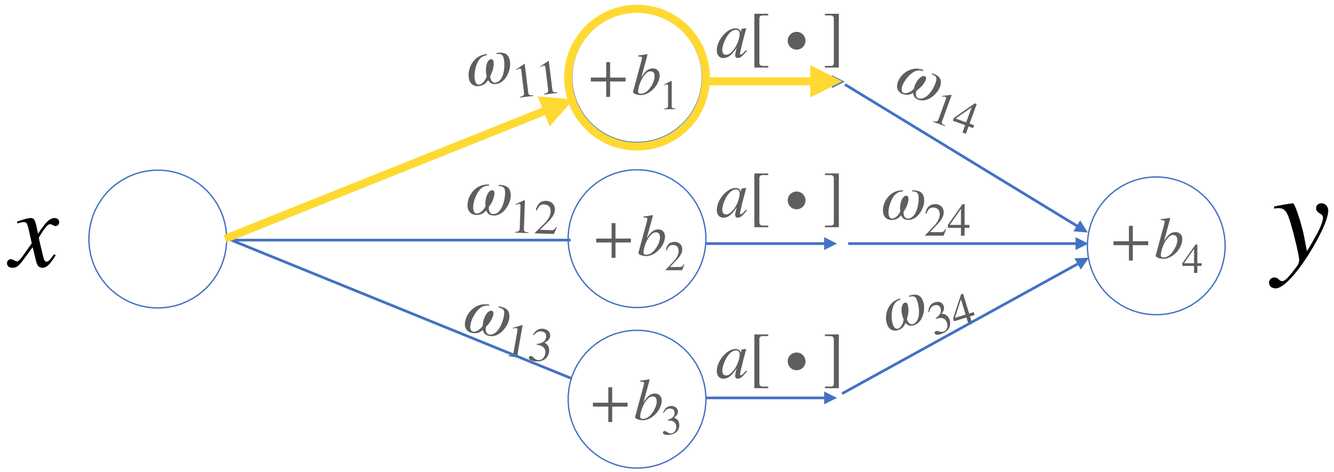



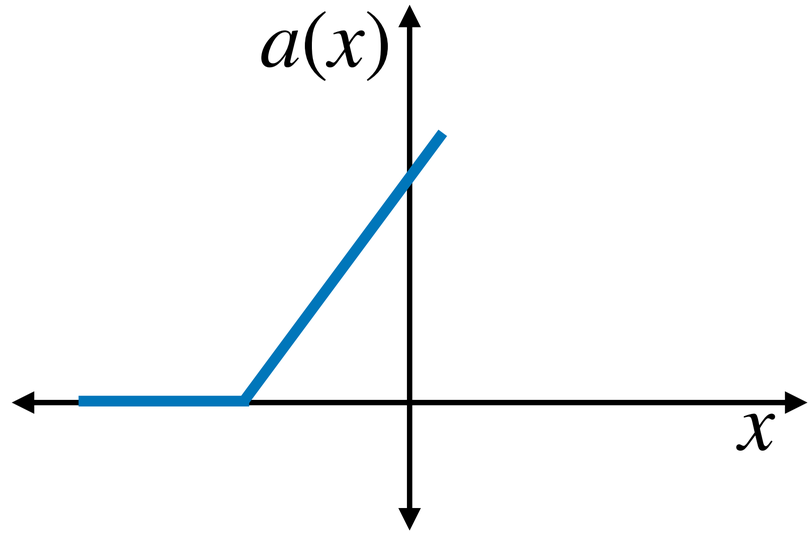

$\omega_{11} = 1.5 \quad b_1 = 3$ $\qquad a(x\,\omega_{11} + b_1) \qquad a(1.5x + 3)$

# Redes Neuronales: Función de Activación

Rectified Linear Unit (ReLU)

$$a(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$$



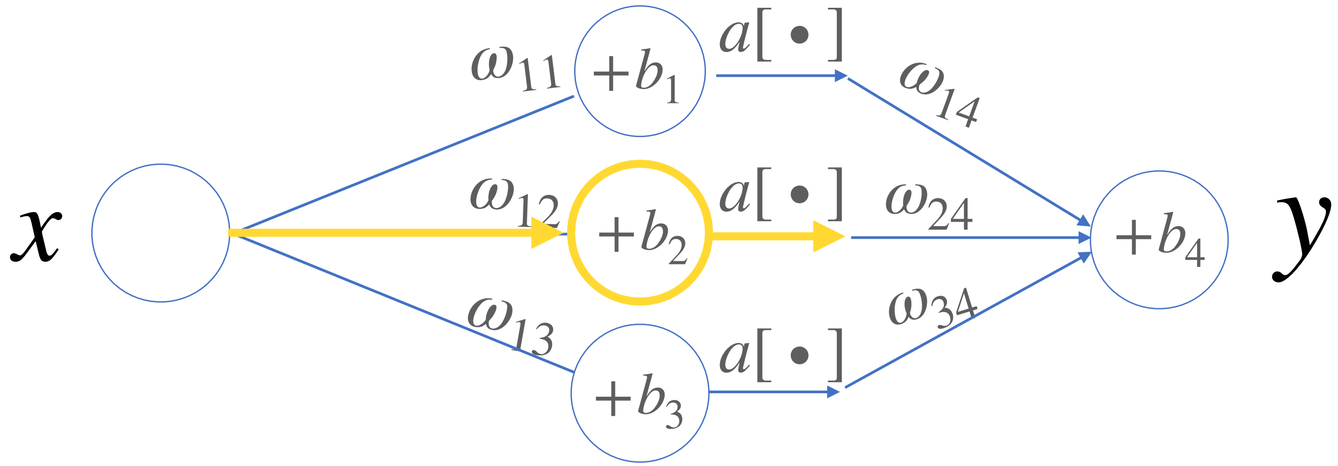



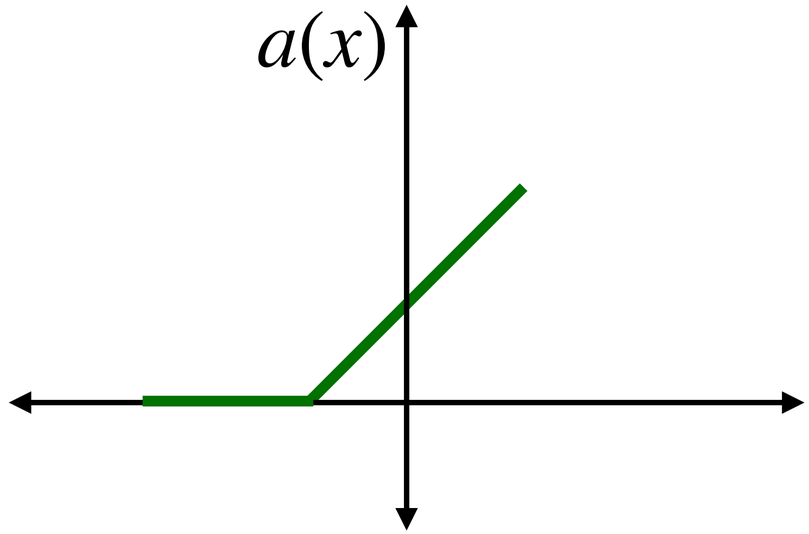

$\omega_{11} = 1.5 \quad b_1 = 3$ $\qquad a(x\,\omega_{11} + b_1) \qquad a(1.5x + 3)$

# Redes Neuronales: Función de Activación

Rectified Linear Unit (ReLU)

$$a(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$$



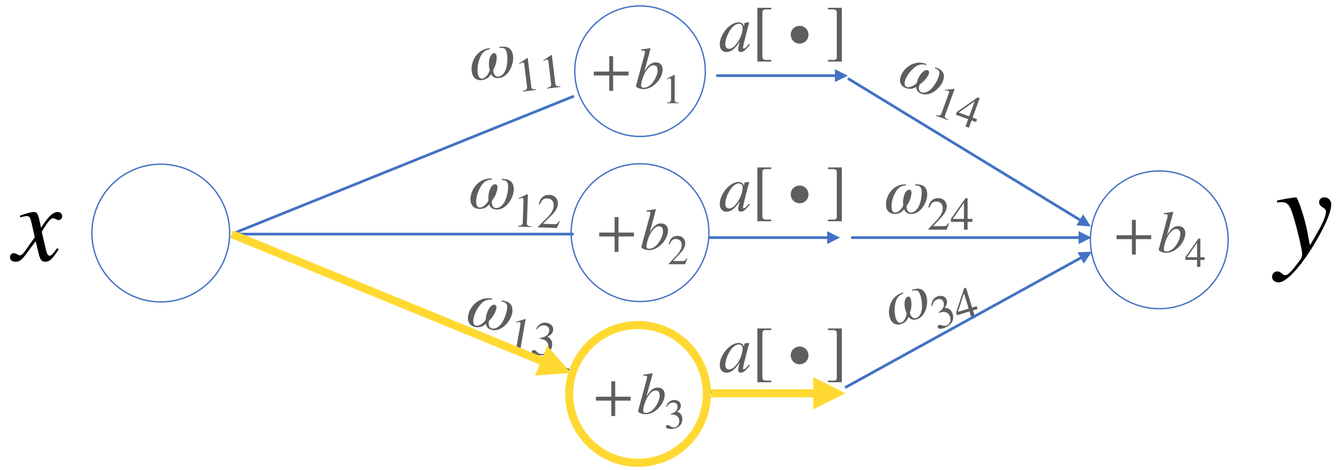

$\omega_{11} = 1.5 \quad b_1 = 3 \qquad a(x\,\omega_{11} + b_1) \qquad a(1.5x + 3)$

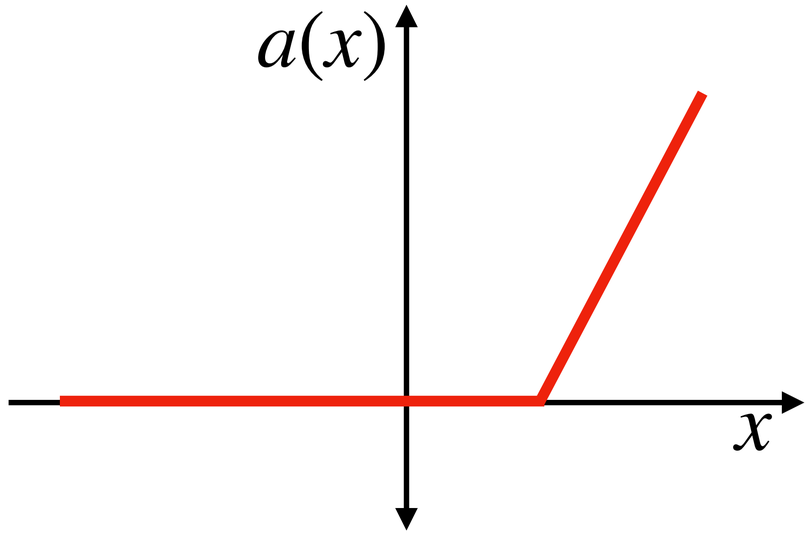
$\omega_{12} = 1 \quad\; b_2 = -1.2 \quad a(x\,\omega_{12} + b_2) \qquad a(x - 1.2)$

# Redes Neuronales: Función de Activación

Rectified Linear Unit (ReLU)

$$a(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$$



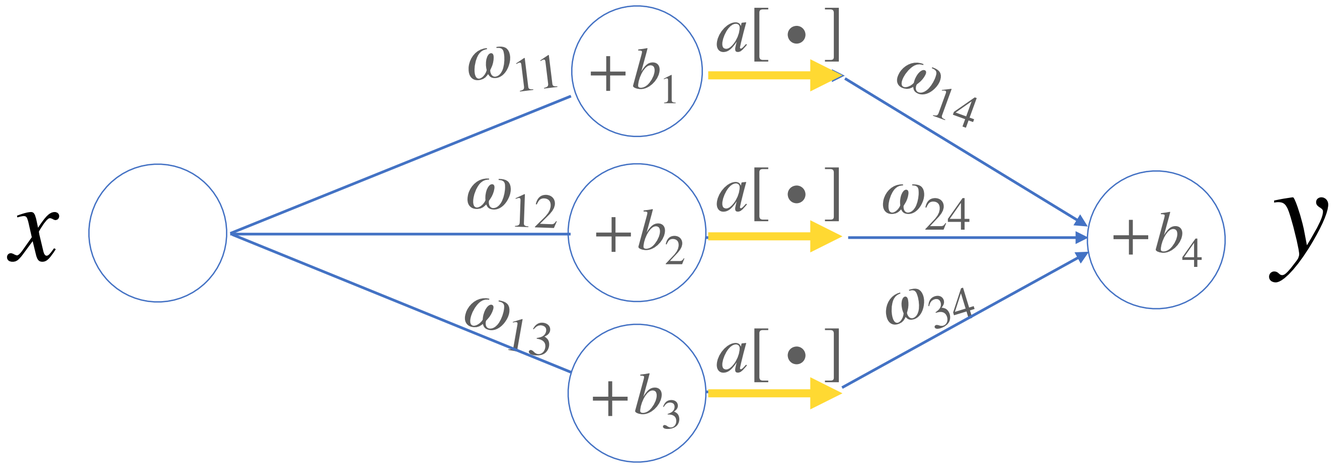



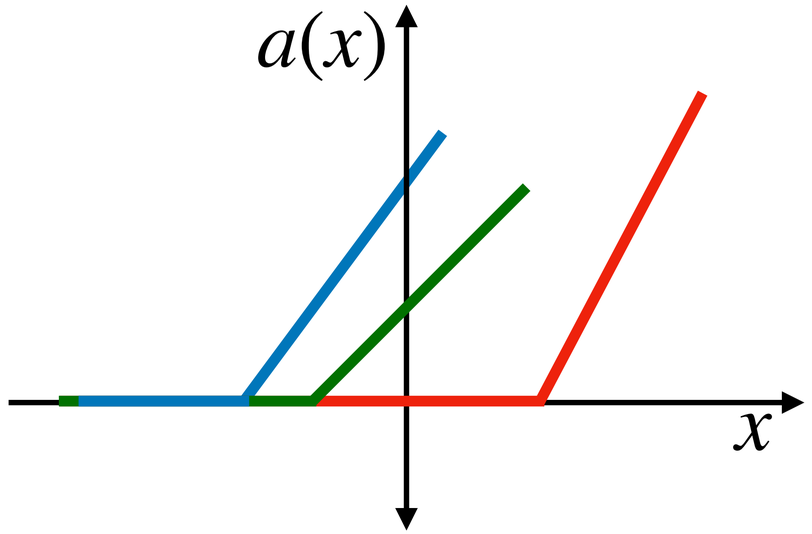

| | | | |
|---|---|---|---|
| $\omega_{11} = 1.5$ | $b_1 = 3$ | $a(x\,\omega_{11} + b_1)$ | $a(1.5x + 3)$ |
| $\omega_{12} = 1$ | $b_2 = -1.2$ | $a(x\,\omega_{12} + b_2)$ | $a(x - 1.2)$ |
| $\omega_{13} = 2$ | $b_3 = -4$ | $a(x\,\omega_{13} + b_3)$ | $a(2x - 4)$ |

# Redes Neuronales: Función de Activación

Rectified Linear Unit (ReLU)

$$a(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$$



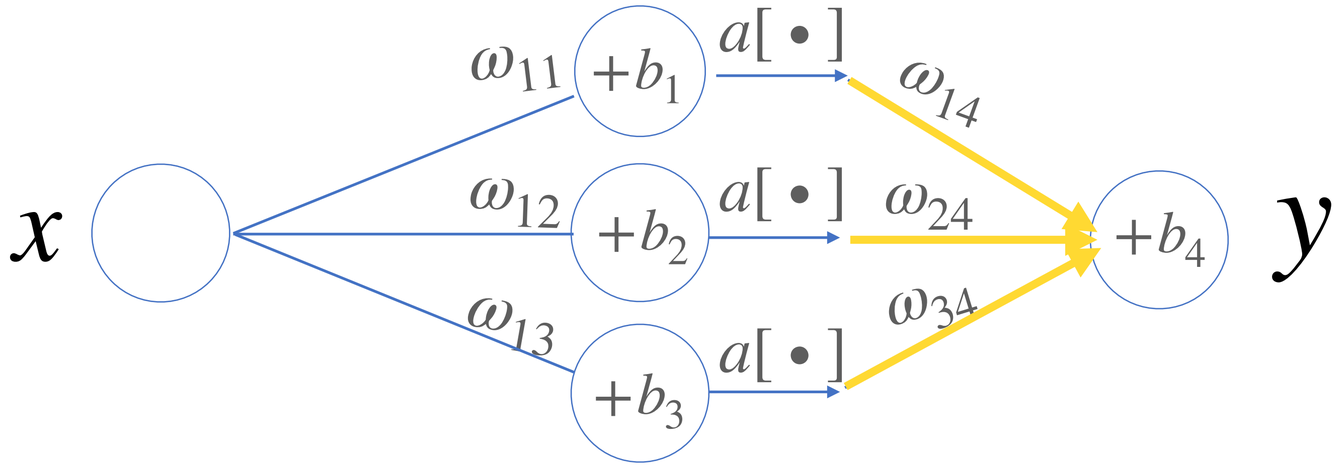



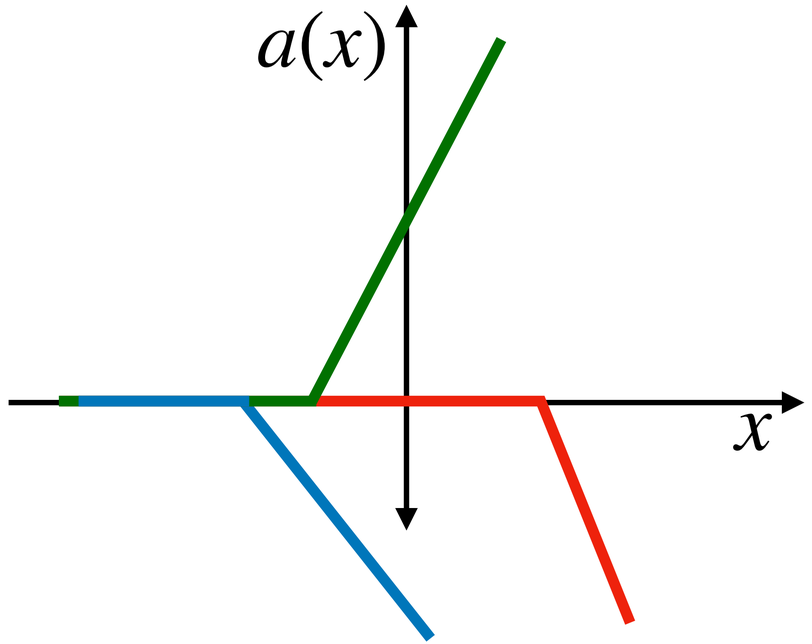

| | | | |
|---|---|---|---|
| $\omega_{11} = 1.5$ | $b_1 = 3$ | $a(x\,\omega_{11} + b_1)$ | $a(1.5x + 3)$ |
| $\omega_{12} = 1$ | $b_2 = -1.2$ | $a(x\,\omega_{12} + b_2)$ | $a(x - 1.2)$ |
| $\omega_{13} = 2$ | $b_3 = -4$ | $a(x\,\omega_{13} + b_3)$ | $a(2x - 4)$ |

# Redes Neuronales: Función de Activación

Rectified Linear Unit (ReLU)

$$a(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$$




$a(1.5x + 3)$ $\quad \omega_{14} = -0.6$

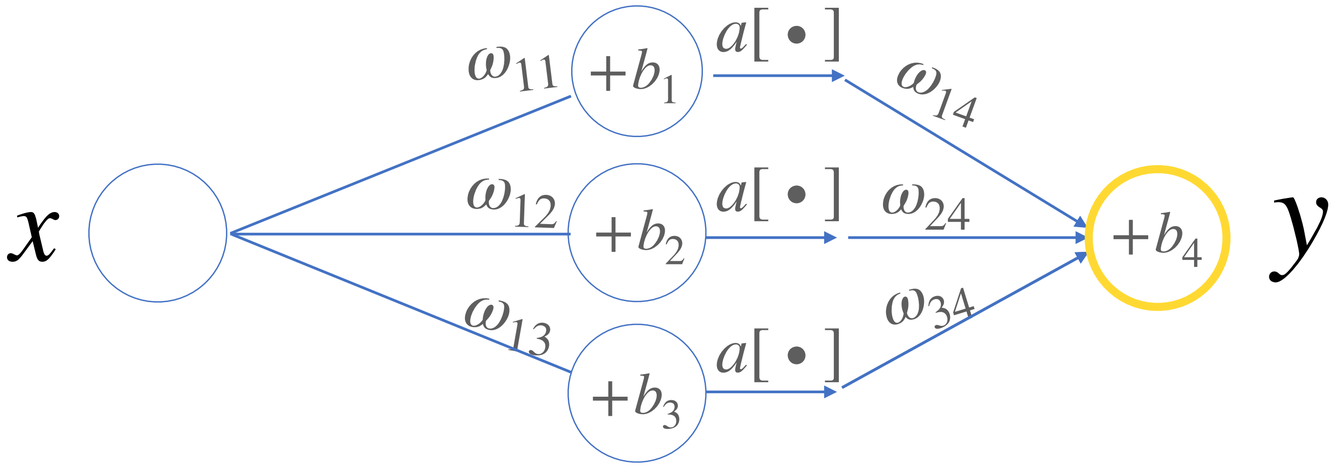
$a(x - 1.2)$ $\quad \omega_{24} = 1.7$

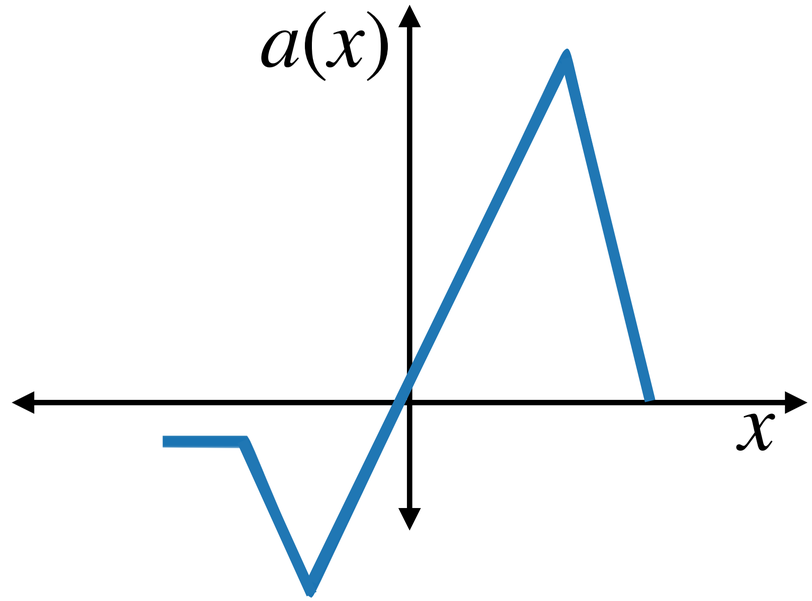
$a(2x - 4)$ $\quad \omega_{34} = -1.2$

# Redes Neuronales: Función de Activación

Rectified Linear Unit (ReLU)

$$a(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$$




$a(1.5x + 3)$ $\quad \omega_{14} = -0.6 \quad b_4 = -0.5$

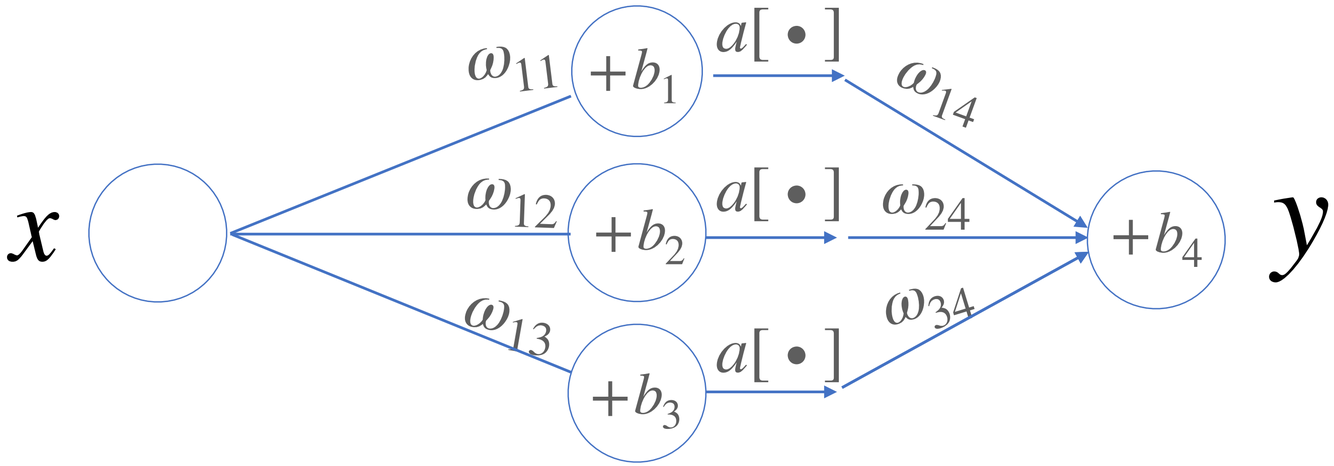
$a(x - 1.2)$ $\quad \omega_{24} = 1.7$

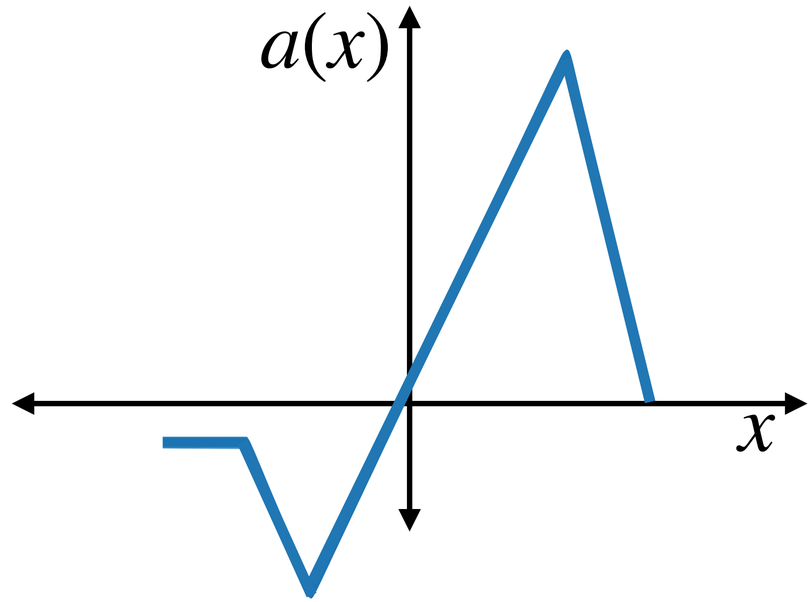
$a(2x - 4)$ $\quad \omega_{34} = -1.2$

# Redes Neuronales: Función de Activación

Rectified Linear Unit (ReLU)

$$a(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$$




$a(1.5x + 3)$ $\quad \omega_{14} = -0.6 \quad b_4 = -0.5$

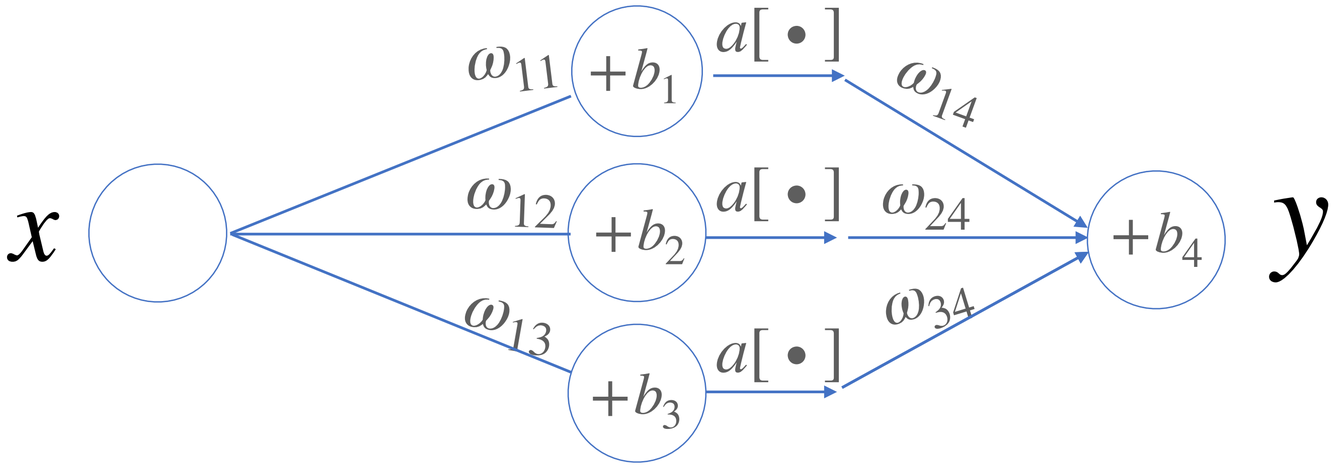
$a(x - 1.2)$ $\quad \omega_{24} = 1.7$

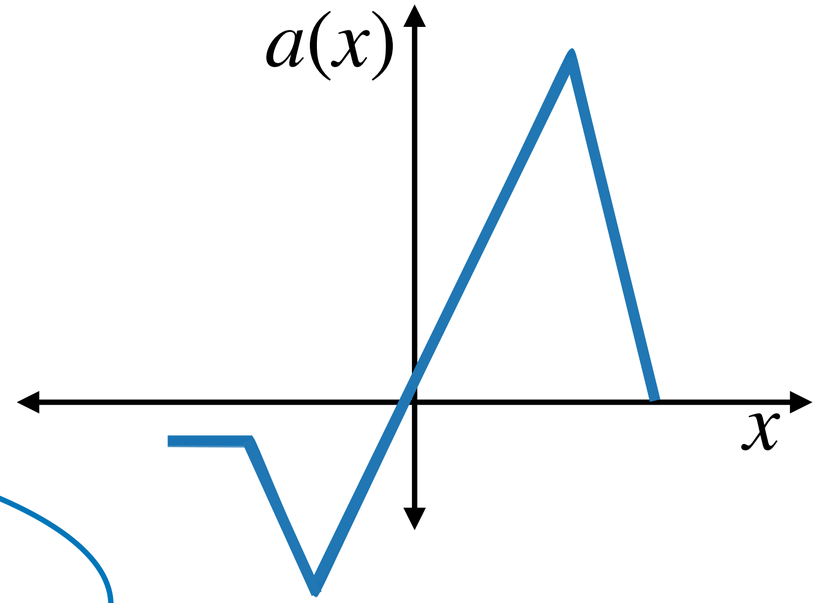
$a(2x - 4)$ $\quad \omega_{34} = -1.2$

# Redes Neuronales: Función de Activación

Rectified Linear Unit (ReLU)

$$a(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$$




$a(1.5x + 3)$

$a(x - 1.2)$

$a(2x - 4)$

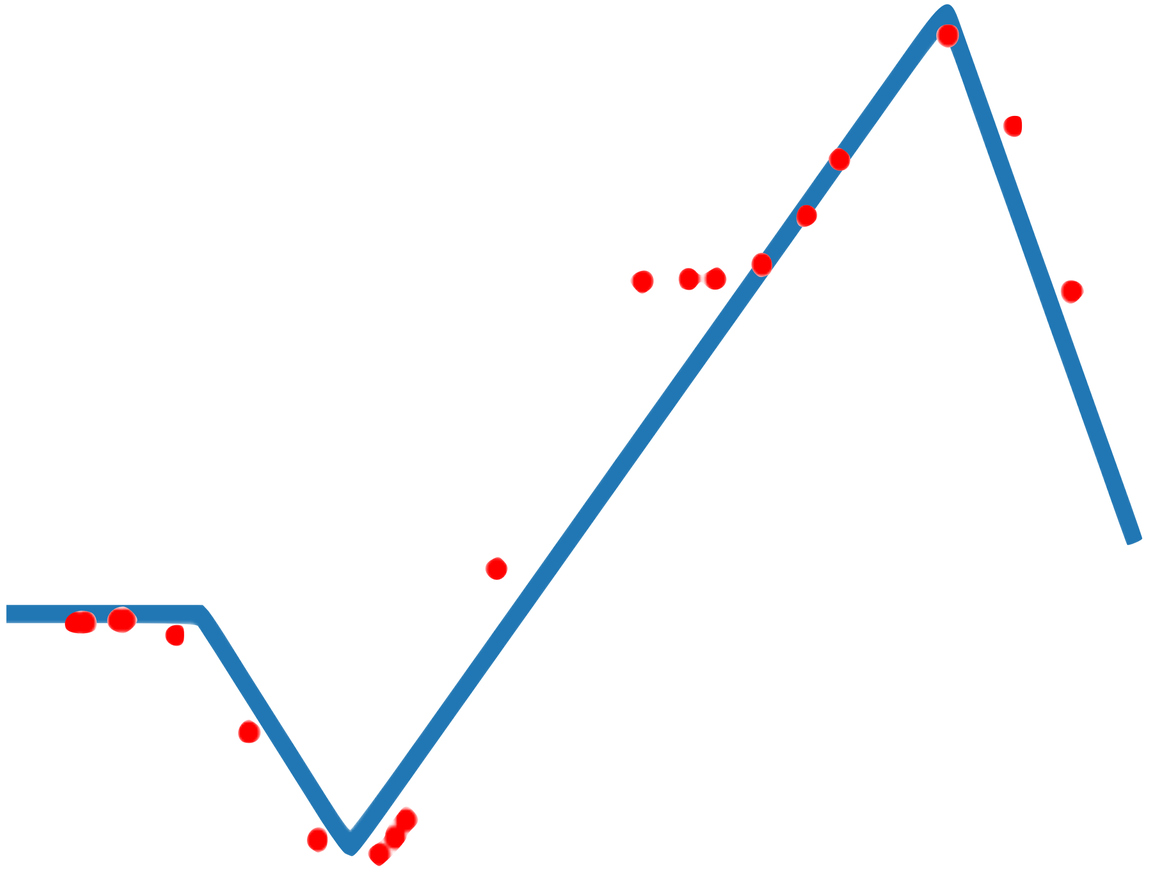
$\omega_{14} = -0.6$
$\omega_{24} = 1.7$
$\omega_{34} = -1.2$
$b_4 = -0.5$

$\omega_{11} = 1.5 \quad b_1 = 3$
$\omega_{12} = 1 \quad b_2 = -1.2$
$\omega_{13} = 2 \quad b_3 = -4$

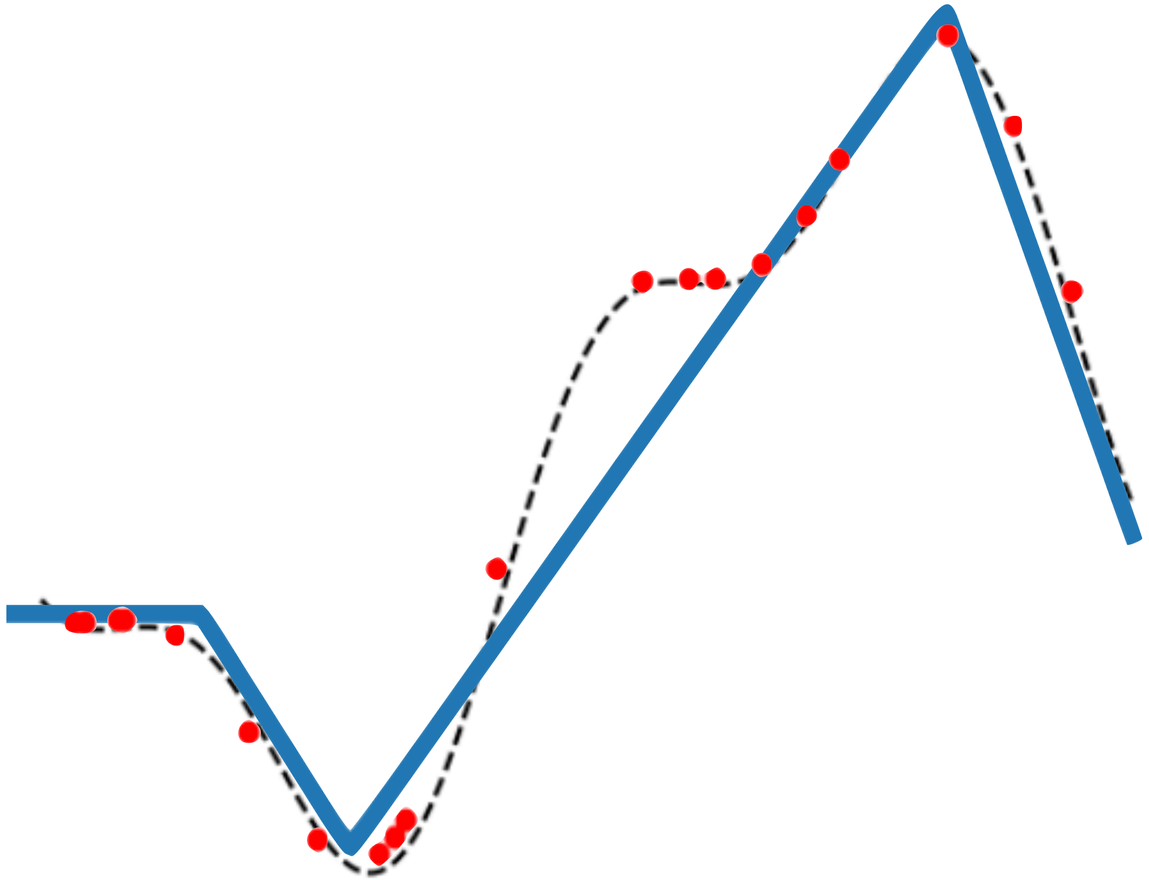
$\theta$ = **Parámetros de nuestro modelo**

# Redes Neuronales: Ajuste

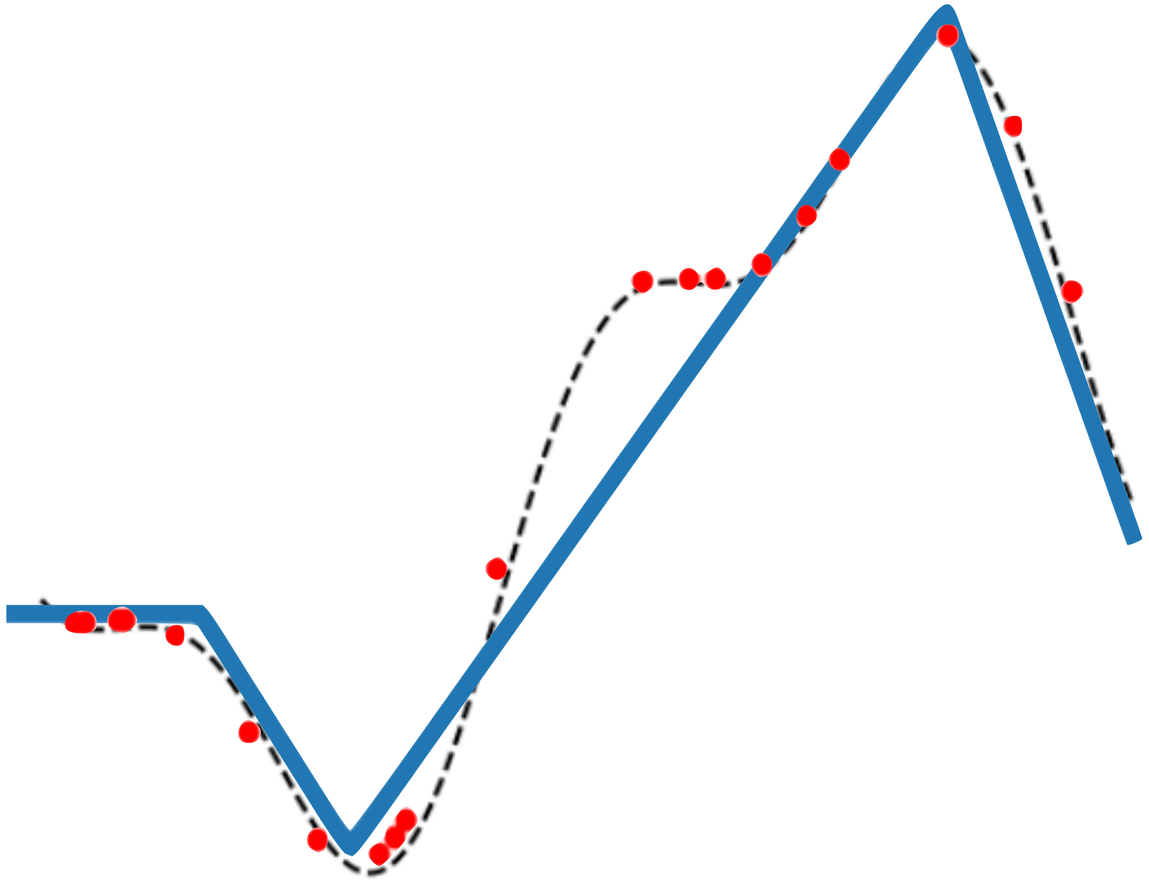
# Redes Neuronales: Ajuste



# Redes Neuronales: Ajuste

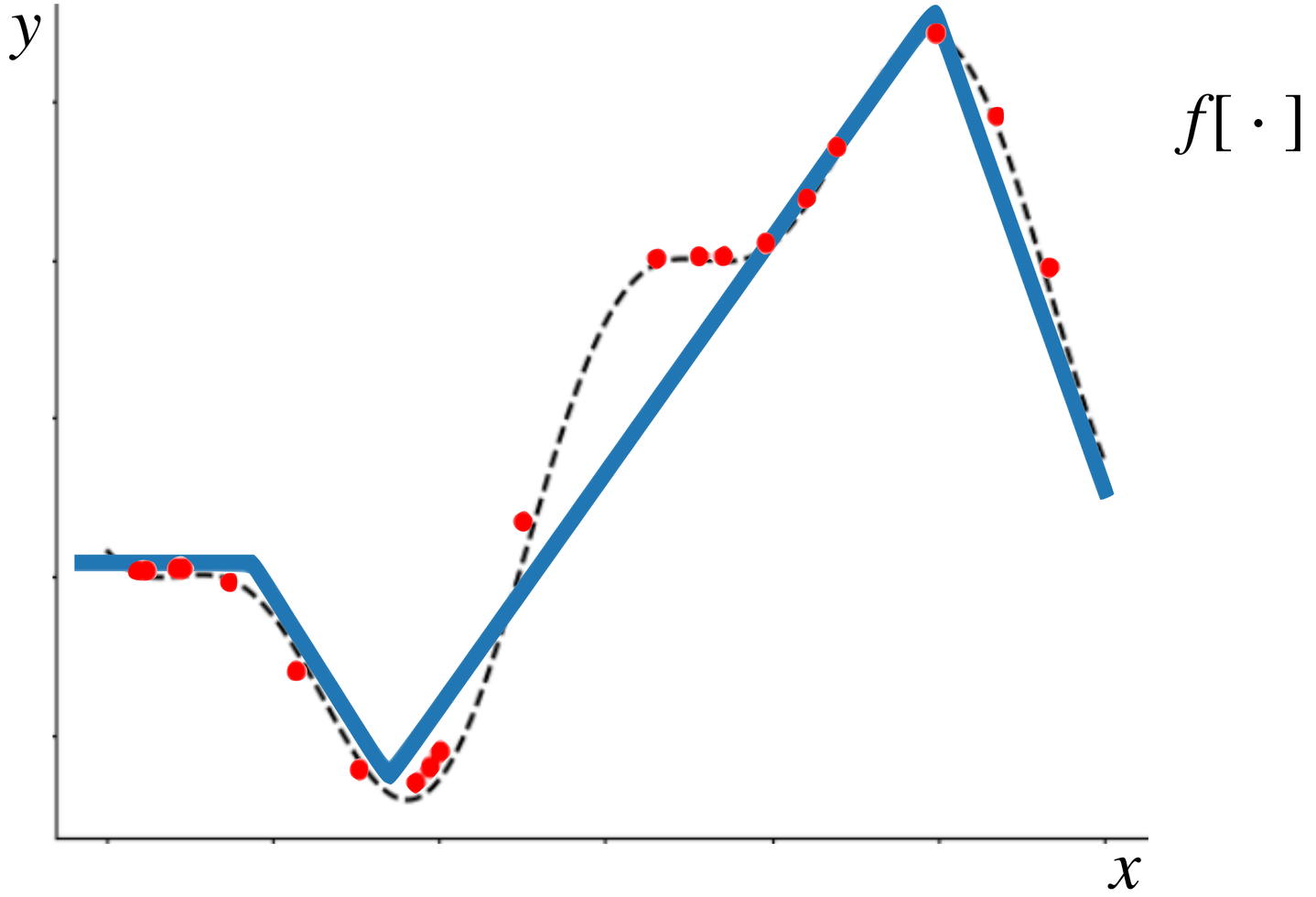


# Redes Neuronales: Ajuste



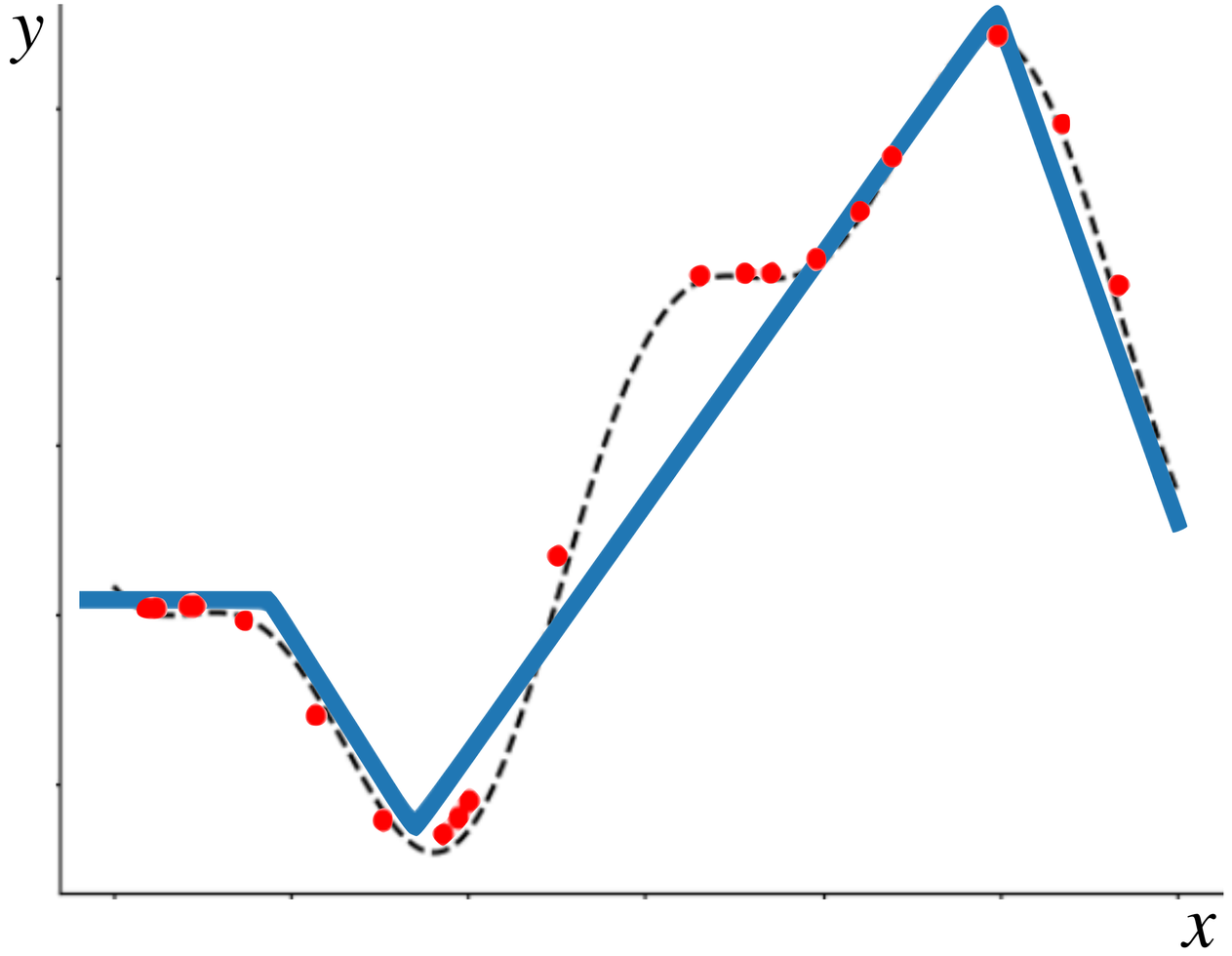
Como obtenemos el fit azul a partir de los datos que vienen de la función en negro

# Redes Neuronales: Ajuste




Como obtenemos el fit azul a partir de los datos que vienen de la función en negro

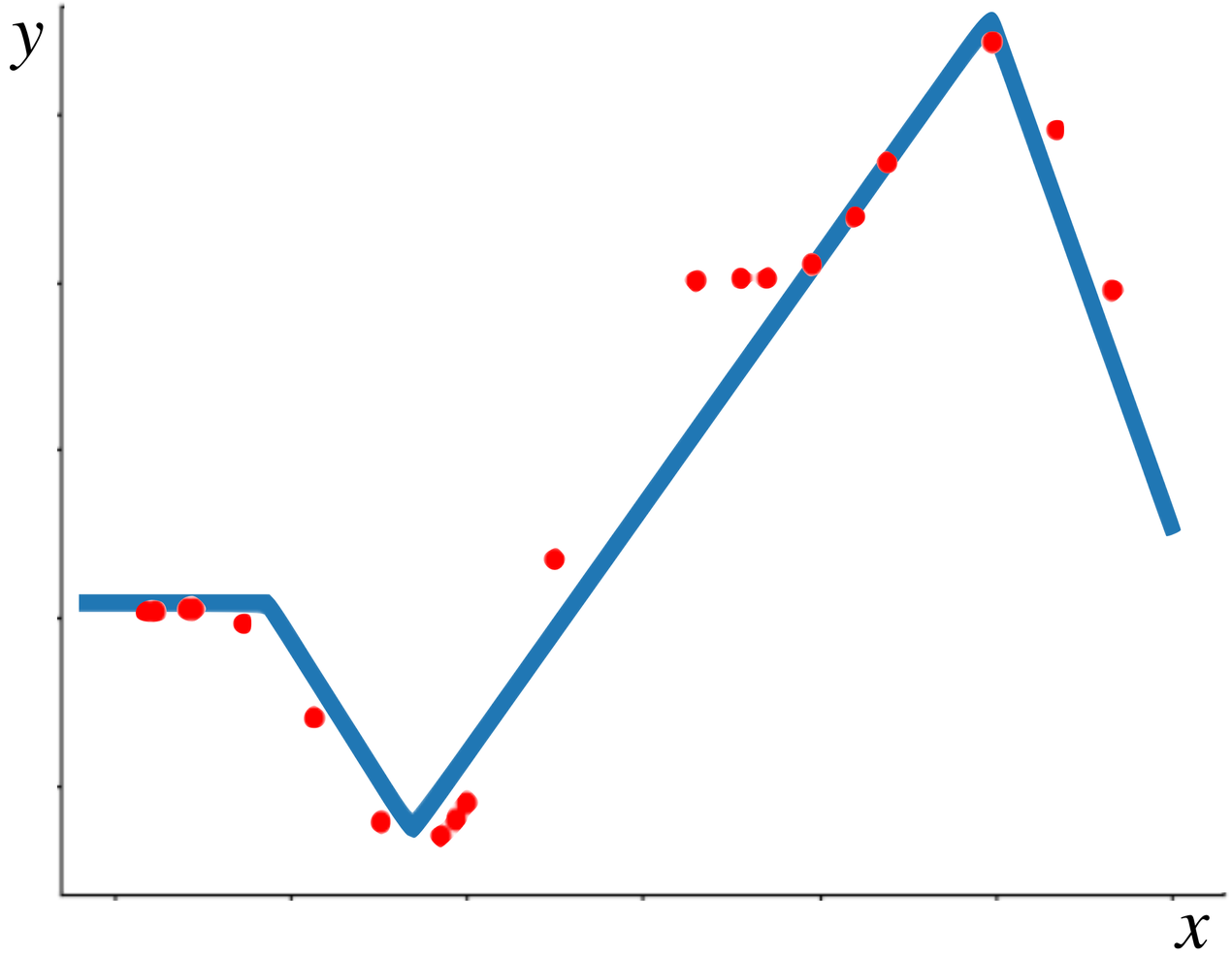
# Redes Neuronales: Ajuste




$$f[\cdot]$$
$$\hat{y} = f[x, \theta]$$

Como obtenemos el fit azul a partir de los datos que vienen de la función en negro

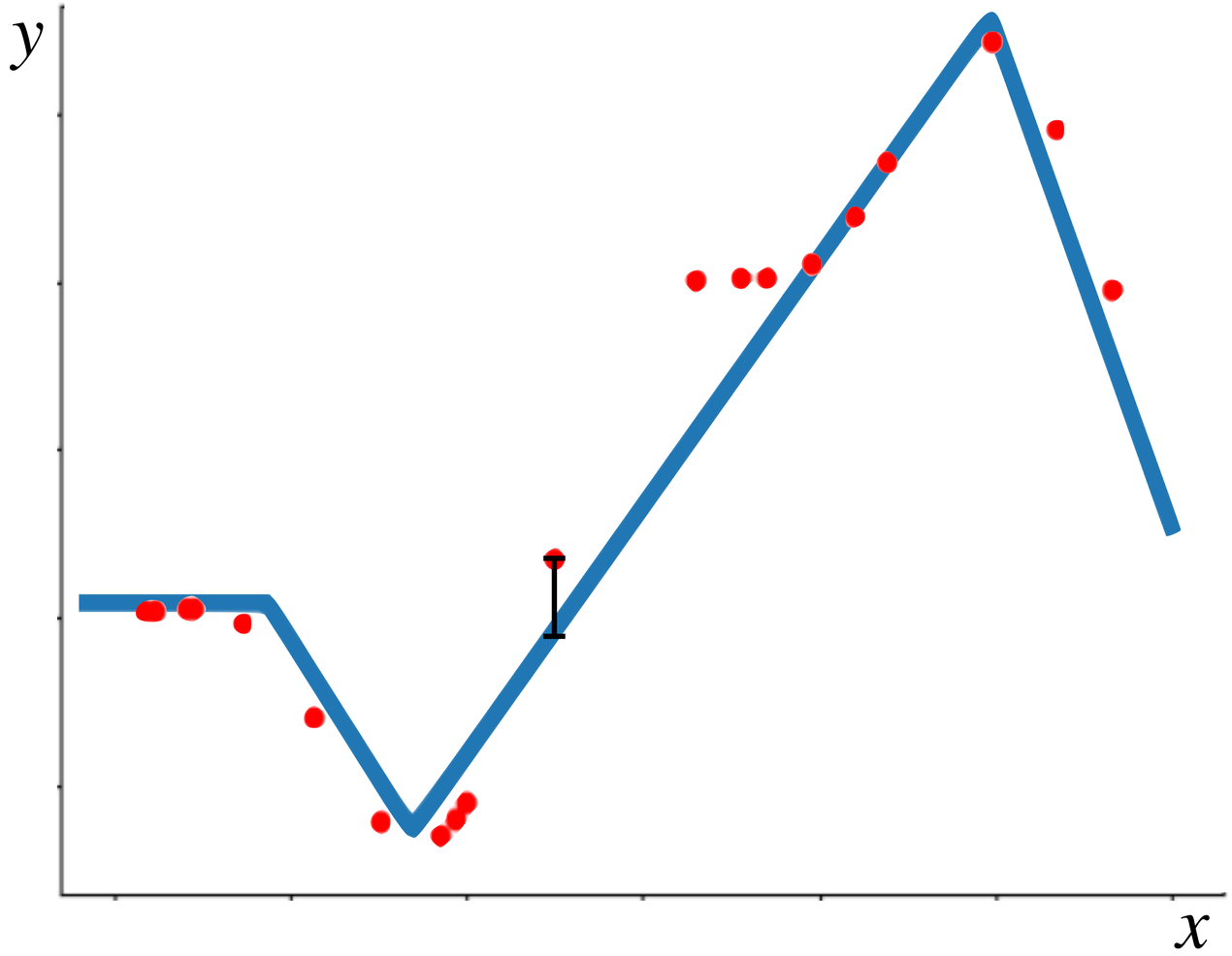
# Redes Neuronales: Ajuste




$$\hat{y} = f[x, \theta]$$

Como obtenemos el fit azul a partir de los datos que vienen de la función en negro

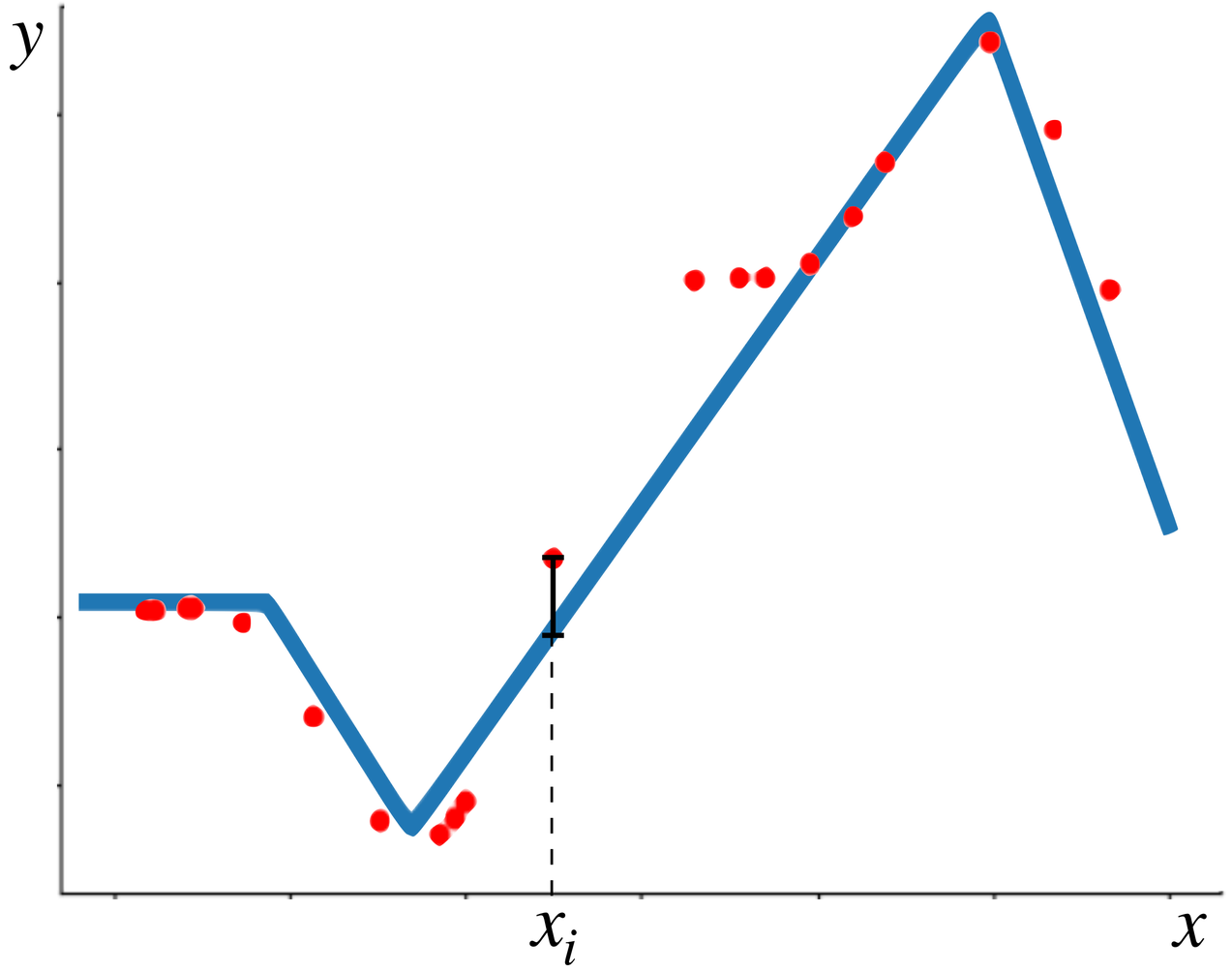
# Redes Neuronales: Ajuste




$$\hat{y} = f[x, \theta]$$

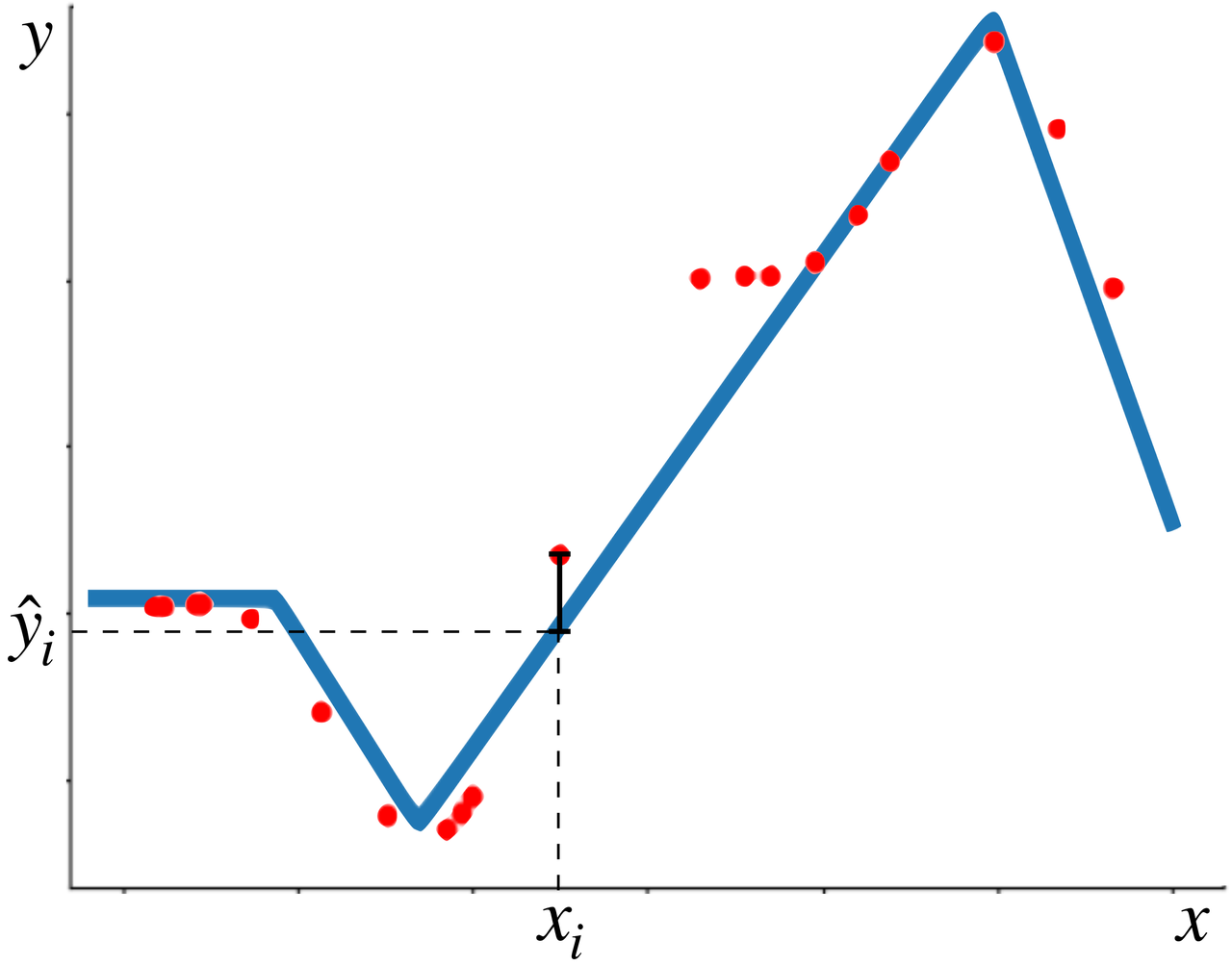
Como obtenemos el fit azul a partir de los datos que vienen de la función en negro

# Redes Neuronales: Ajuste



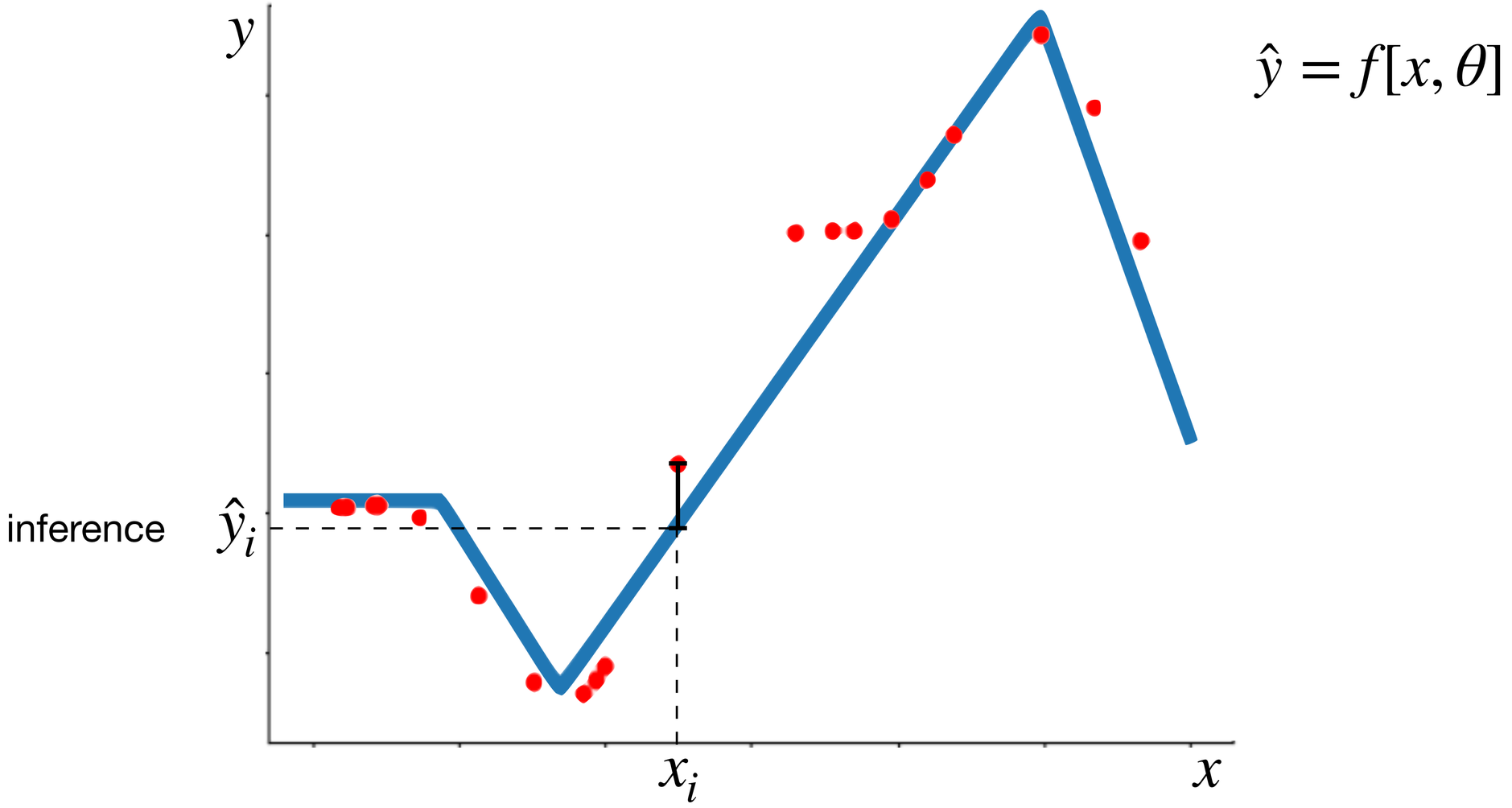

$$\hat{y} = f[x, \theta]$$

# Redes Neuronales: Ajuste

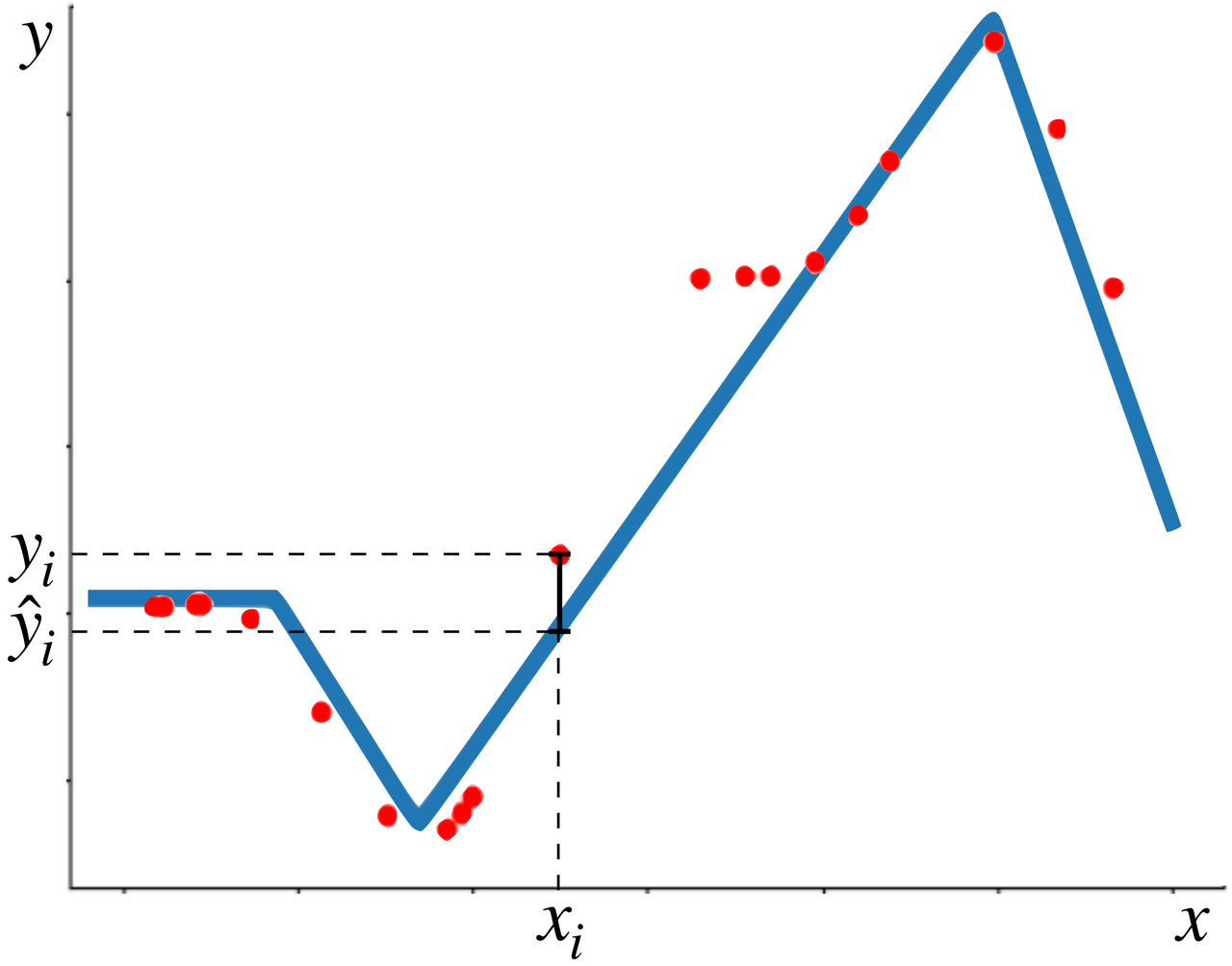



$$\hat{y} = f[x, \theta]$$

# Redes Neuronales: Ajuste

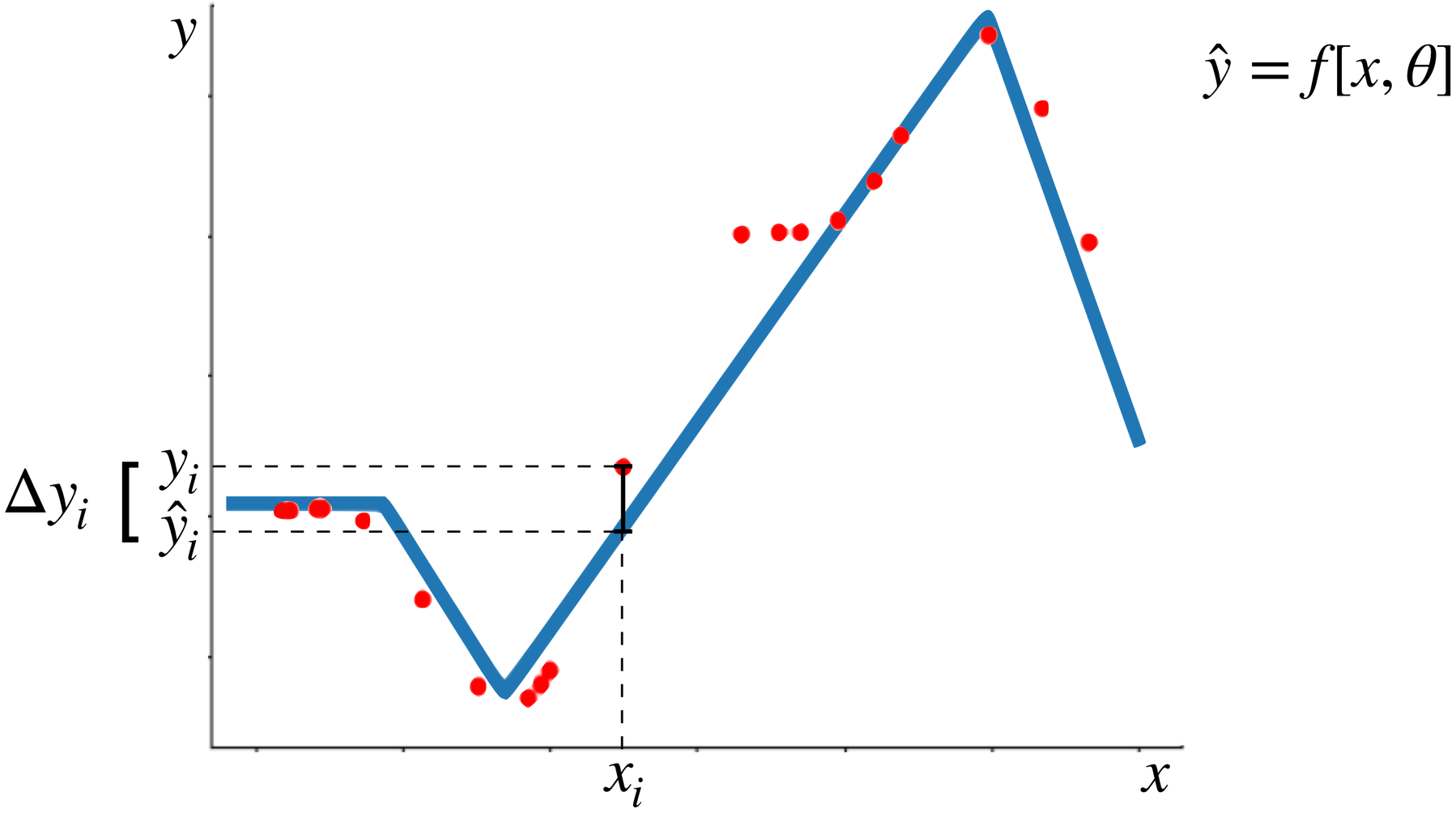


# Redes Neuronales: Ajuste

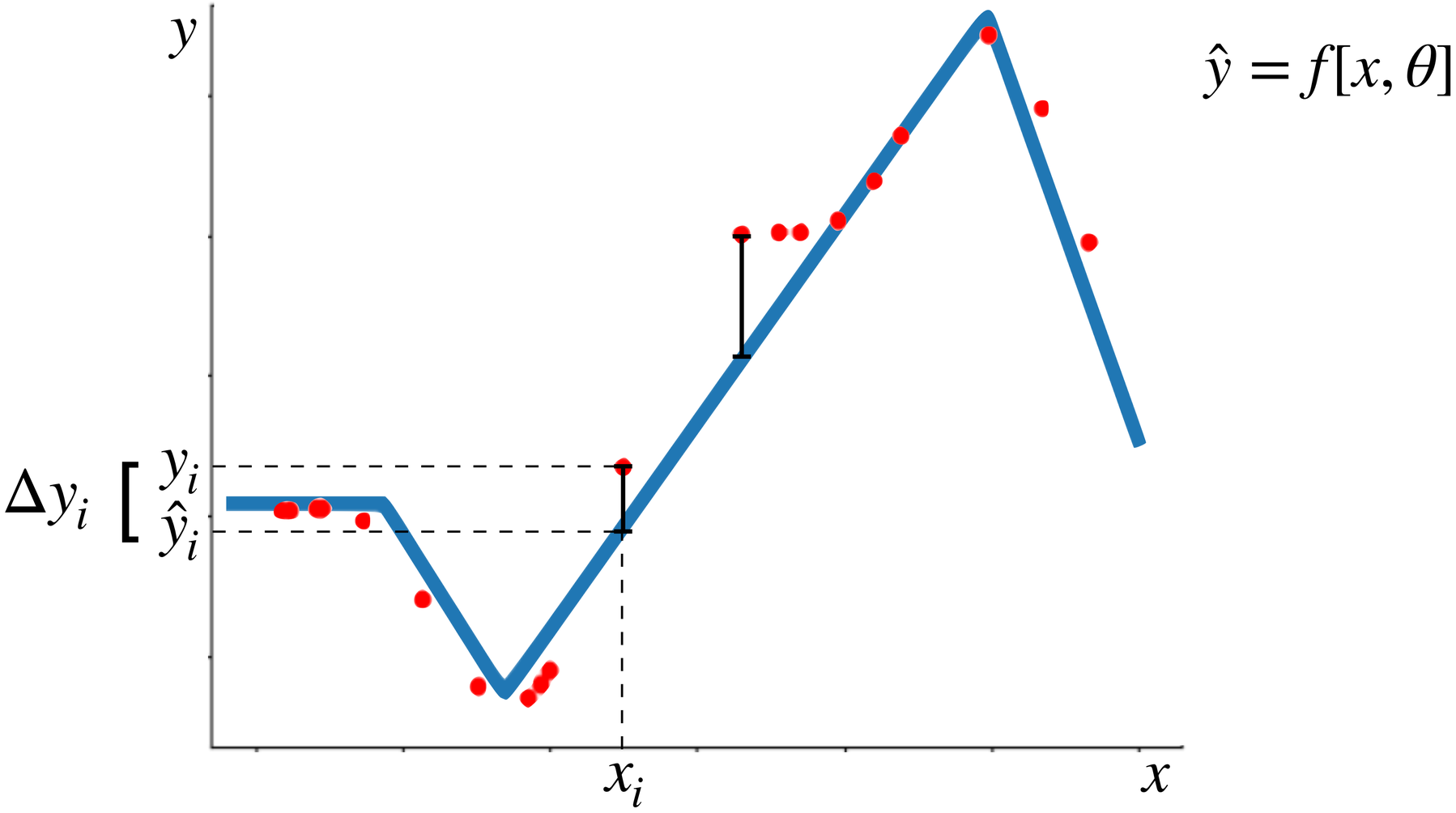



$$\hat{y} = f[x, \theta]$$

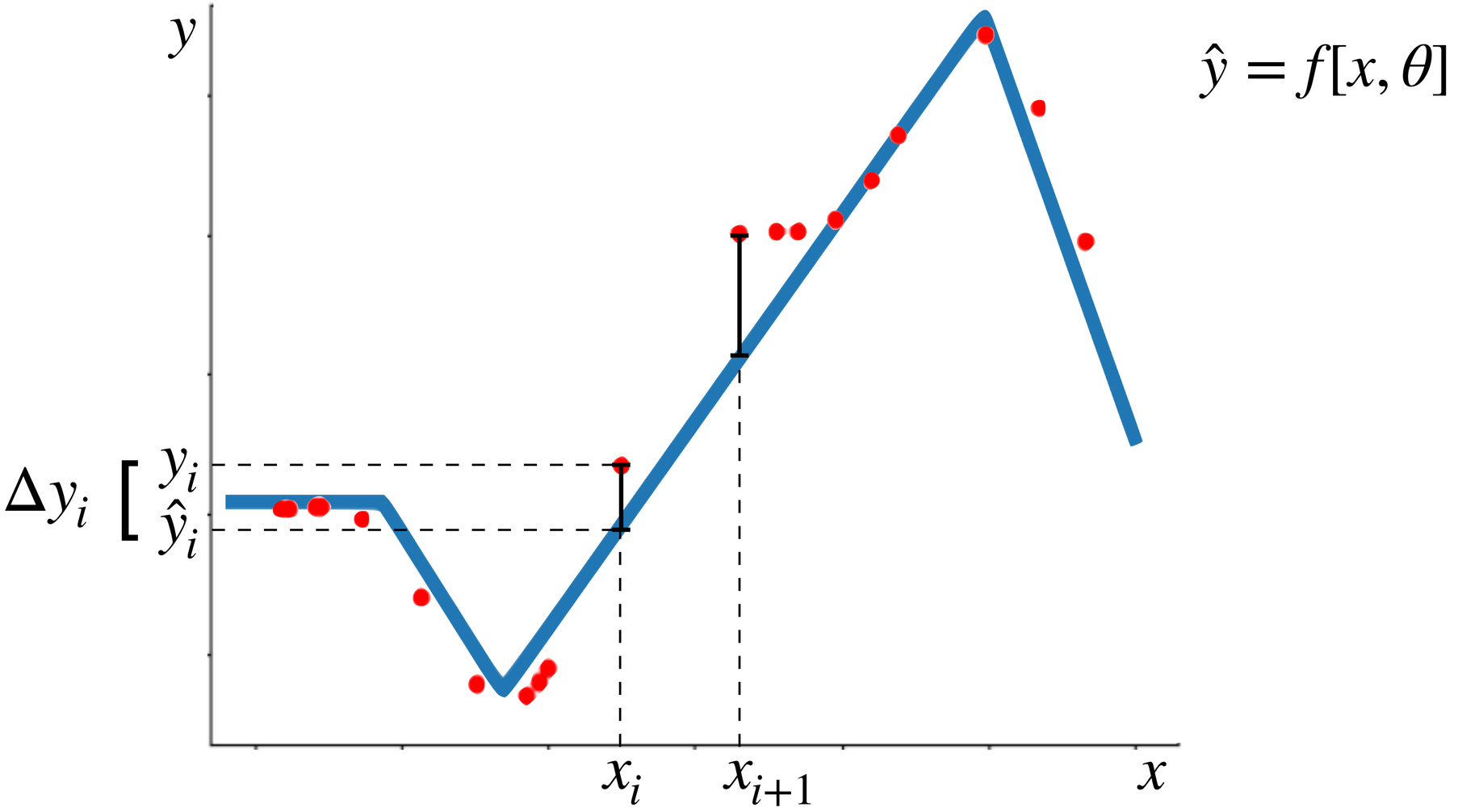
# Redes Neuronales: Ajuste



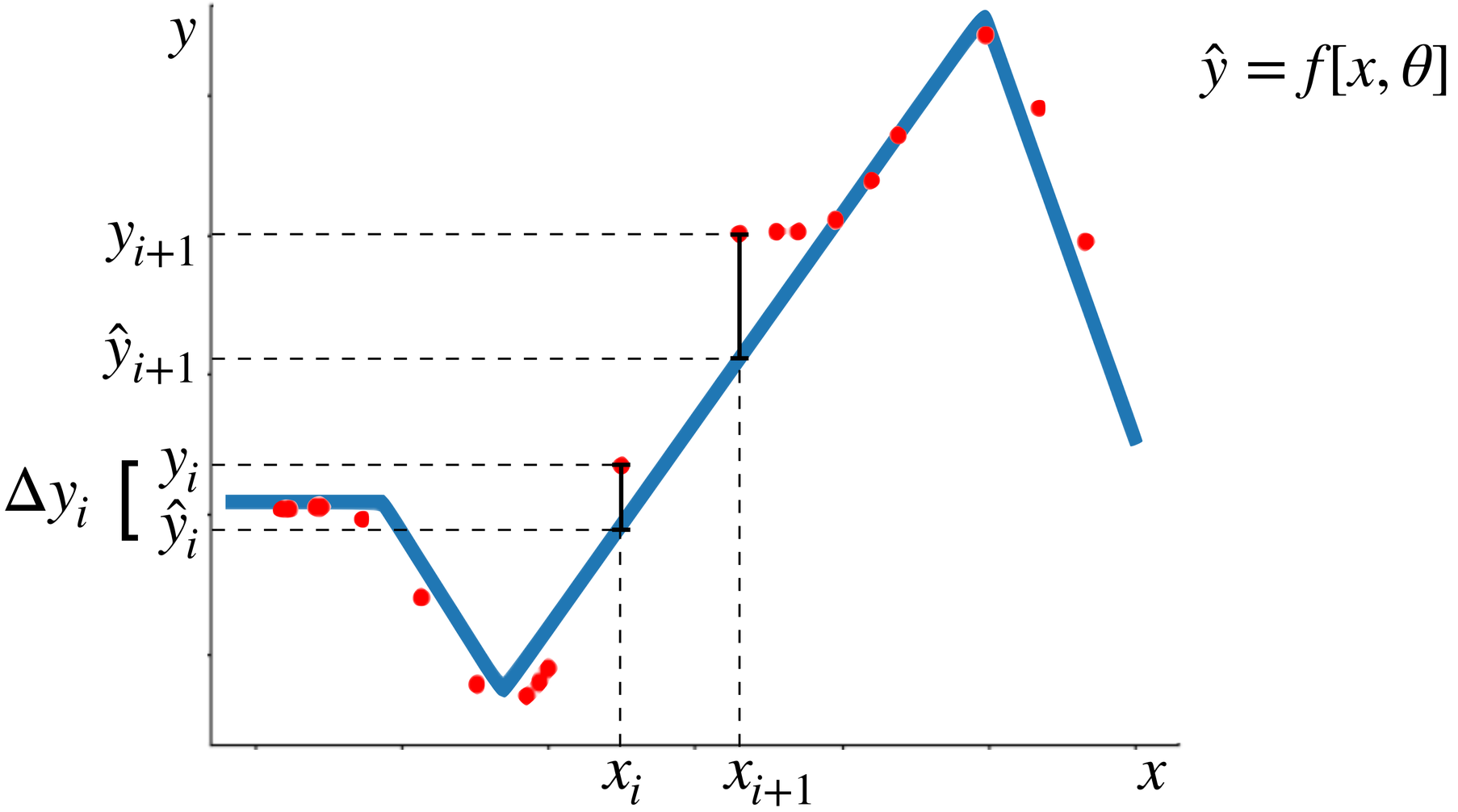
# Redes Neuronales: Ajuste



# Redes Neuronales: Ajuste

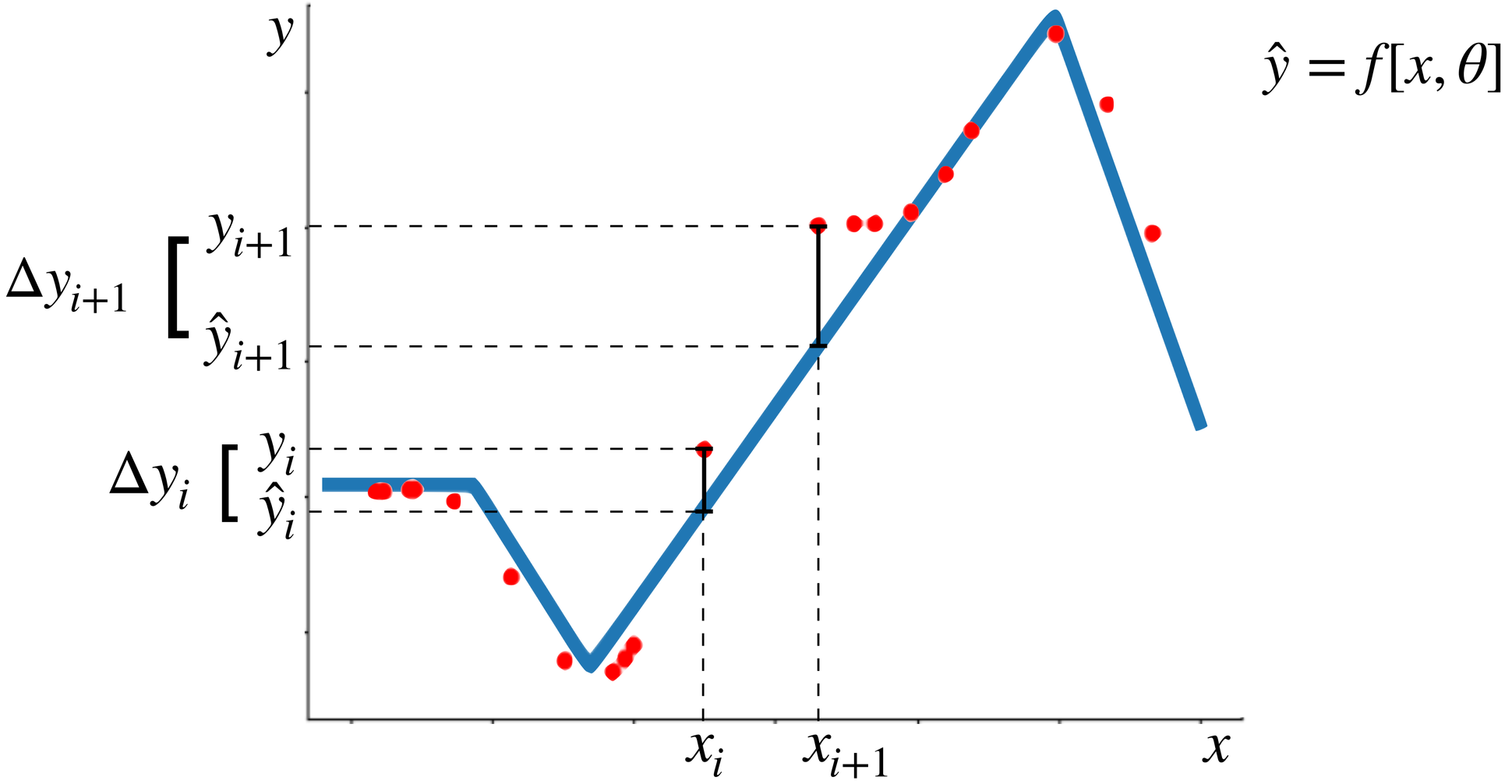


# Redes Neuronales: Ajuste



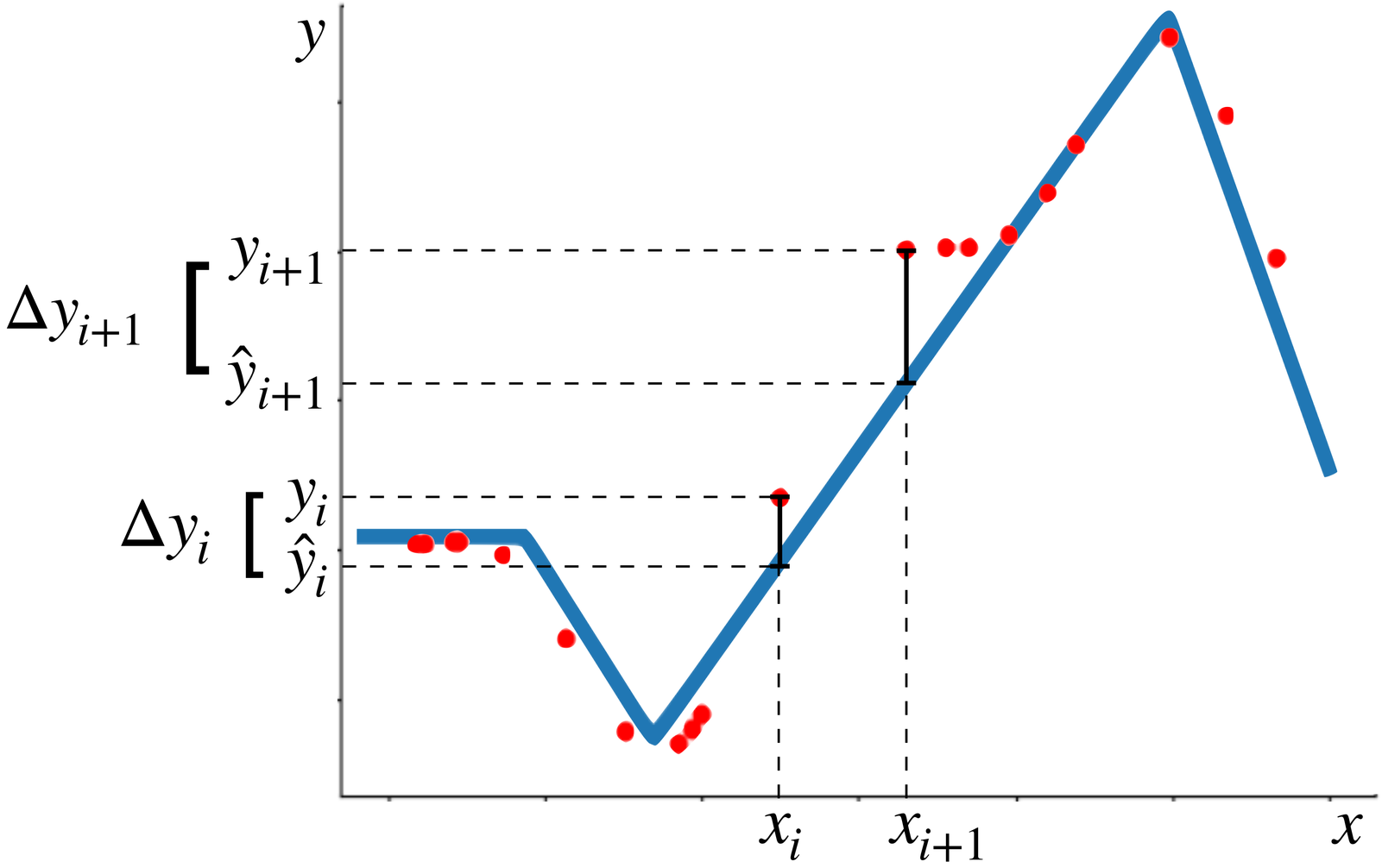

$$\hat{y} = f[x, \theta]$$

# Redes Neuronales: Ajuste




$$\hat{y} = f[x, \theta]$$

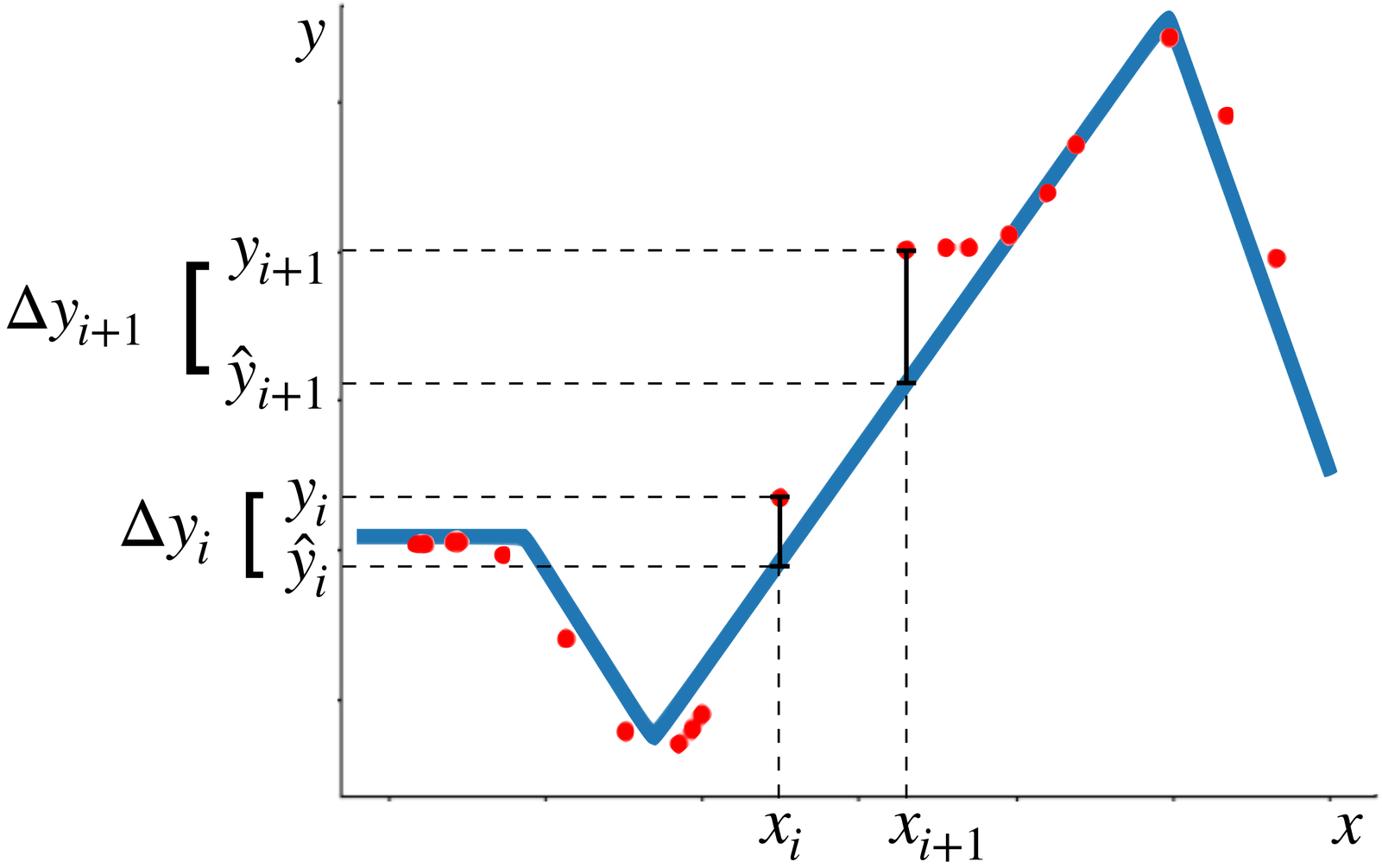
# Redes Neuronales: Ajuste




$$\hat{y} = f[x, \theta]$$

La mejor función $f[x, \theta]$ será la que minimice los $\Delta y_i$.

# Redes Neuronales: Ajuste



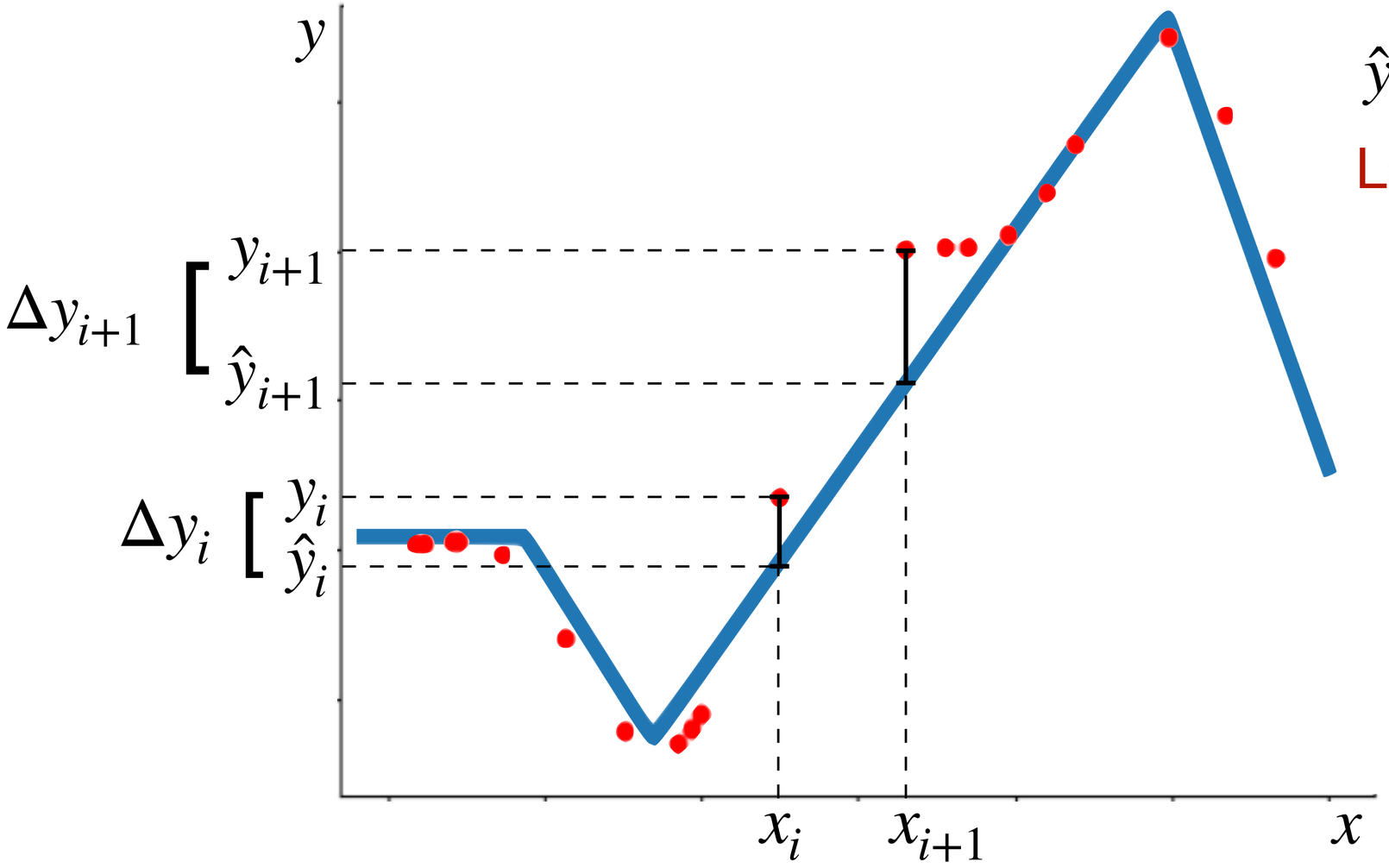

$$\hat{y} = f[x, \theta]$$

La mejor función $f[x, \theta]$ será la que minimice los $\Delta y_i$.

Pero qué combinación de $\Delta y_i$ minimizamos?

La función de pérdida

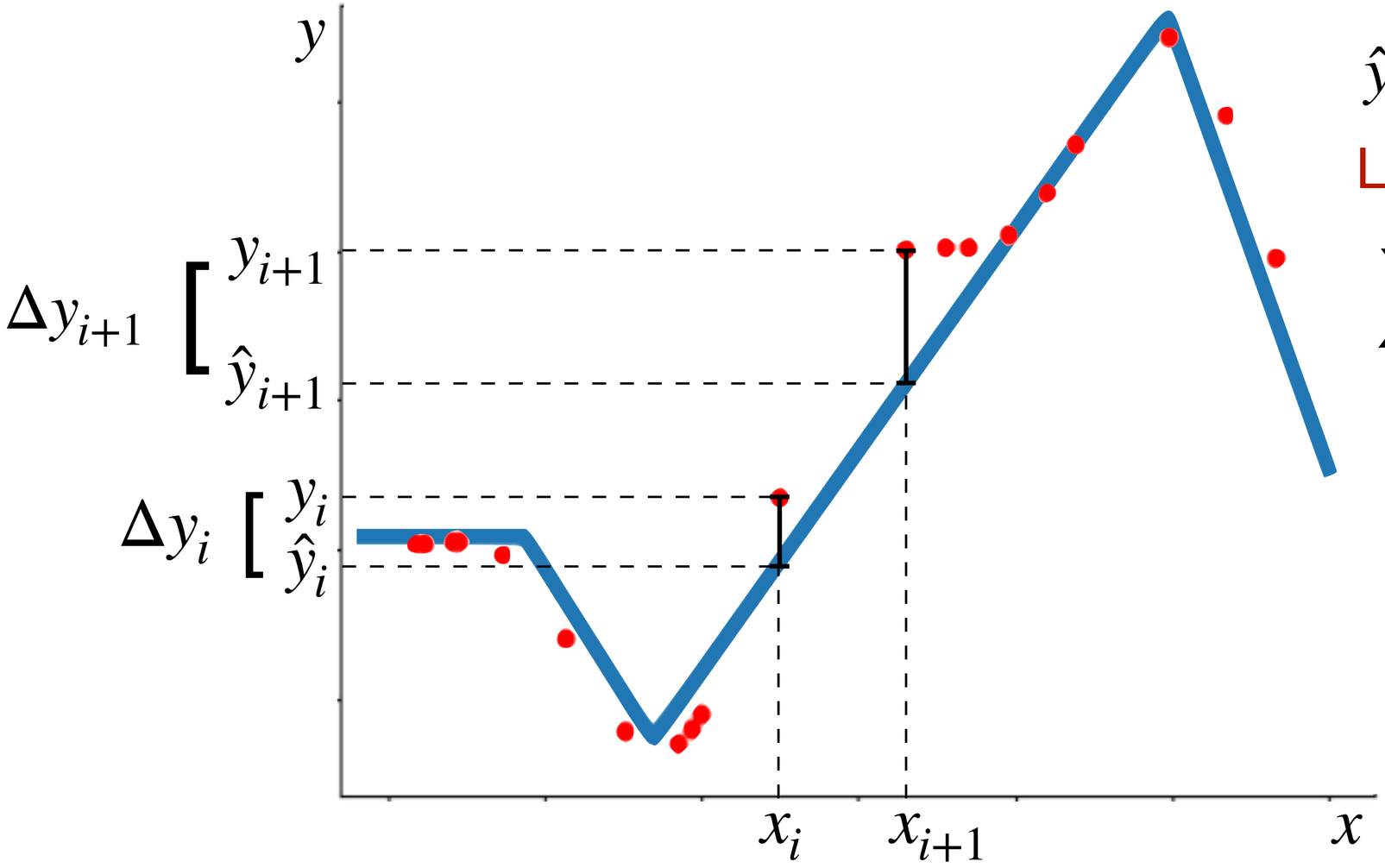
# Redes Neuronales: Función Pérdida




$$\hat{y} = f[x, \theta]$$

La función de pérdida $\mathscr{L}$

# Redes Neuronales: Función Pérdida

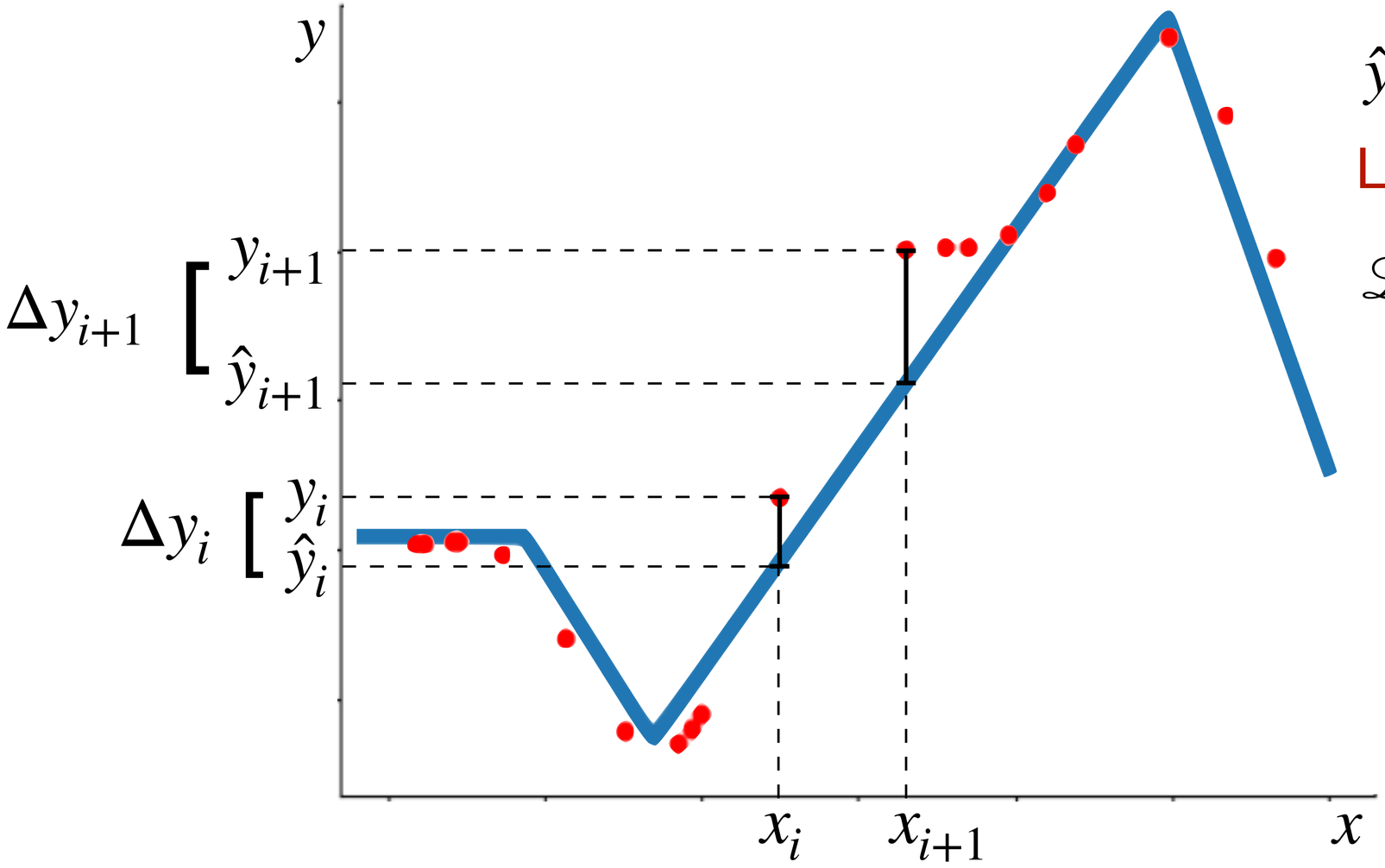



$$\hat{y} = f[x, \theta]$$

La función de pérdida $\mathscr{L}$

$$\sum_i (y_i - \hat{y}_i)^2$$

# Redes Neuronales: Función Pérdida

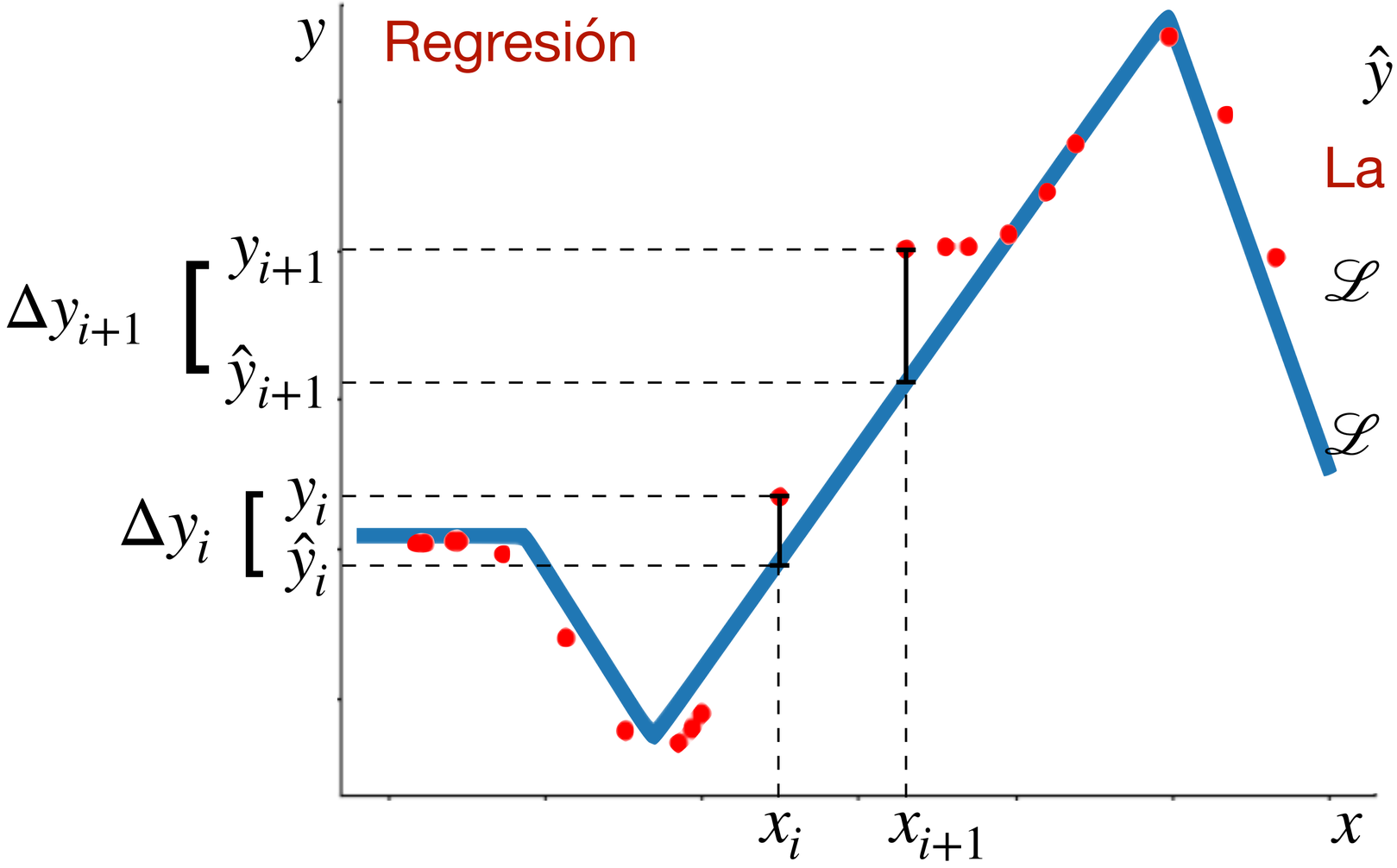



$$\hat{y} = f[x, \theta]$$

La función de pérdida $\mathscr{L}$

$$\mathscr{L} = \sum_i (y_i - \hat{y}_i)^2 = MSE$$

# Redes Neuronales: Función Pérdida



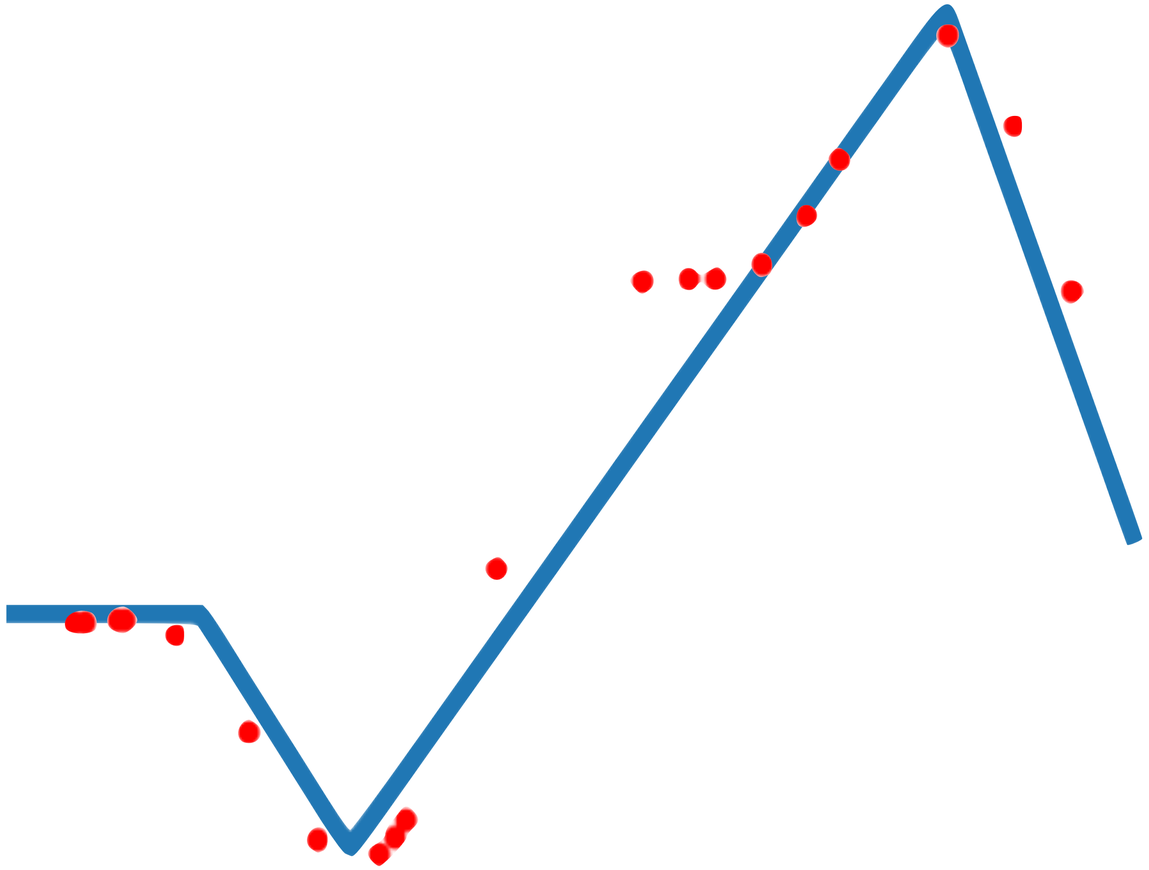

$$\hat{y} = f[x, \theta]$$

La función de pérdida $\mathscr{L}$

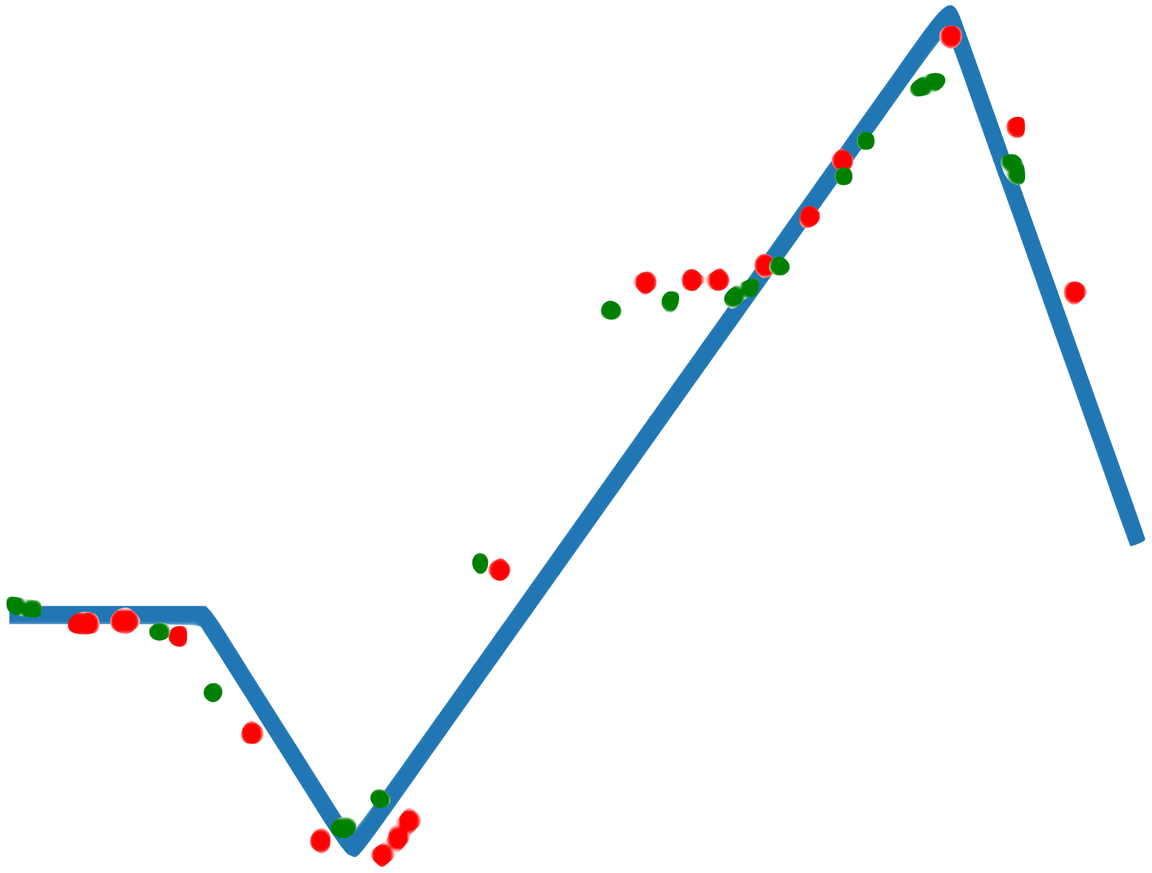
$$\mathscr{L} = \frac{1}{n}\sum_i (y_i - \hat{y}_i)^2 = MSE$$

$$\mathscr{L} = \frac{1}{n}\sum_i |y_i - \hat{y}_i| = MAE$$

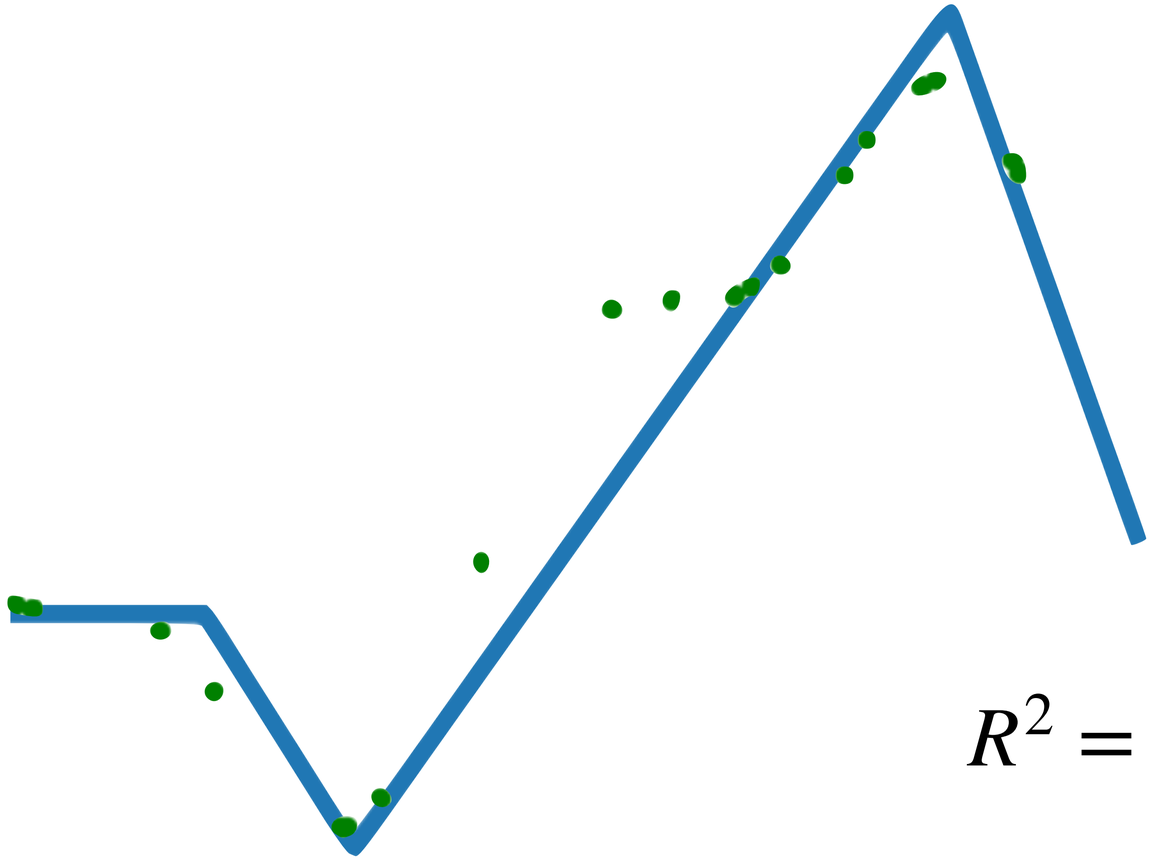
# Redes Neuronales: Métricas de Evaluación



# Redes Neuronales: Métricas de Evaluación

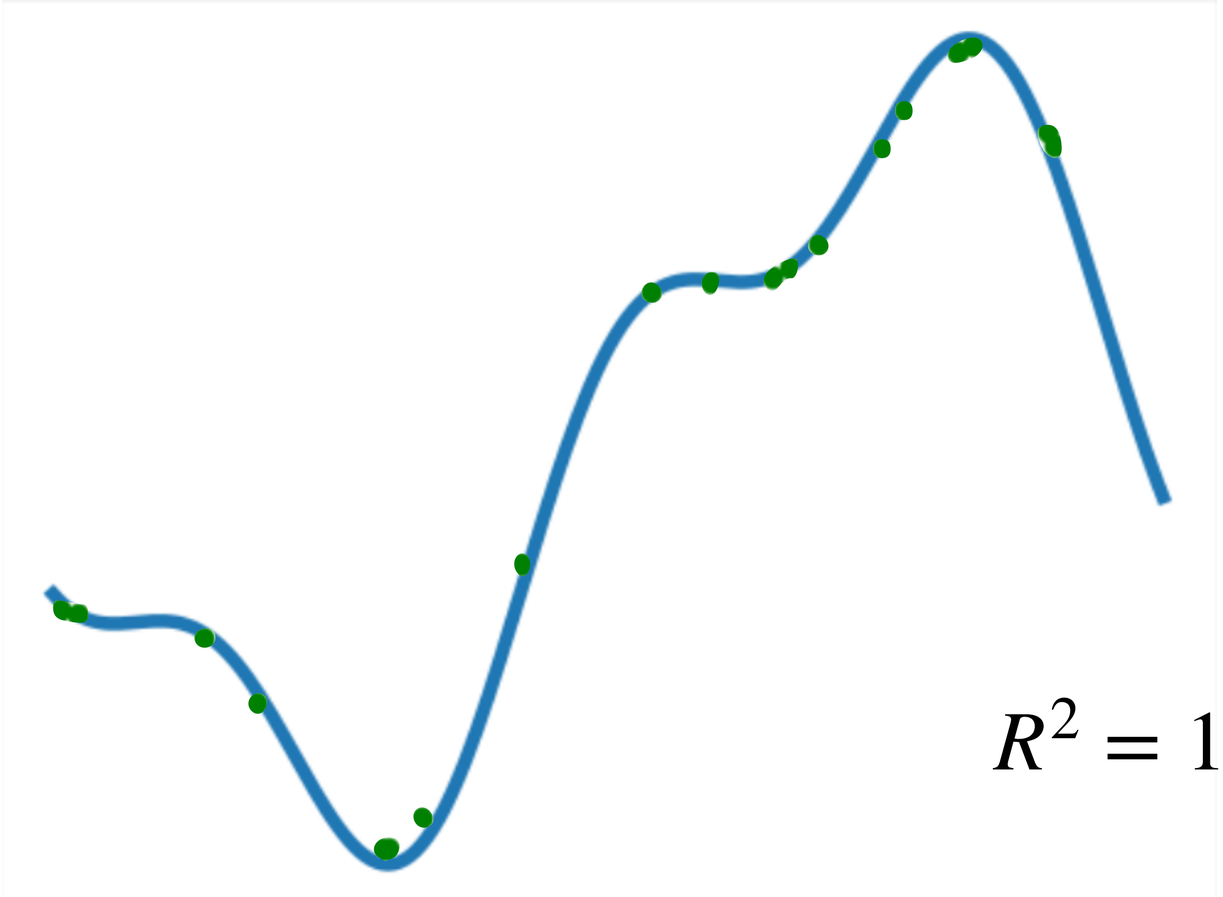


# Redes Neuronales: Métricas de Evaluación



$$R^2 = 1 - \frac{\text{Error del modelo}}{\text{Variabilidad de los datos}}$$

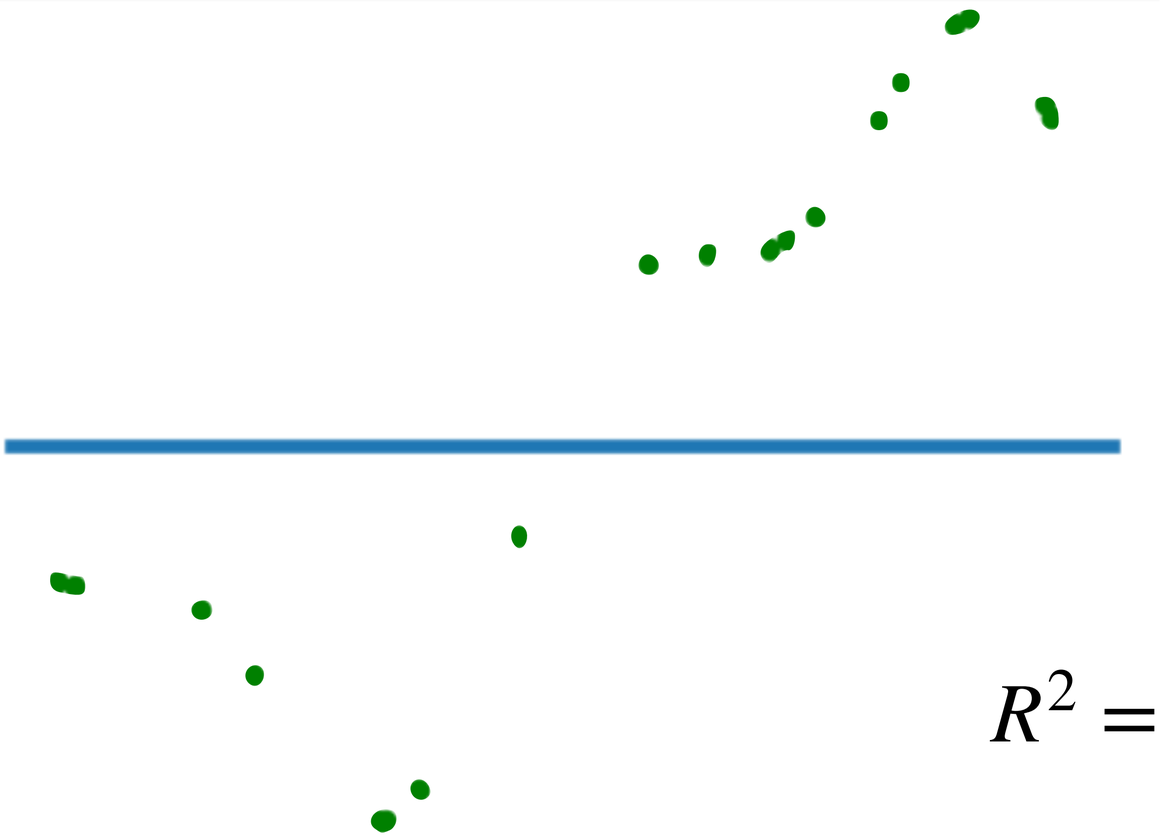
# Redes Neuronales: Métricas de Evaluación



$$R^2 = 1$$

$$R^2 = 1 - \frac{\text{Error del modelo}}{\text{Variabilidad de los datos}}$$

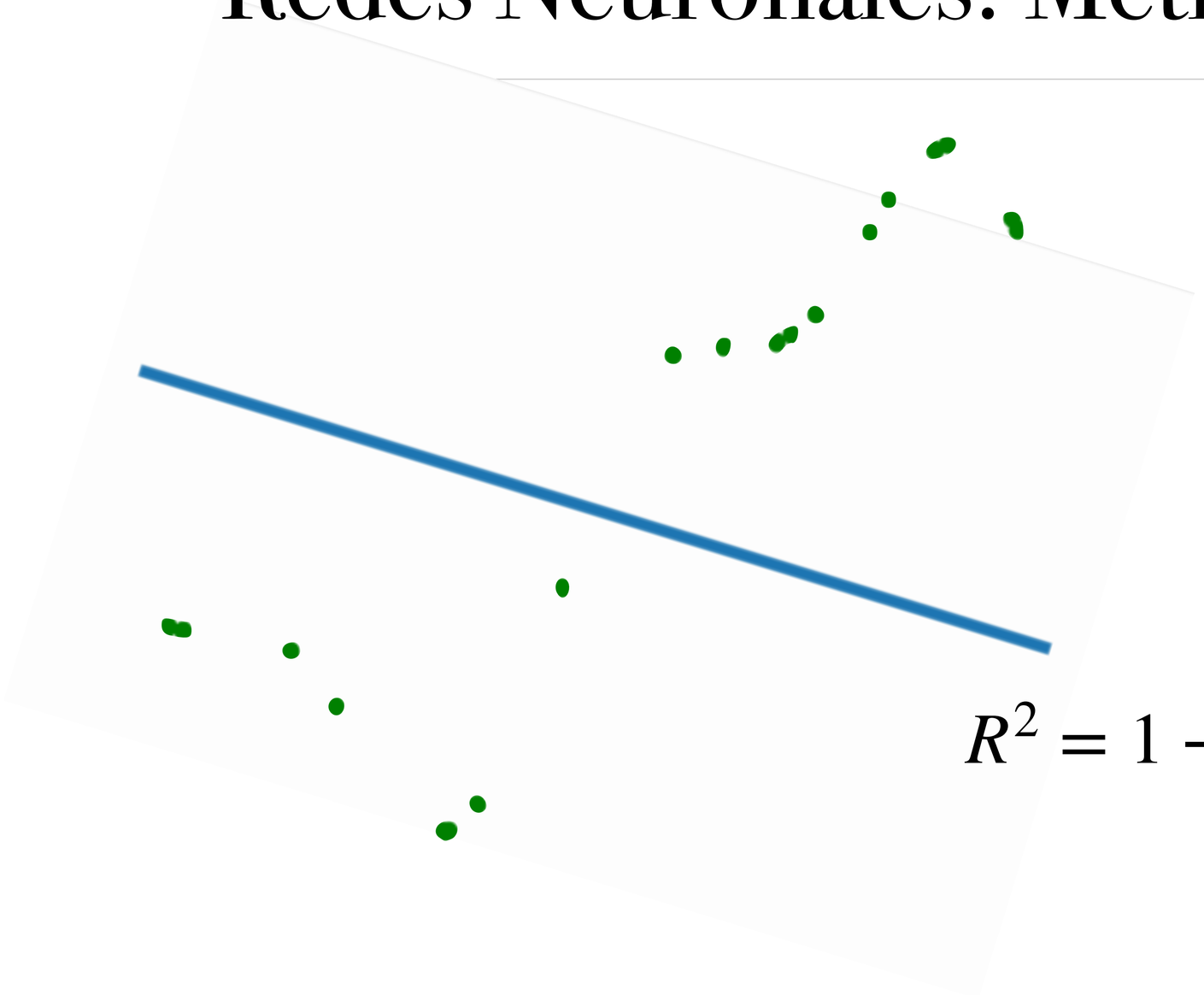
# Redes Neuronales: Métricas de Evaluación



$$R^2 = 0$$

$$R^2 = 1 - \frac{\text{Error del modelo}}{\text{Variabilidad de los datos}}$$

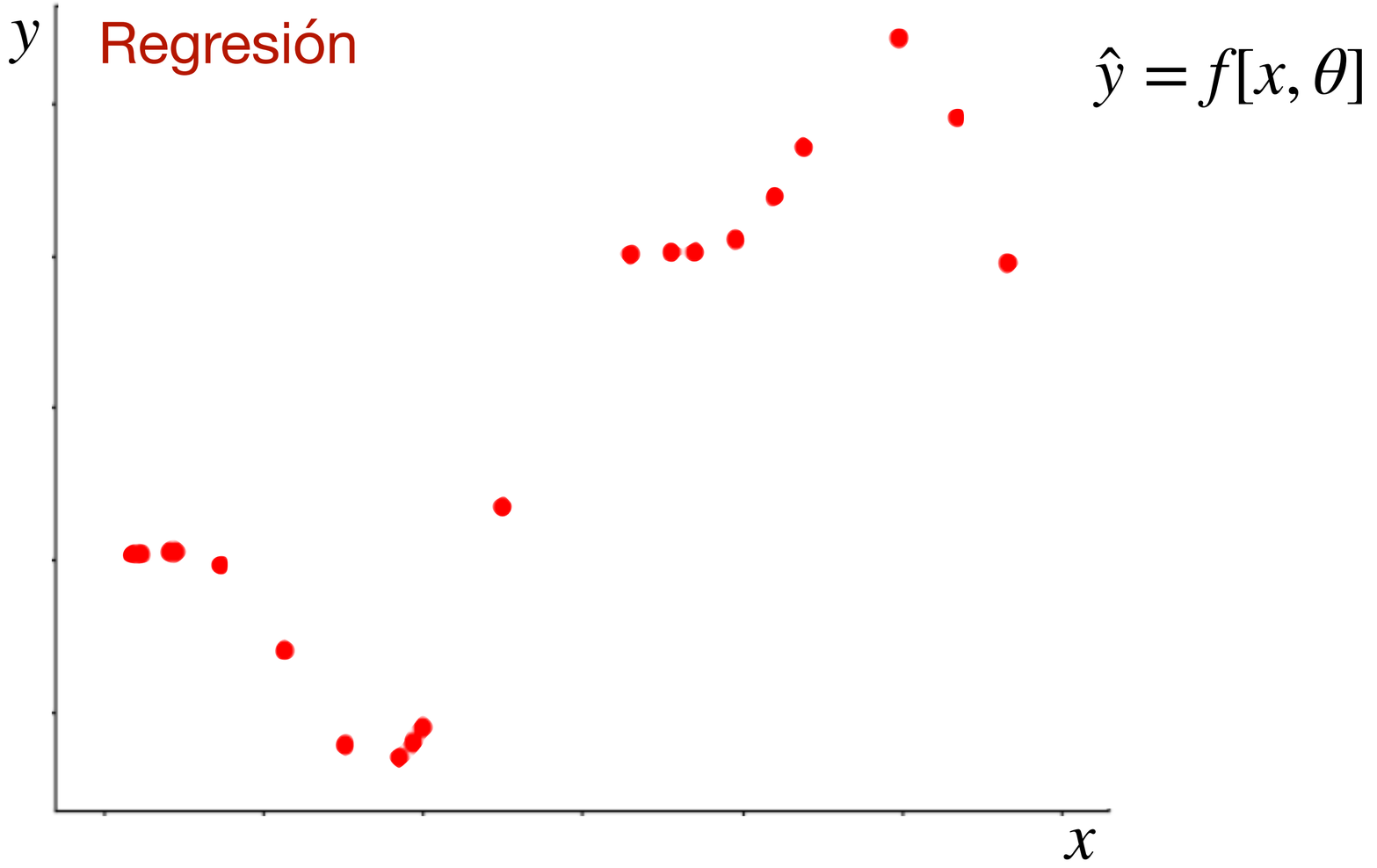
# Redes Neuronales: Métricas de Evaluación



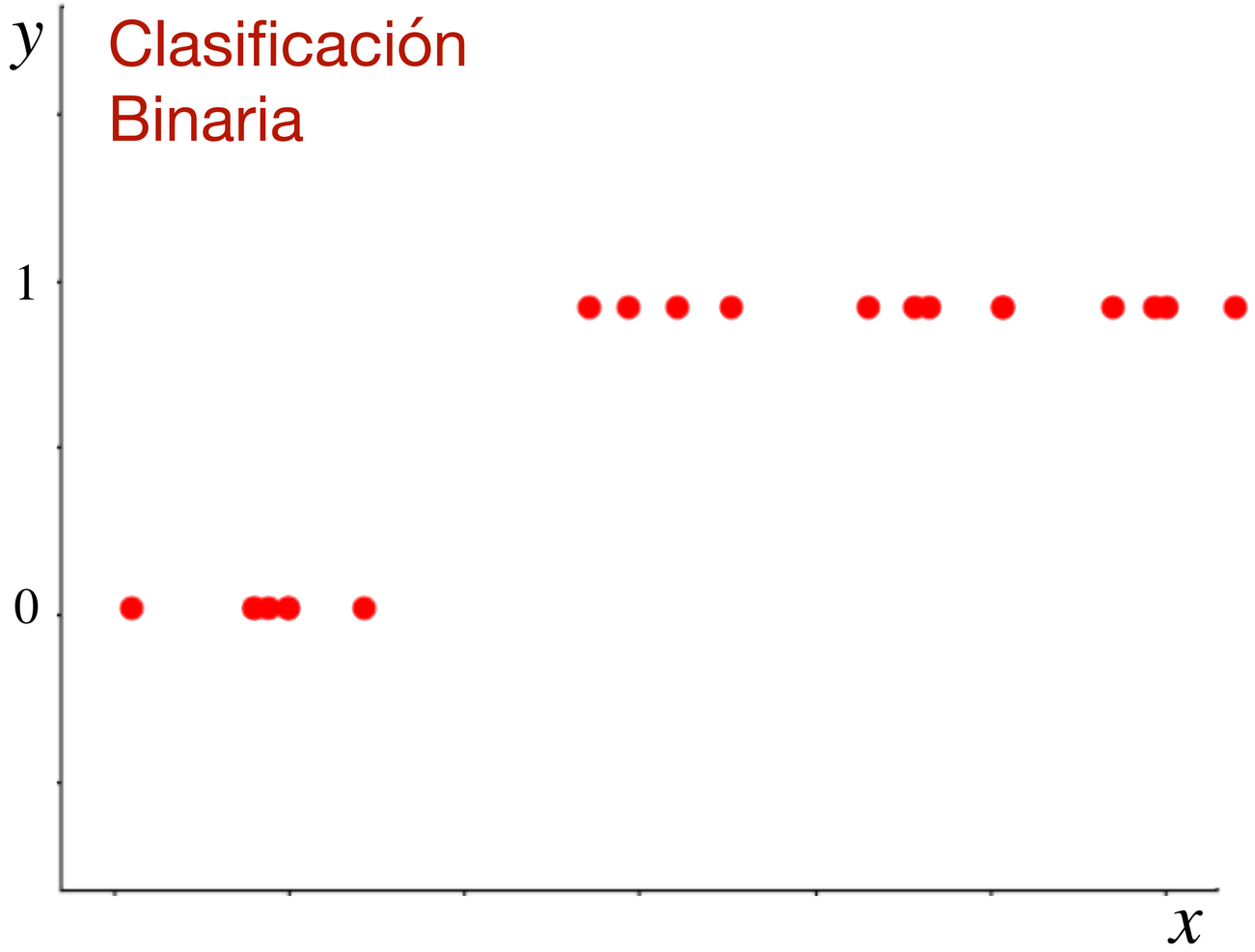
$$R^2 < 0$$

$$R^2 = 1 - \frac{\text{Error del modelo}}{\text{Variabilidad de los datos}}$$

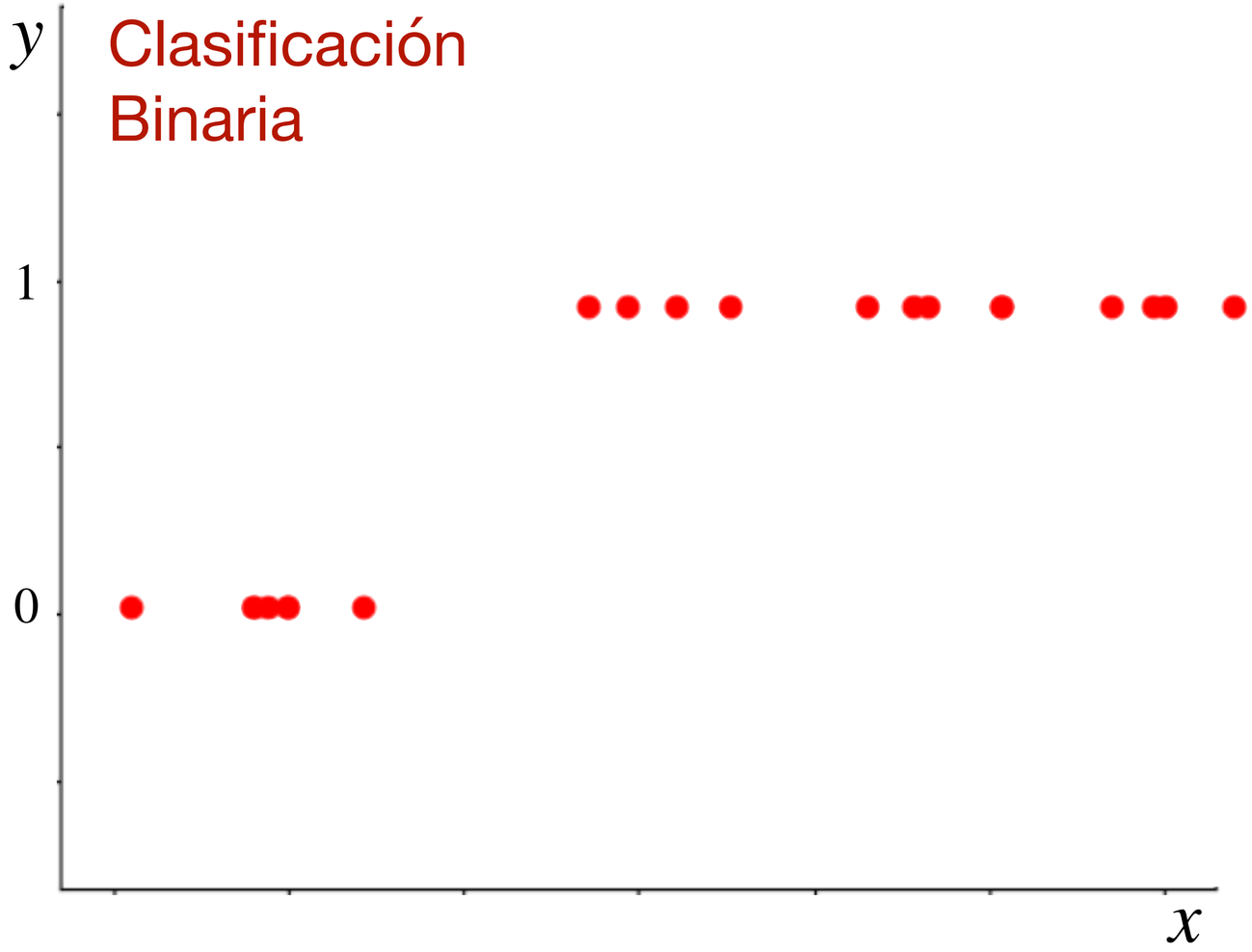
# Redes Neuronales: Función Pérdida



# Redes Neuronales: Función Pérdida

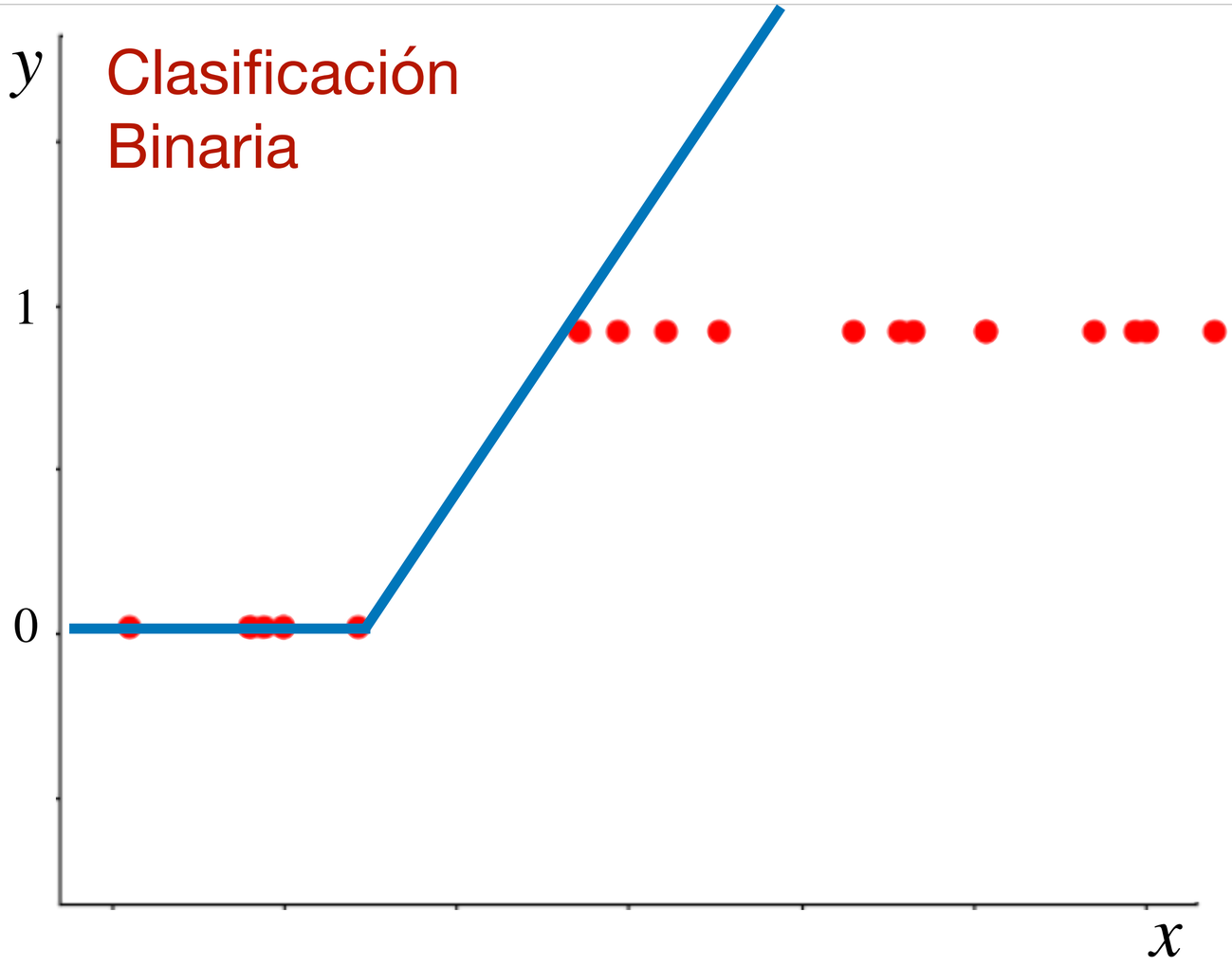


# Redes Neuronales: Función Pérdida



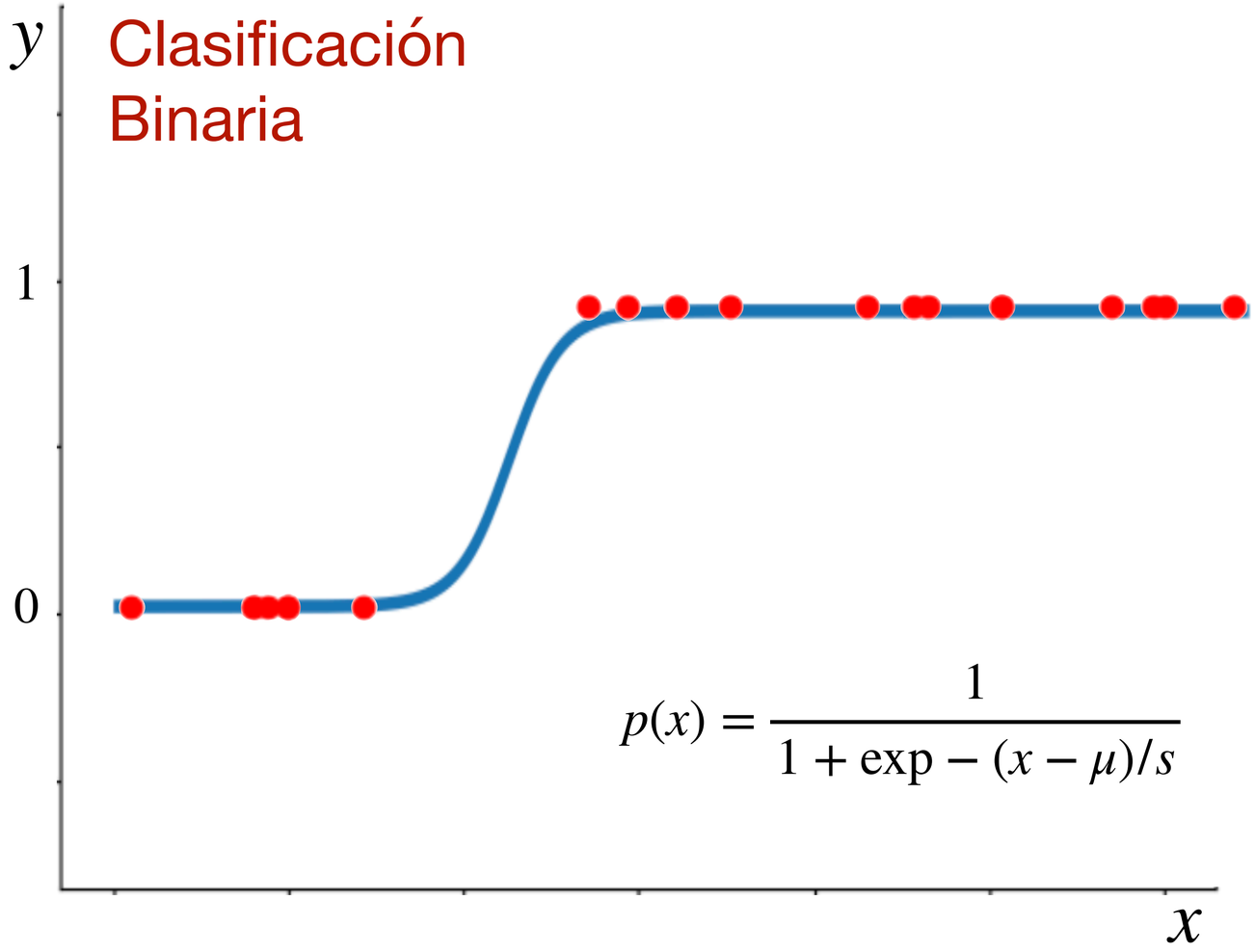

$$\hat{y} = \begin{cases} 1 \\ 0 \end{cases}$$

# Redes Neuronales: Función Pérdida




$$\hat{y} = \begin{cases} 1 \\ 0 \end{cases}$$

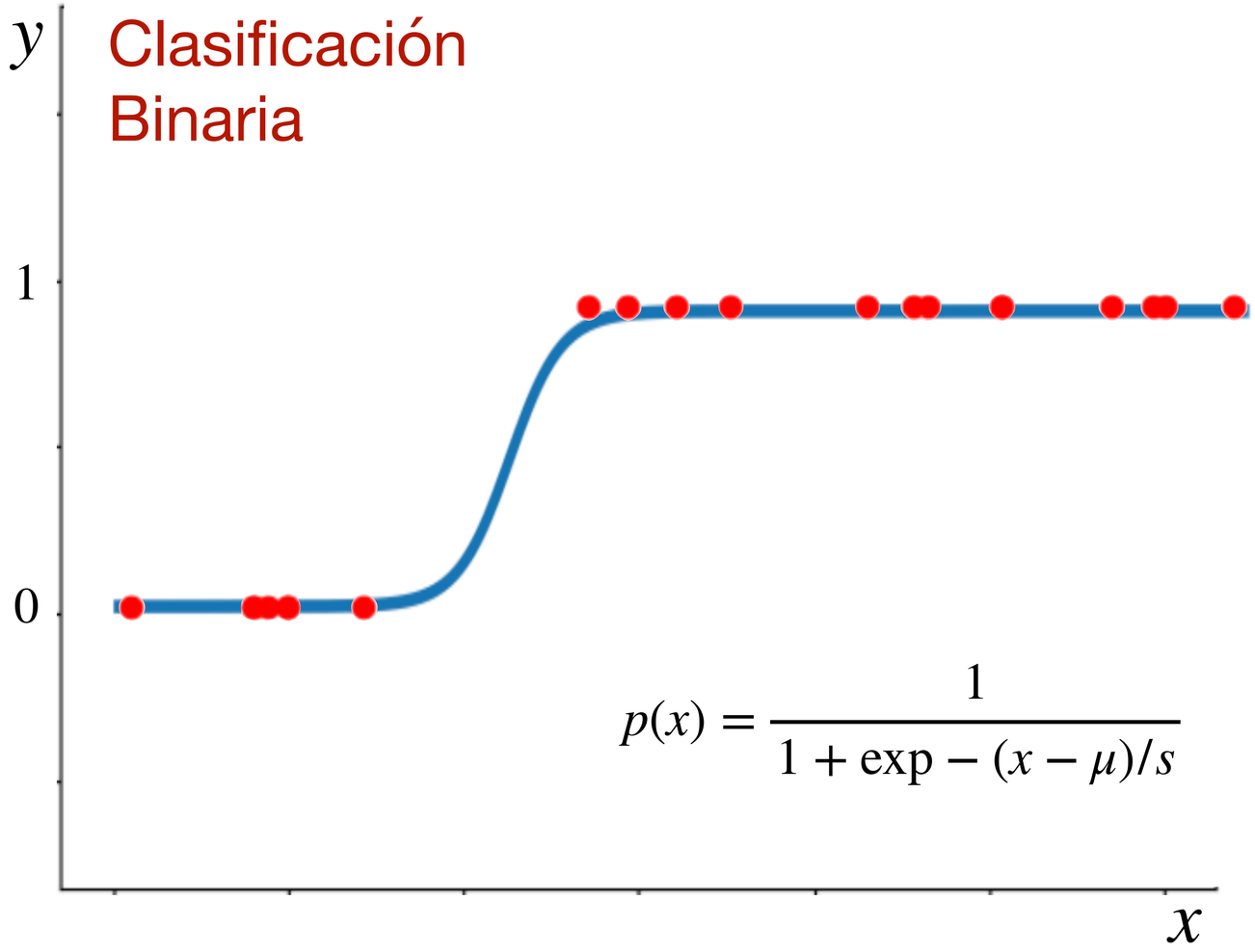
# Redes Neuronales: Función Pérdida




$$p(x) = \frac{1}{1 + \exp - (x - \mu)/s}$$

$$\hat{y} = \begin{cases} 1 \\ 0 \end{cases}$$

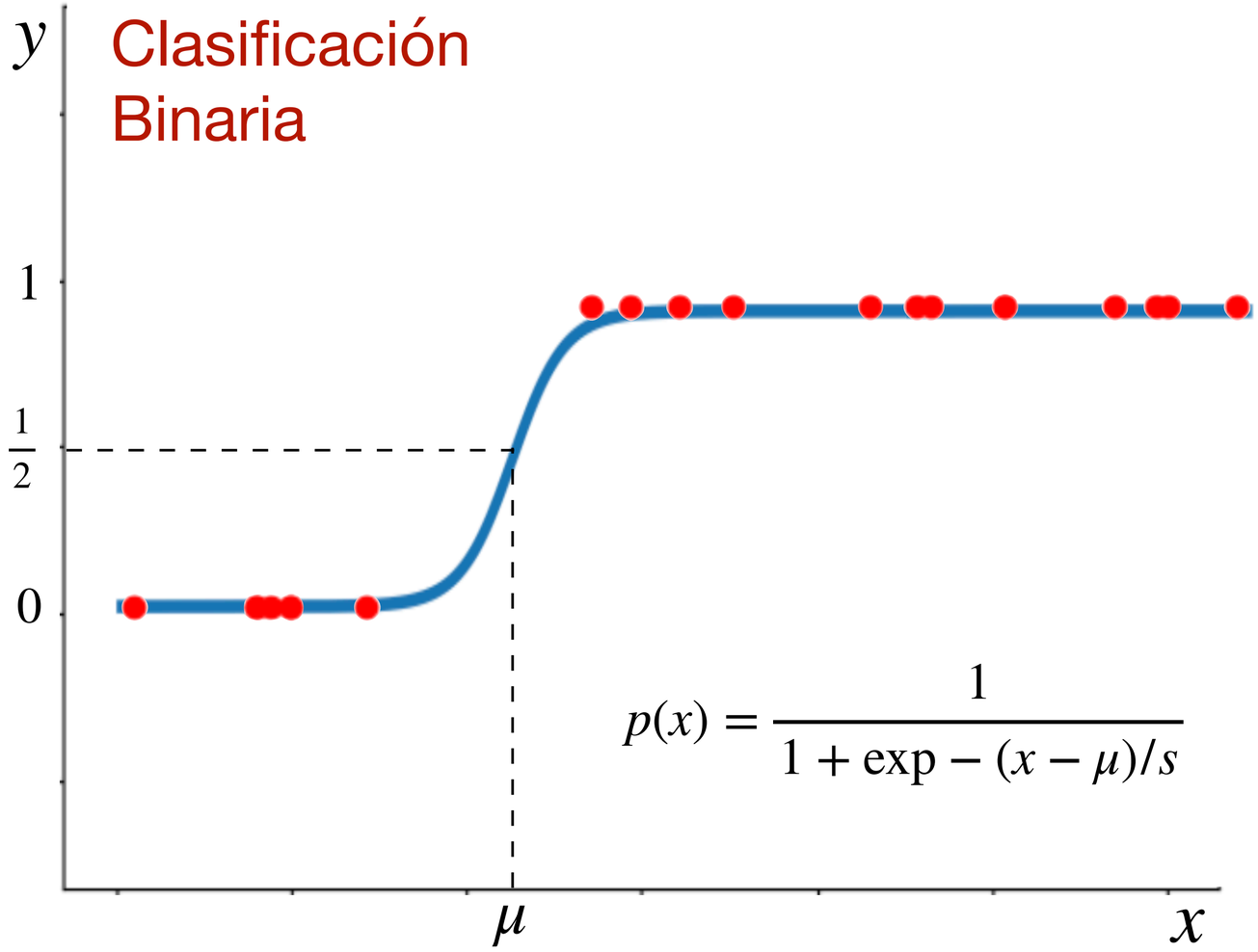
# Redes Neuronales: Función Pérdida




$$p(x) = \frac{1}{1 + \exp - (x - \mu)/s}$$

$$\hat{y} = \begin{cases} 1 & p(x) \geq 0.5 \\ 0 & p(x) < 0.5 \end{cases}$$

# Redes Neuronales: Función Pérdida




$$p(x) = \frac{1}{1 + \exp - (x - \mu)/s}$$

$$\hat{y} = \begin{cases} 1 & p(x) \geq 0.5 \\ 0 & p(x) < 0.5 \end{cases}$$

# Redes Neuronales: Función Pérdida

$$P(y = 1 \mid x) = \frac{1}{1 + \exp^{-(x-\mu)/s}} = \frac{1}{1 + \exp^{-\Theta^T \vec{x}}} = \sigma(\Theta^T \vec{x})$$

# Redes Neuronales: Función Pérdida

$$P(y = 1 \,|\, x) = \frac{1}{1 + \exp^{-(x-\mu)/s}} = \frac{1}{1 + \exp^{-\Theta^T \vec{x}}} = \sigma(\Theta^T \vec{x})$$

$$P(y = 0 \,|\, x) = 1 - P(y = 1 \,|\, x)$$

# Redes Neuronales: Función Pérdida

$$P(y=1\,|\,x) = \frac{1}{1+\exp^{-(x-\mu)/s}} = \frac{1}{1+\exp^{-\Theta^T\vec{x}}} = \sigma(\Theta^T\vec{x})$$

$$P(y=0\,|\,x) = 1 - P(y=1\,|\,x)$$

$$P(y\,|\,x) = P(y=1\,|\,x)^y\, P(y=0\,|\,x)^{(1-y)} = P(y=1\,|\,x)^y\,(1-P(y=1\,|\,x))^{(1-y)}$$

Bernoulli Distribution

# Redes Neuronales: Función Pérdida

$$P(y=1\,|\,x) = \frac{1}{1+\exp^{-(x-\mu)/s}} = \frac{1}{1+\exp^{-\Theta^T\vec{x}}} = \sigma(\Theta^T\vec{x})$$

$$P(y=0\,|\,x) = 1 - P(y=1\,|\,x)$$

$$P(y\,|\,x) = P(y=1\,|\,x)^y\, P(y=0\,|\,x)^{(1-y)} = P(y=1\,|\,x)^y\,(1-P(y=1\,|\,x))^{(1-y)}$$

Bernoulli Distribution

Observaciones Independientes

$$L(\Theta) = P(Y\,|\,X,\Theta) = \Pi_i\, P(y_i=1\,|\,x_i)_i^y\,(1-P(y_i=1\,|\,x_i))^{(1-y_i)}$$

# Redes Neuronales: Función Pérdida

$$P(y=1\,|\,x) = \frac{1}{1+\exp^{-(x-\mu)/s}} = \frac{1}{1+\exp^{-\Theta^T\vec{x}}} = \sigma(\Theta^T\vec{x})$$

$$P(y=0\,|\,x) = 1 - P(y=1\,|\,x)$$

$$P(y\,|\,x) = P(y=1\,|\,x)^y\, P(y=0\,|\,x)^{(1-y)} = P(y=1\,|\,x)^y\,(1-P(y=1\,|\,x))^{(1-y)}$$ Bernoulli Distribution

Observaciones Independientes

$$L(\Theta) = P(Y|X,\Theta) = \Pi_i\, P(y_i=1\,|\,x_i)_i^y\,(1-P(y_i=1\,|\,x_i))^{(1-y_i)} = \Pi_i\,\sigma(x_i)^{y_i}\,(1-\sigma(x_i))^{(1-y_i)}$$ Likelihood

# Redes Neuronales: Función Pérdida

$$P(y=1\,|\,x) = \frac{1}{1+\exp^{-(x-\mu)/s}} = \frac{1}{1+\exp^{-\Theta^T\vec{x}}} = \sigma(\Theta^T\vec{x})$$

$$P(y=0\,|\,x) = 1 - P(y=1\,|\,x)$$

$$P(y\,|\,x) = P(y=1\,|\,x)^y\, P(y=0\,|\,x)^{(1-y)} = P(y=1\,|\,x)^y\,(1-P(y=1\,|\,x))^{(1-y)}$$ Bernoulli Distribution

Observaciones Independientes

$$L(\Theta) = P(Y|X,\Theta) = \Pi_i\, P(y_i=1\,|\,x_i)^y_i\,(1-P(y_i=1\,|\,x_i))^{(1-y_i)} \quad = \Pi_i\,\sigma(x_i)^{y_i}\,(1-\sigma(x_i))^{(1-y_i)}$$ Likelihood

$$\log(L(\Theta)) = \log(P(Y|X,\Theta) = \log(\sum_i P(y_i=1\,|\,x_i)^y_i\,(1-P(y_i=1\,|\,x_i))^{(1-y_i)})$$

# Redes Neuronales: Función Pérdida

$$P(y=1\,|\,x) = \frac{1}{1+\exp^{-(x-\mu)/s}} = \frac{1}{1+\exp^{-\Theta^T\vec{x}}} = \sigma(\Theta^T\vec{x})$$

$$P(y=0\,|\,x) = 1 - P(y=1\,|\,x)$$

$$P(y\,|\,x) = P(y=1\,|\,x)^y\, P(y=0\,|\,x)^{(1-y)} = P(y=1\,|\,x)^y\,(1-P(y=1\,|\,x))^{(1-y)}$$ Bernoulli Distribution

Observaciones Independientes

$$L(\Theta) = P(Y|X,\Theta) = \Pi_i\, P(y_i=1\,|\,x_i)^{y_i}\,(1-P(y_i=1\,|\,x_i))^{(1-y_i)} \quad = \Pi_i\,\sigma(x_i)^{y_i}\,(1-\sigma(x_i))^{(1-y_i)}$$ Likelihood

$$\log(L(\Theta)) = \log(P(Y|X,\Theta) = \log(\sum_i P(y_i=1\,|\,x_i)_i^y\,(1-P(y_i=1\,|\,x_i))^{(1-y_i)})$$

$$\log(L(\Theta)) = \sum_i (\log\ \sigma(x_i)^{y_i}\,(1-\sigma(x_i))^{(1-y_i)})$$ Log Likelihood

# Redes Neuronales: Función Pérdida

$$P(y=1\,|\,x) = \frac{1}{1+\exp^{-(x-\mu)/s}} = \frac{1}{1+\exp^{-\Theta^T\vec{x}}} = \sigma(\Theta^T\vec{x})$$

$$P(y=0\,|\,x) = 1 - P(y=1\,|\,x)$$

$$P(y\,|\,x) = P(y=1\,|\,x)^y\, P(y=0\,|\,x)^{(1-y)} = P(y=1\,|\,x)^y\,(1-P(y=1\,|\,x))^{(1-y)}$$ Bernoulli Distribution

Observaciones Independientes

$$L(\Theta) = P(Y|X,\Theta) = \Pi_i\, P(y_i=1\,|\,x_i)^y_i\,(1-P(y_i=1\,|\,x_i))^{(1-y_i)} \quad = \Pi_i\,\sigma(x_i)^{y_i}\,(1-\sigma(x_i))^{(1-y_i)}$$ Likelihood

$$\log(L(\Theta)) = \log(P(Y|X,\Theta) = \log(\sum_i P(y_i=1\,|\,x_i)^y_i\,(1-P(y_i=1\,|\,x_i))^{(1-y_i)})$$

$$\log(L(\Theta)) = \sum_i (\log\ \sigma(x_i)^{y_i}\,(1-\sigma(x_i))^{(1-y_i)})$$ Log Likelihood

$$\mathscr{L}_{BCE} = -\left[\, y\log(\sigma(x)) + (1-y)\log(1-\sigma(x))\,\right]$$ Función de Pérdida

# Redes Neuronales: Métricas de Evaluación

| | **Detección de GW** | **Ruido** |
|---|---|---|
| **Hay GW** | TP | FN |
| **No hay GW** | FP | TN |

# Redes Neuronales: Métricas de Evaluación

| | **Detección de GW** | **Ruido** |
|---|---|---|
| **Hay GW** | TP | FN |
| **No hay GW** | FP | TN |

Precisión =

# Redes Neuronales: Métricas de Evaluación

| | **Detección de GW** | **Ruido** |
|---|---|---|
| **Hay GW** | TP | FN |
| **No hay GW** | FP | TN |

$$\text{Precisión} = \frac{TP}{TP + FP}$$

# Redes Neuronales: Métricas de Evaluación

| | **Detección de GW** | **Ruido** |
|---|---|---|
| **Hay GW** | TP | FN |
| **No hay GW** | FP | TN |

$$\text{Precisión} = \frac{TP}{TP + FP}$$

$$\text{Recall} =$$

# Redes Neuronales: Métricas de Evaluación

| | **Detección de GW** | **Ruido** |
|---|---|---|
| **Hay GW** | TP | FN |
| **No hay GW** | FP | TN |

$$\text{Precisión} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

# Redes Neuronales: Función Pérdida

$$\mathscr{L}_{MSE} = \frac{1}{n}\sum_{i}(y_i - \hat{y}_i)^2$$

$$\mathscr{L}_{MAE} = \frac{1}{n}\sum_{i}|y_i - \hat{y}_i|$$

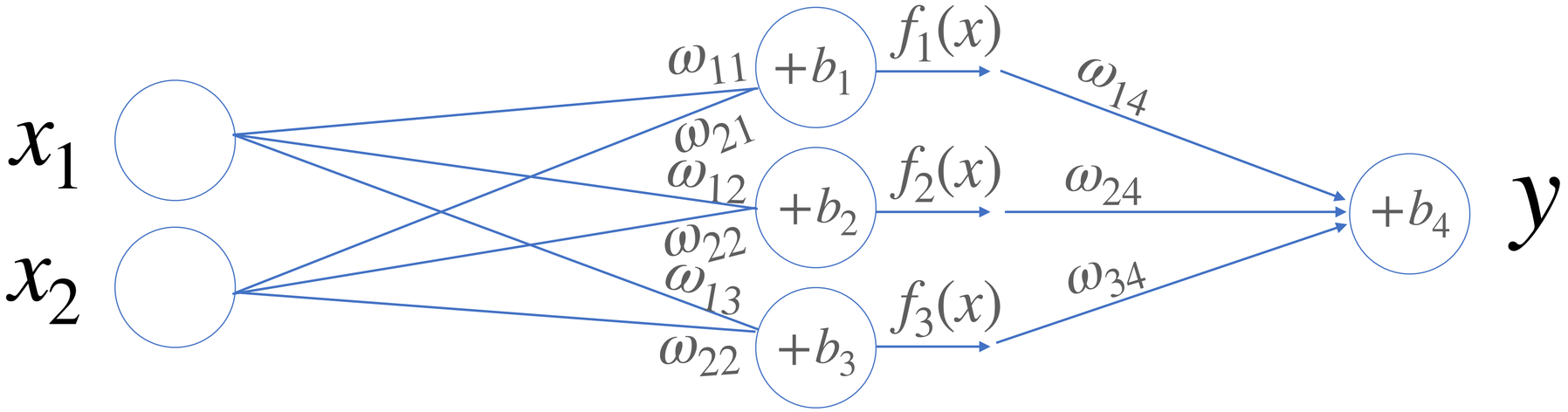
$$\mathscr{L}_{BCE} = -\left[y\log(\hat{y}) + (1-y)\log(1-\sigma(x))\right]$$

$$\mathscr{L}_{CE} = -\sum_{k=1}^{K} y_k \log(\sigma(x_k))$$

• • •

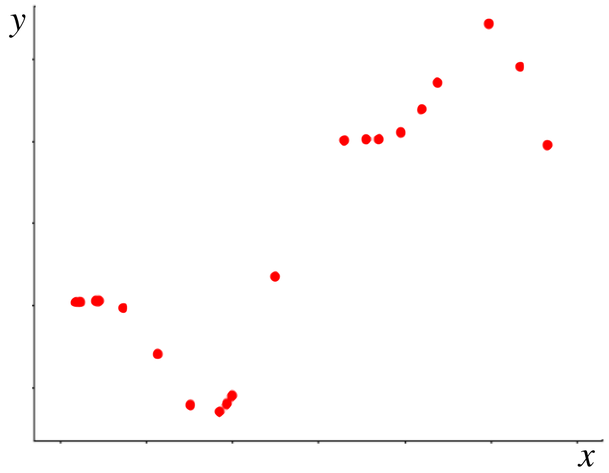
# Redes Neuronales: Shallow

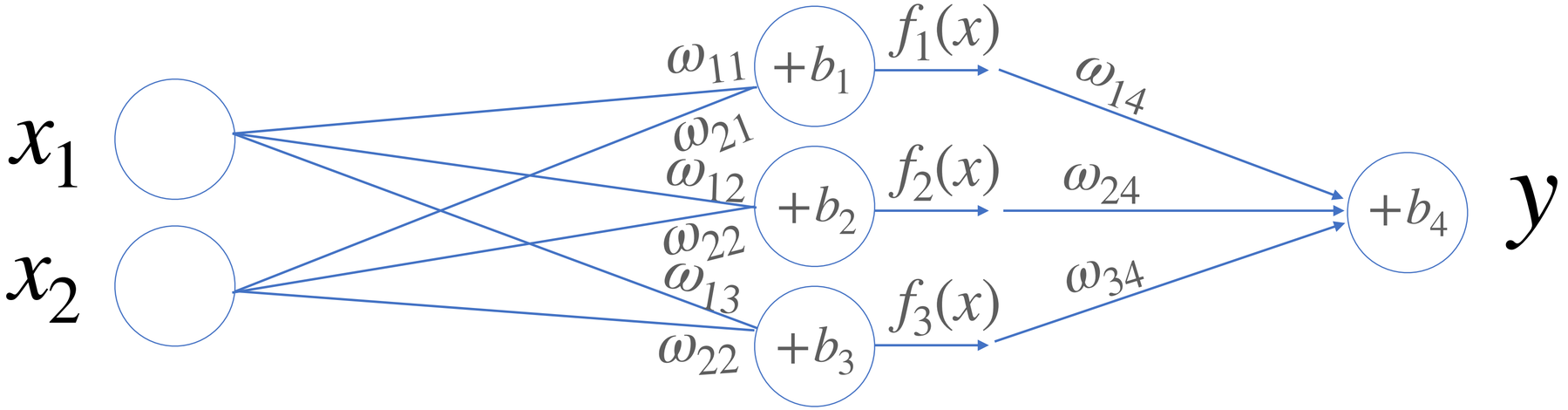
Optimizar los parámetros de la red: minimizar la función de pérdida



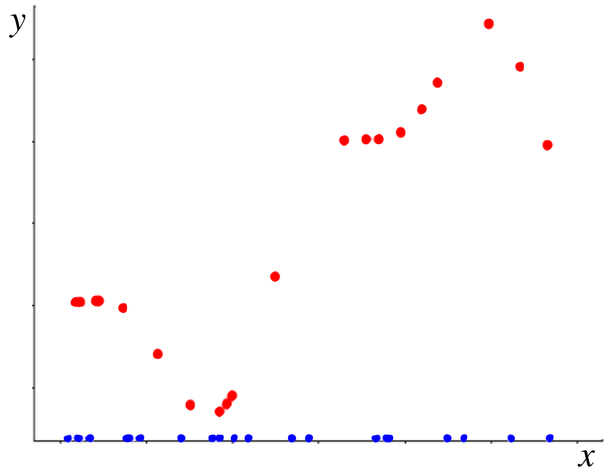

$$\mathscr{L}_{MSE} = \frac{1}{n}\sum_{i}(y_i - \hat{y}_i)^2$$

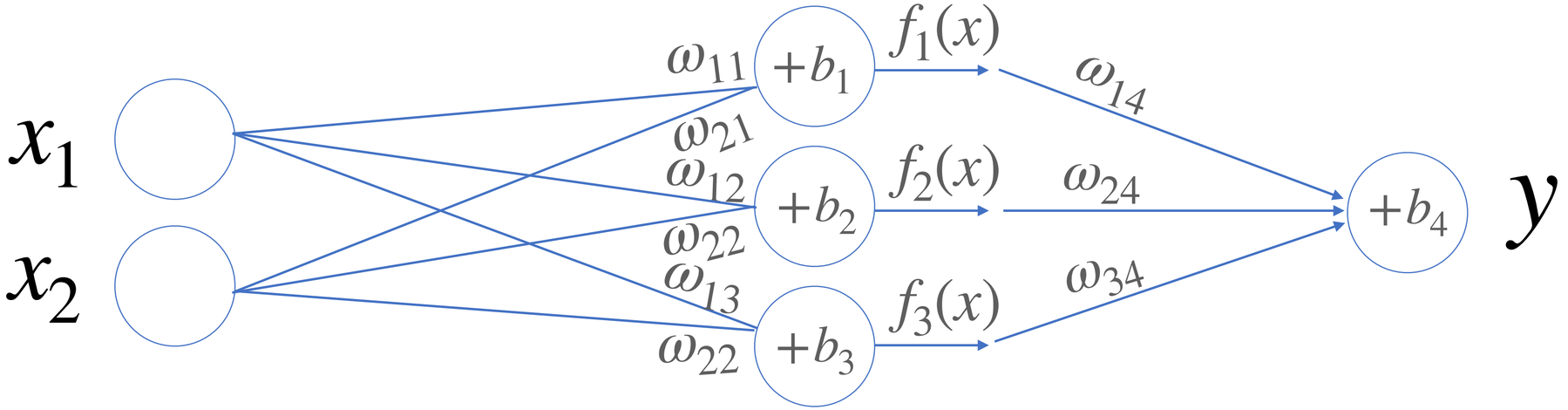
# Redes Neuronales: Función Pérdida



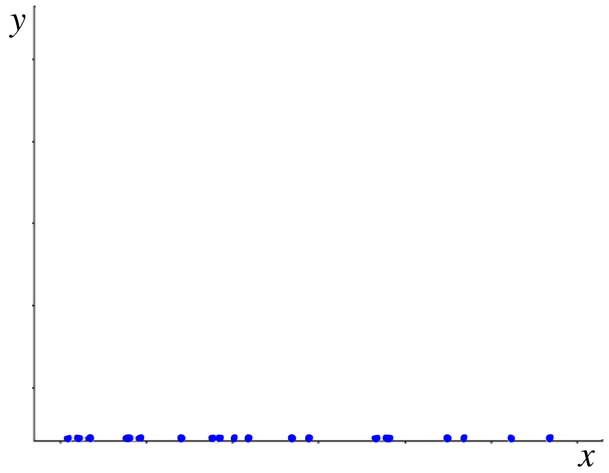



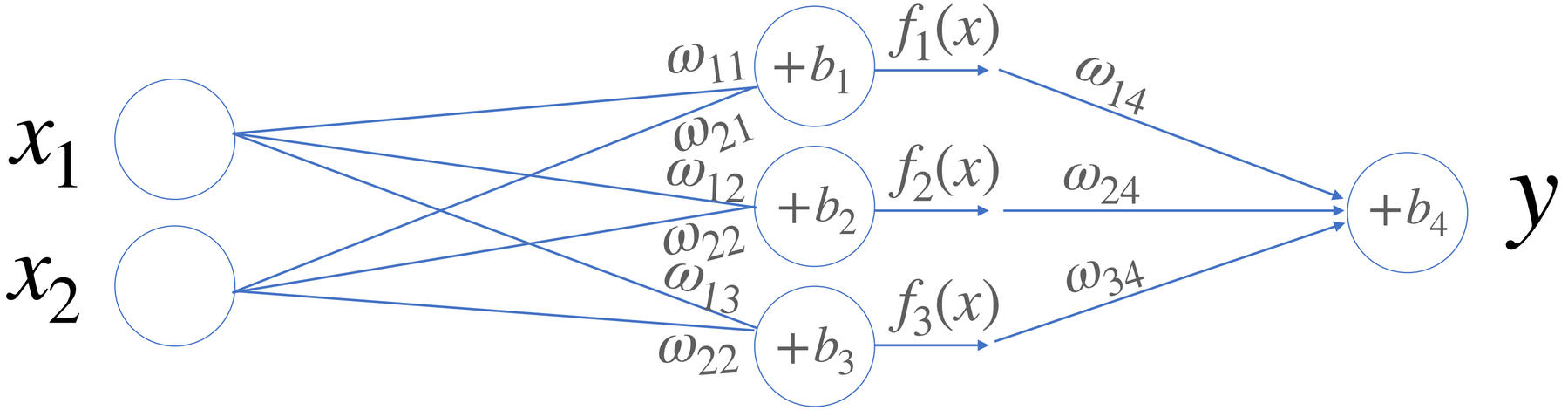
# Redes Neuronales: Función Pérdida



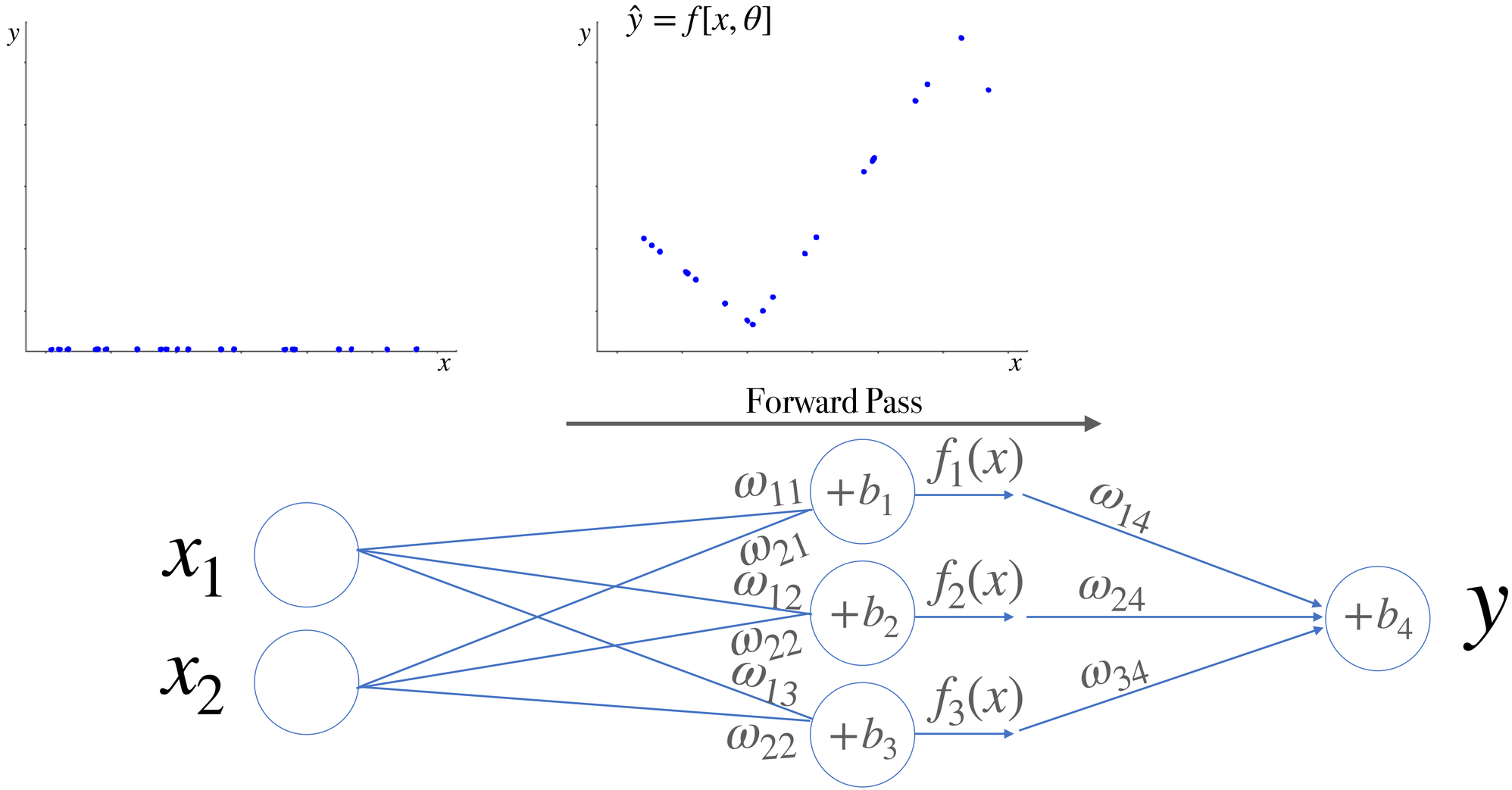



# Redes Neuronales: Función Pérdida

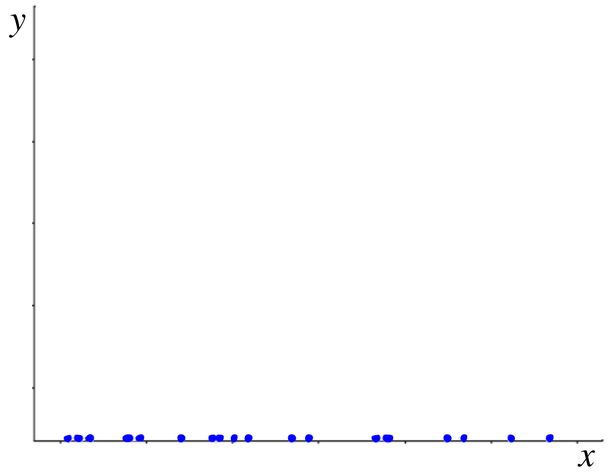


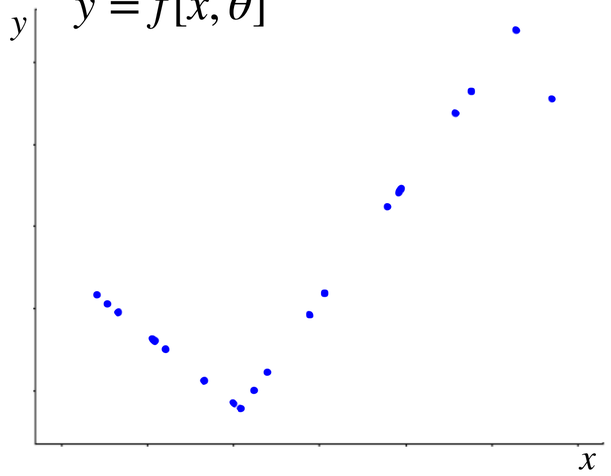



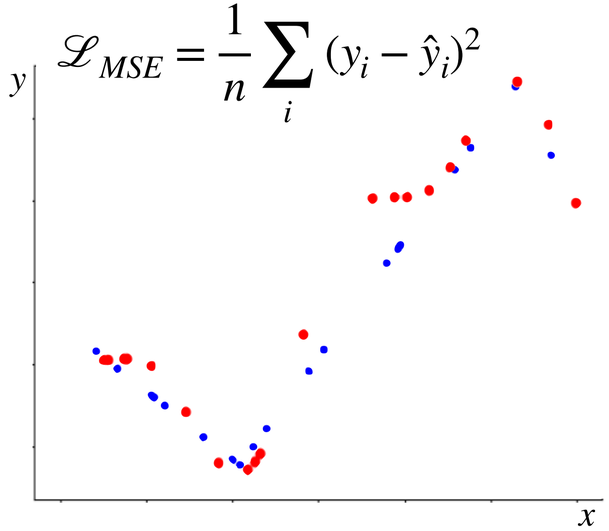
# Redes Neuronales: Función Pérdida

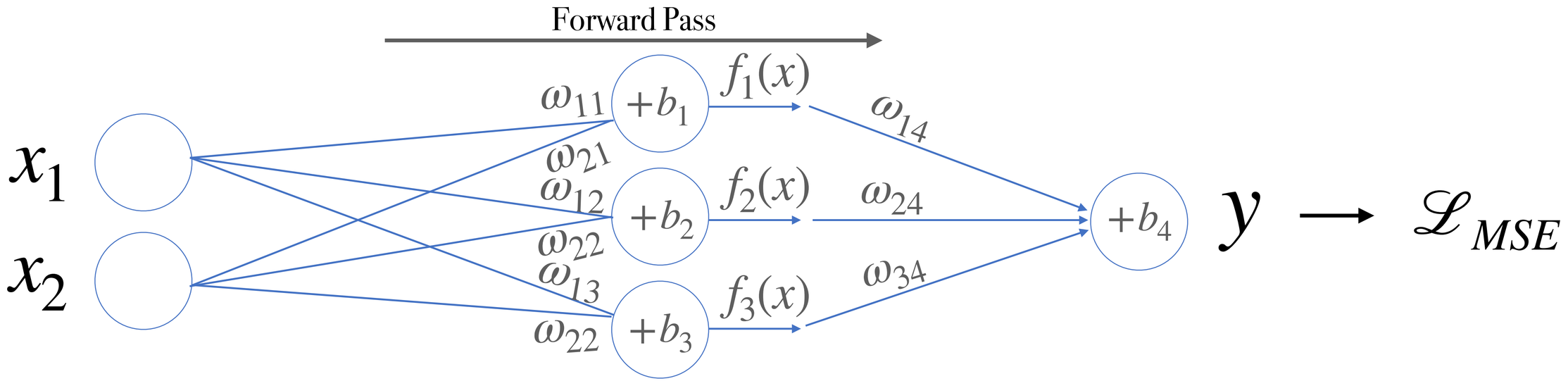


# Redes Neuronales: Función Pérdida



# Redes Neuronales: Bias-Varianza

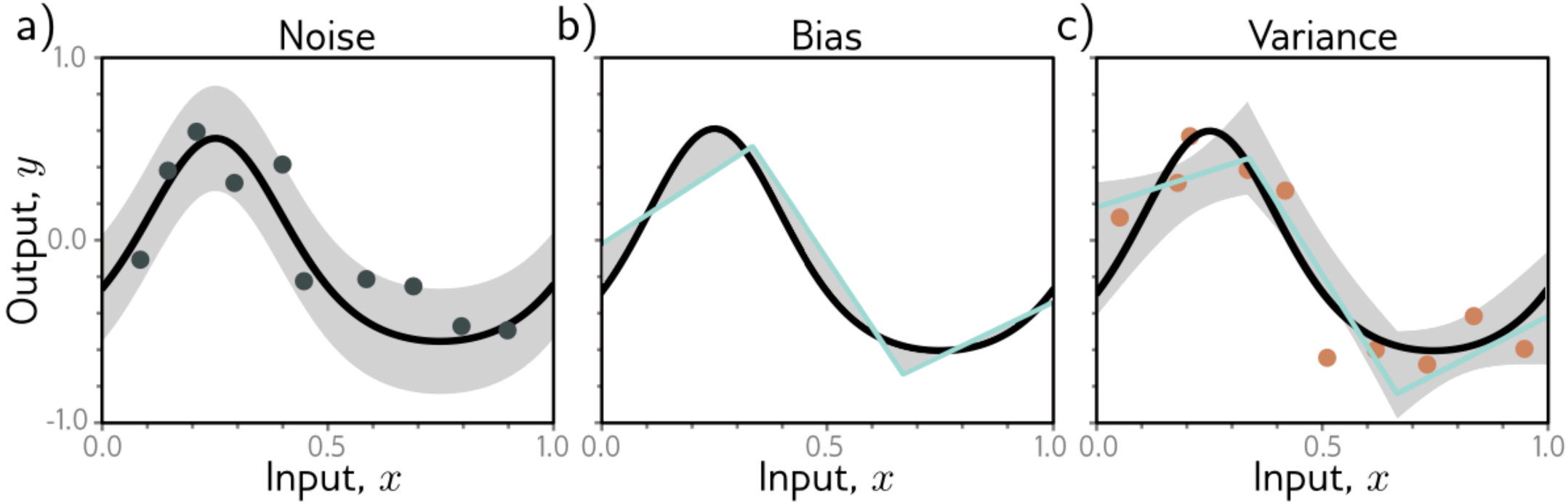



Figura de Understanding Deep Learning

# Redes Neuronales: Reducir la Varianza

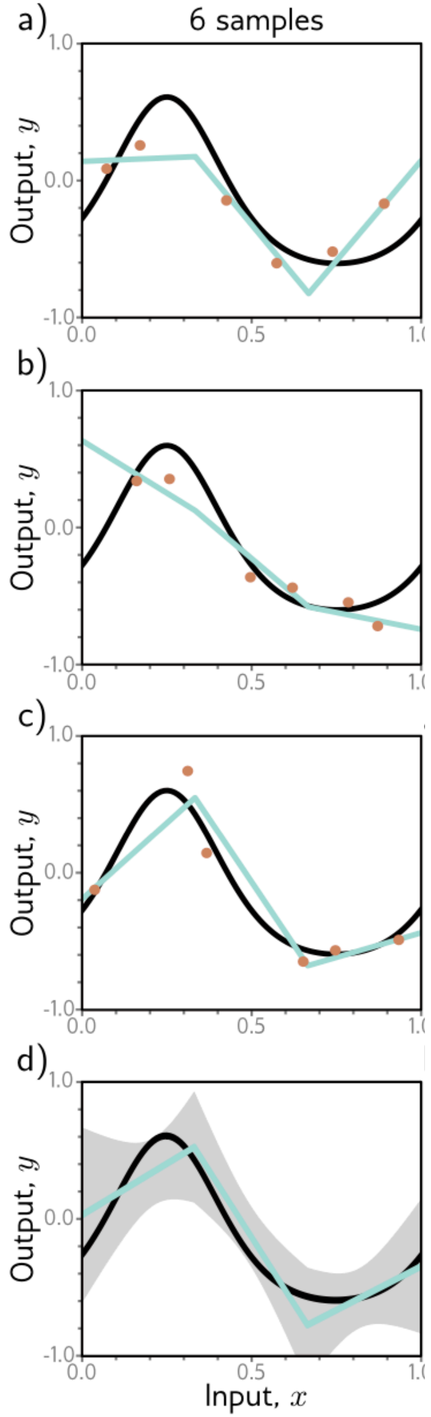



Figura de Understanding Deep Learning

# Redes Neuronales: Reducir la Varianza

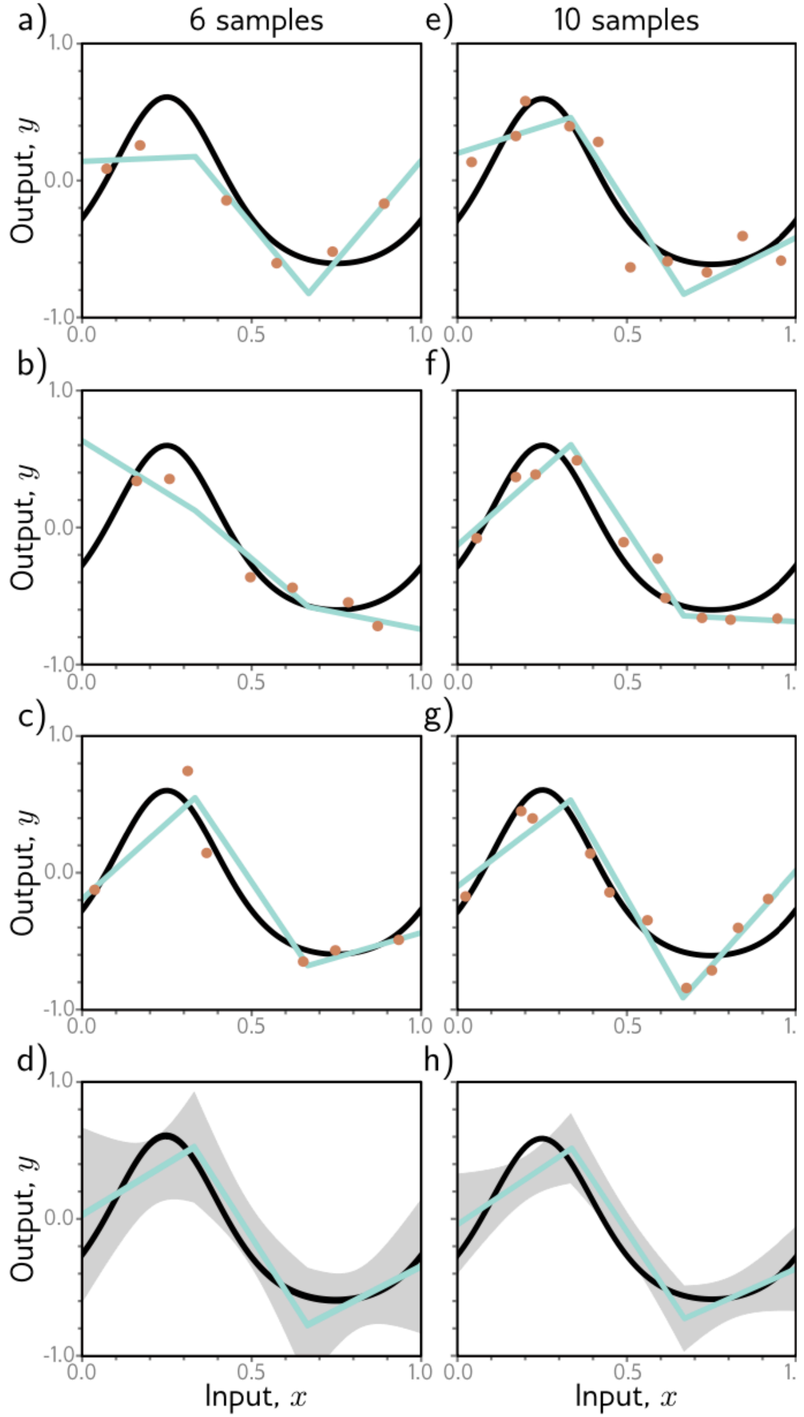



Figura de Understanding Deep Learning

# Redes Neuronales: Reducir la Varianza

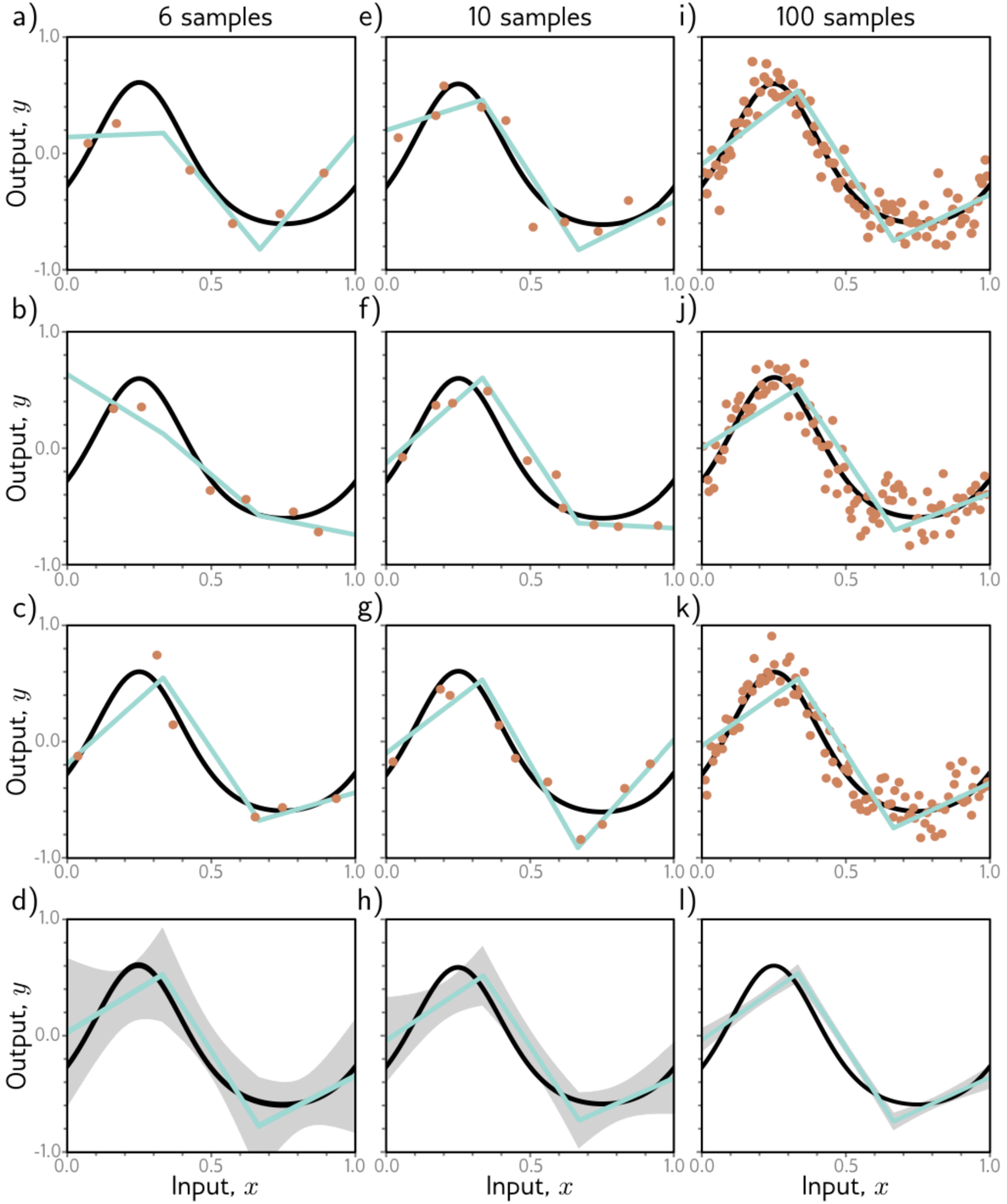



Figura de Understanding Deep Learning

# Redes Neuronales: Reducir el Bias

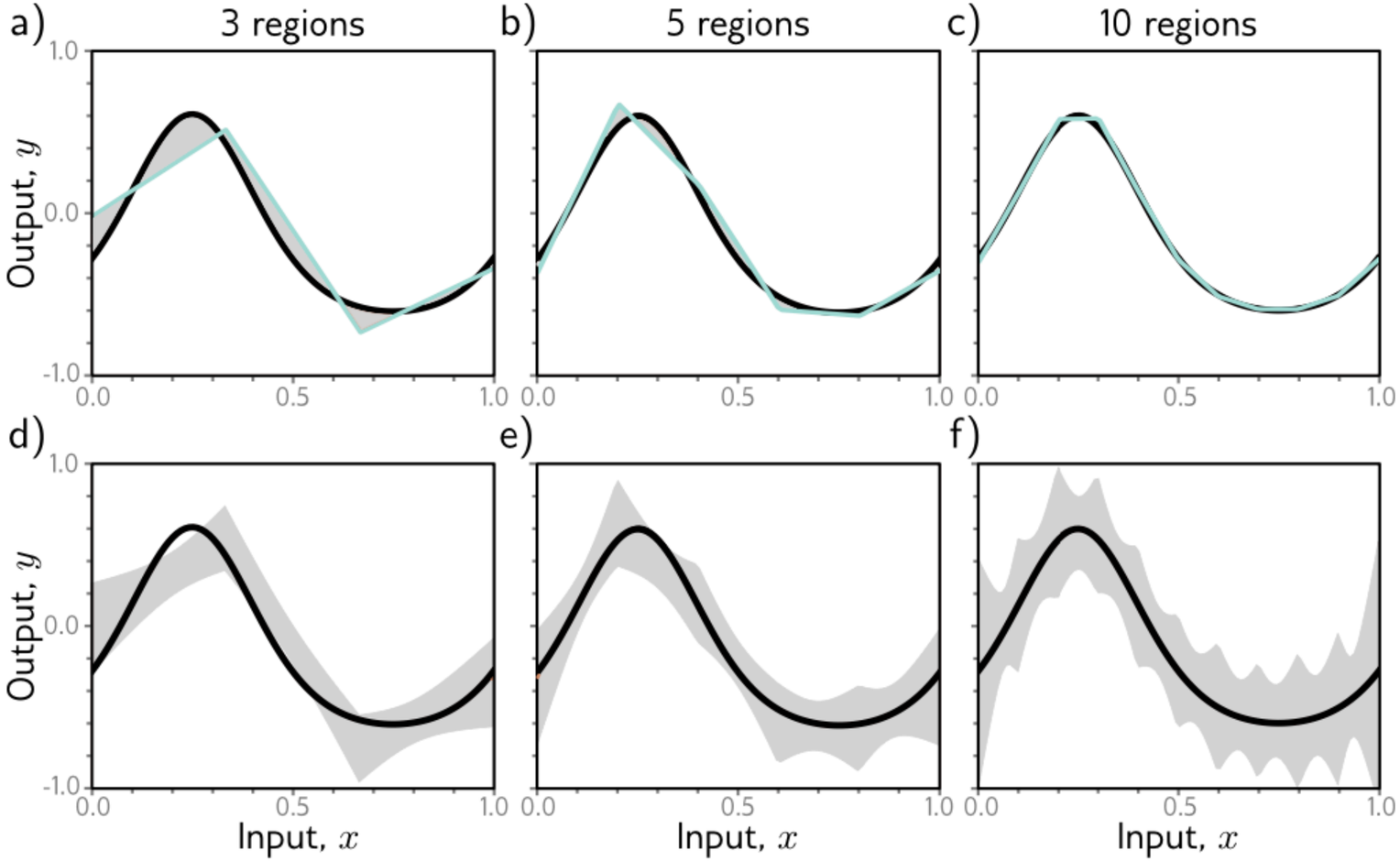



Figura de Understanding Deep Learning

# Redes Neuronales: Overfitting

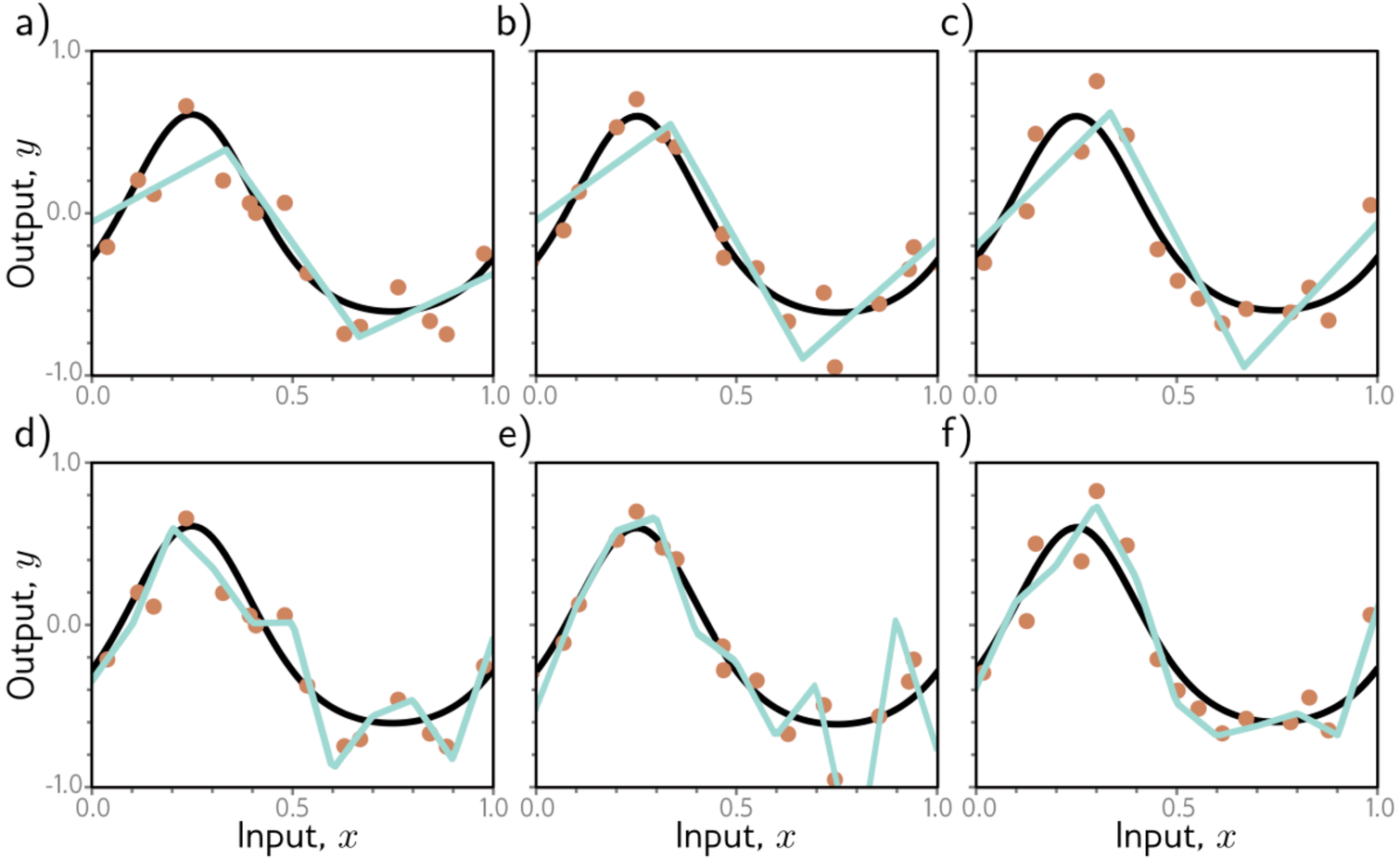



Figura de Understanding Deep Learning

# Redes Neuronales: Trade-off

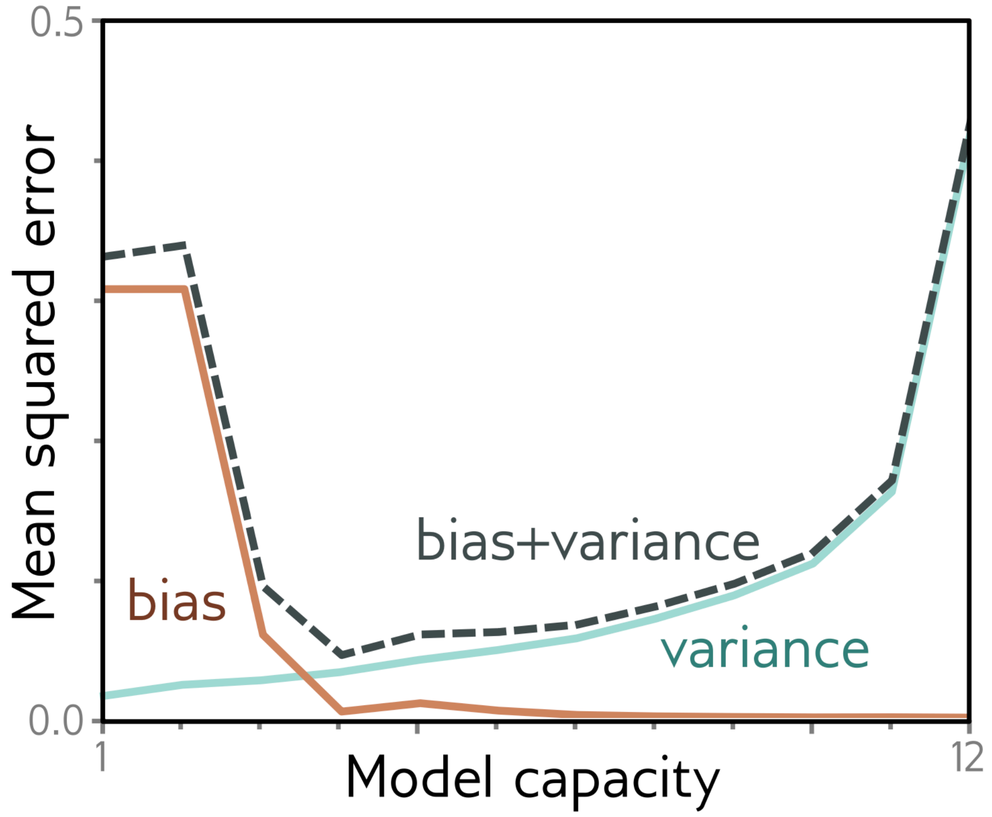



Figura de Understanding Deep Learning

# Redes Neuronales: The Curse of Dimensionality

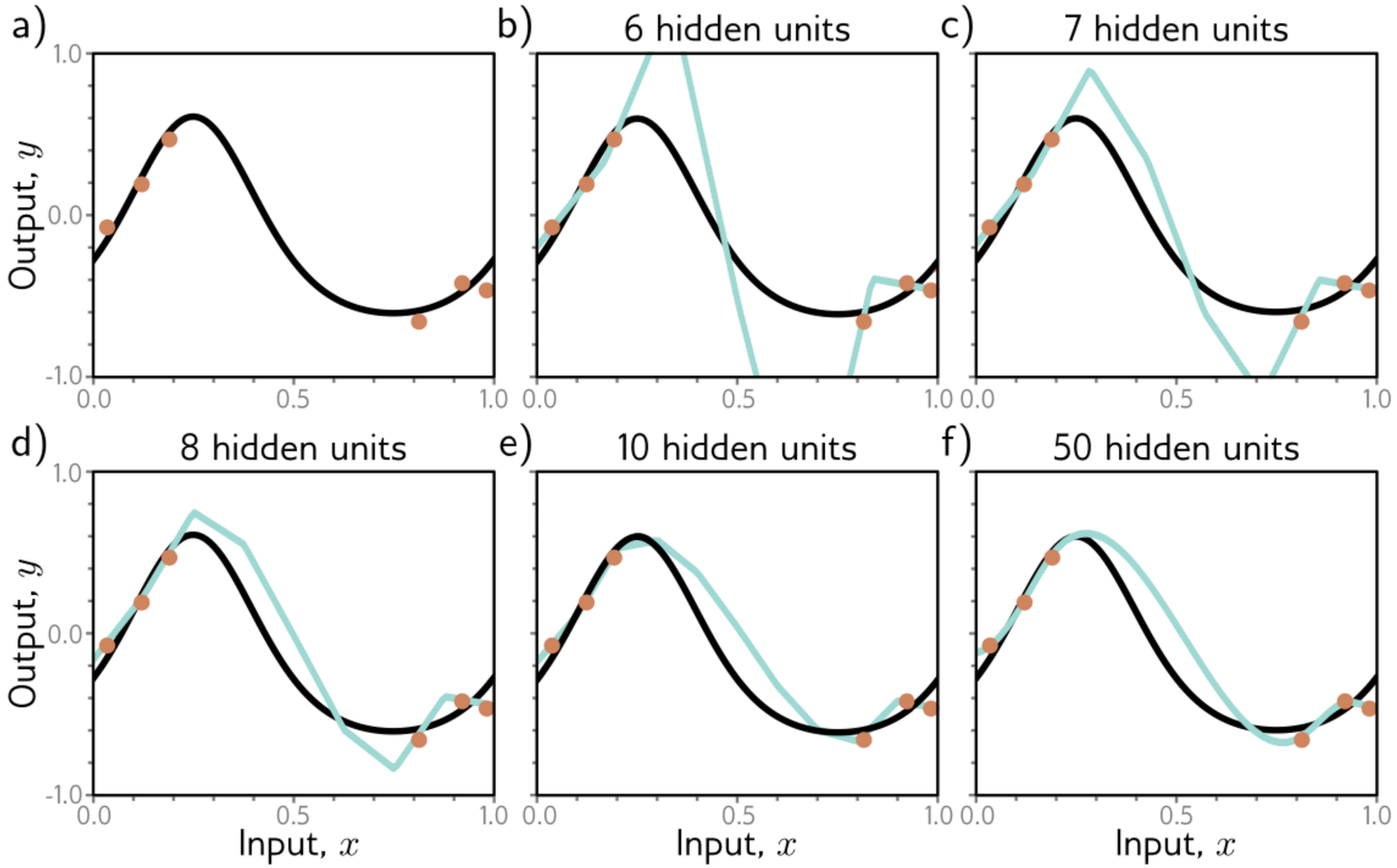
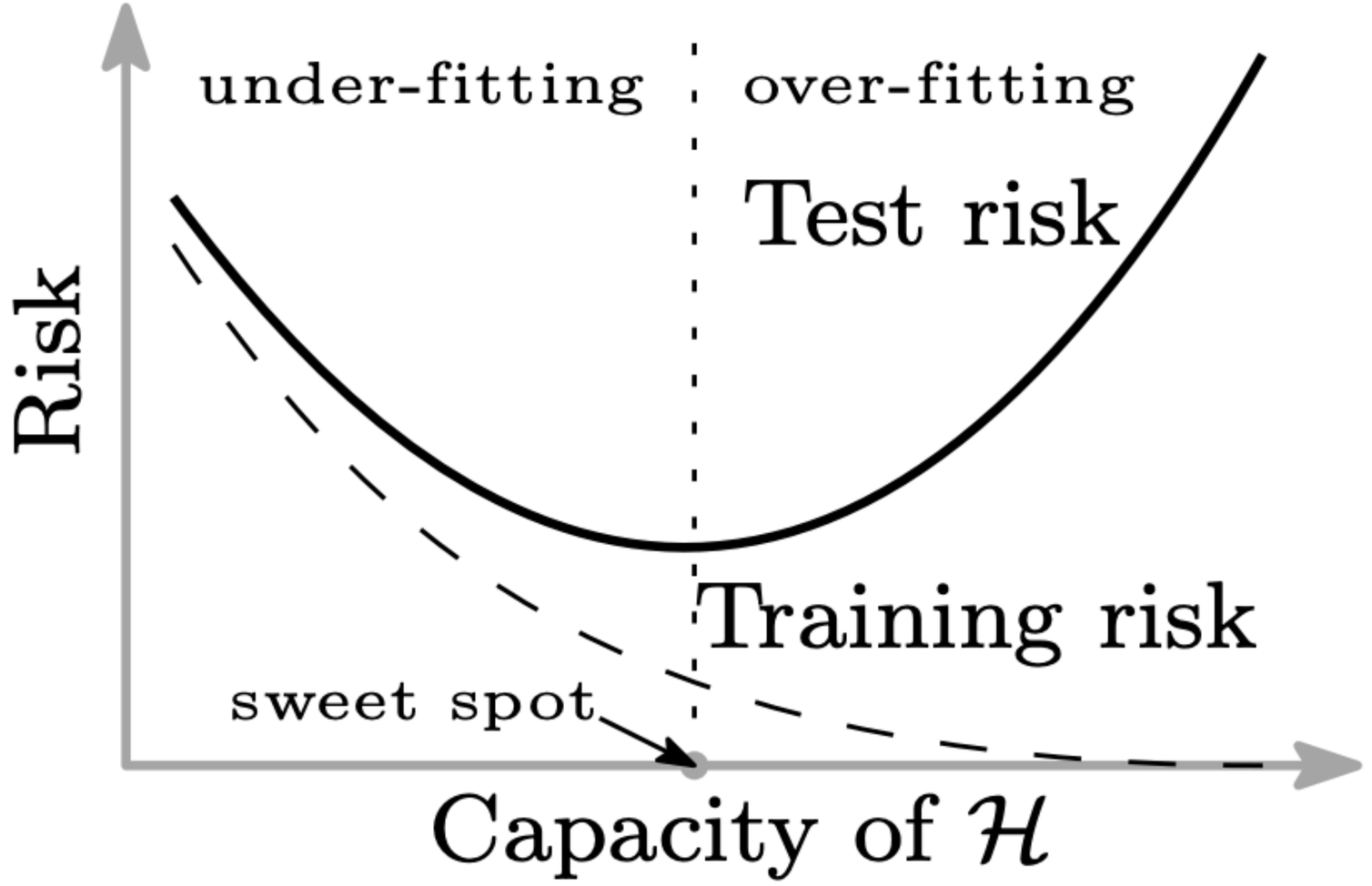


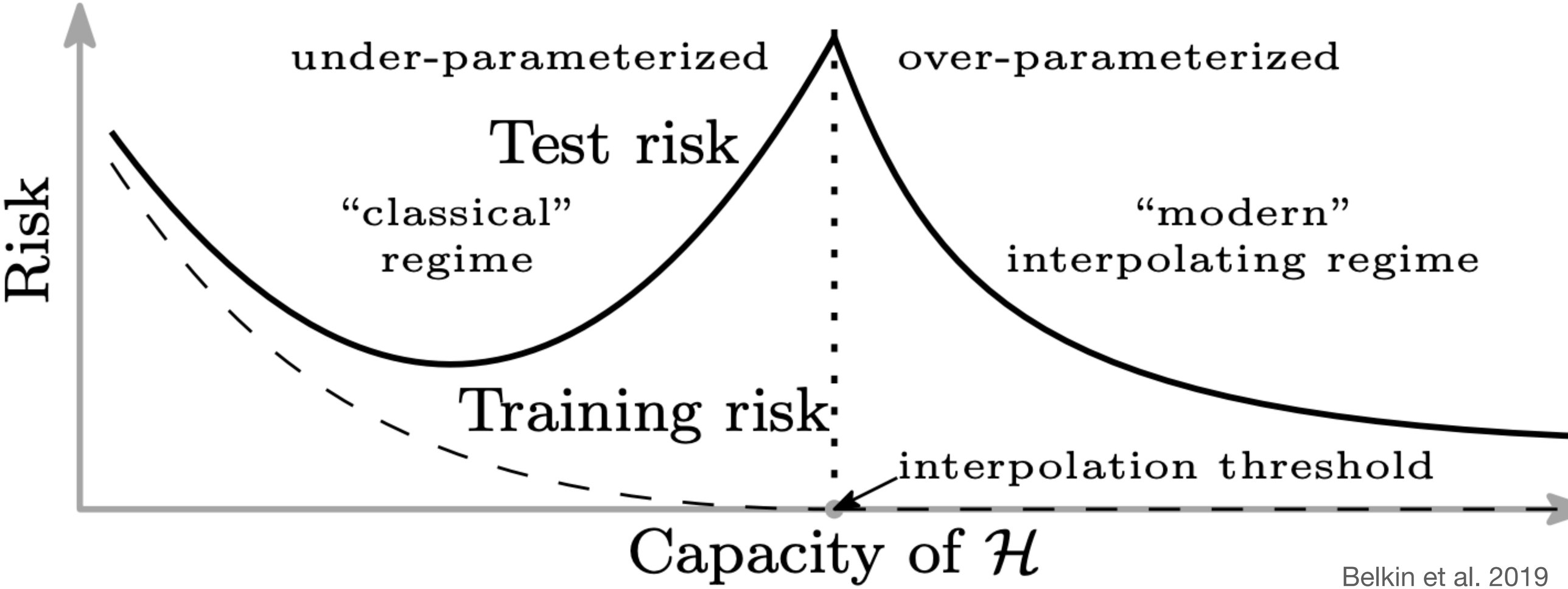

Figura de Understanding Deep Learning

# Redes Neuronales: Trade Off Bias-Variance



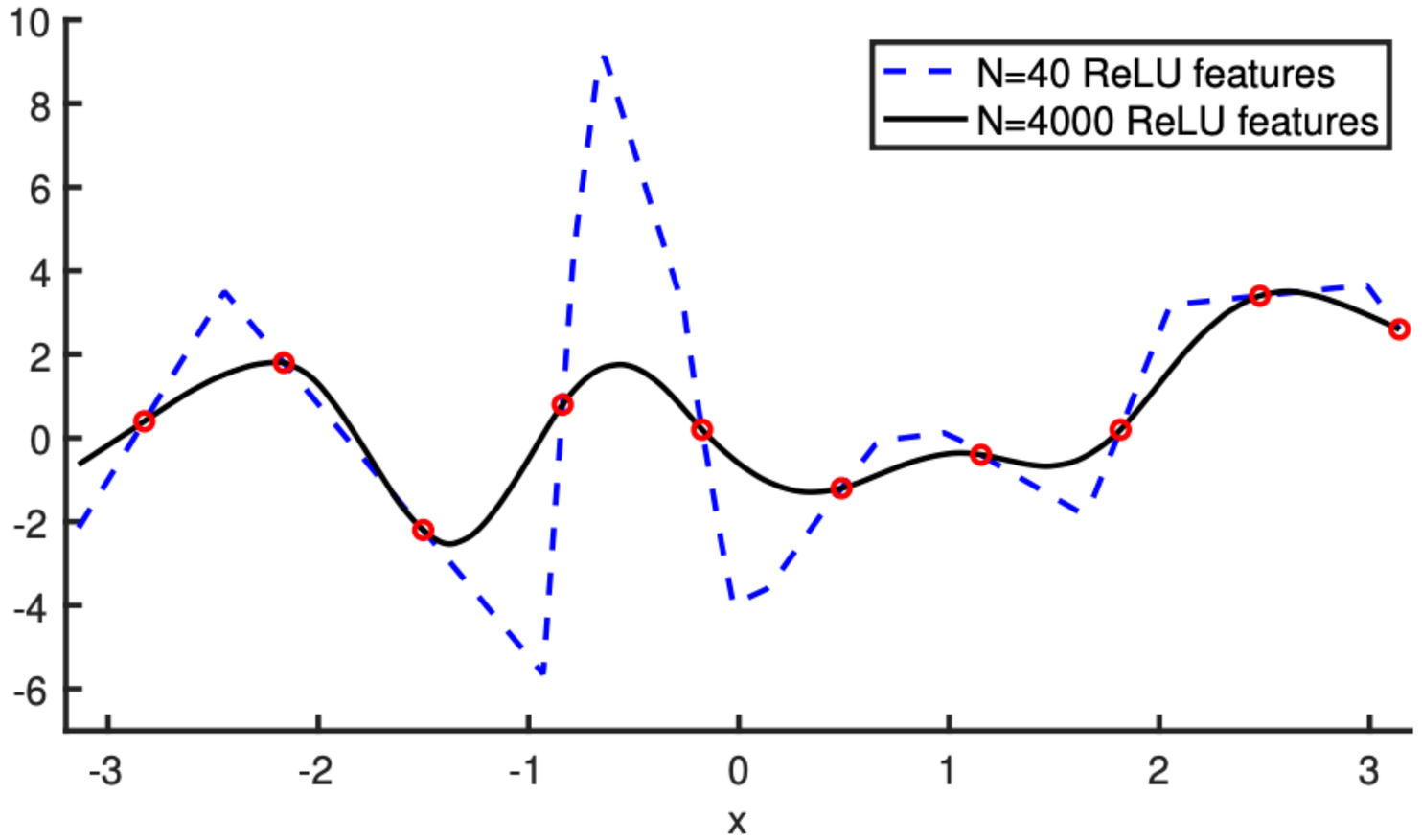

Belkin et al. 2019

# Redes Neuronales: Trade Off Bias-Variance




Belkin et al. 2019

# Redes Neuronales: Double Descent




Para mas detalles: Rocks & Mehta 2022

Belkin et al. 2019

# Redes Neuronales: Double Descent

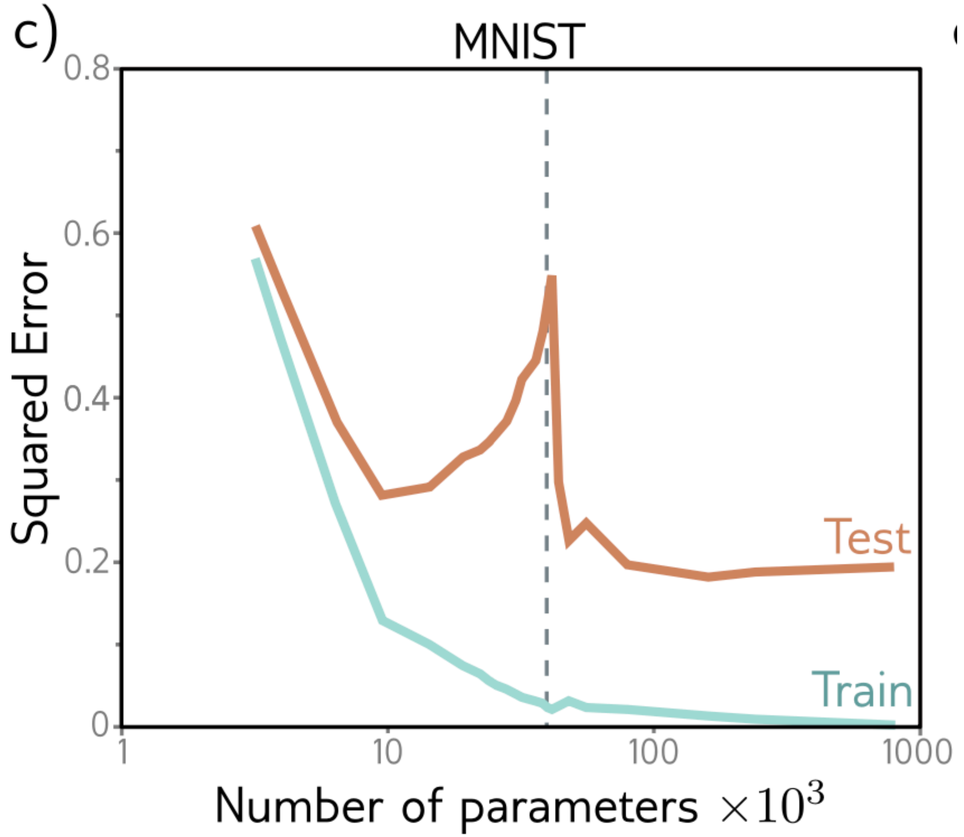


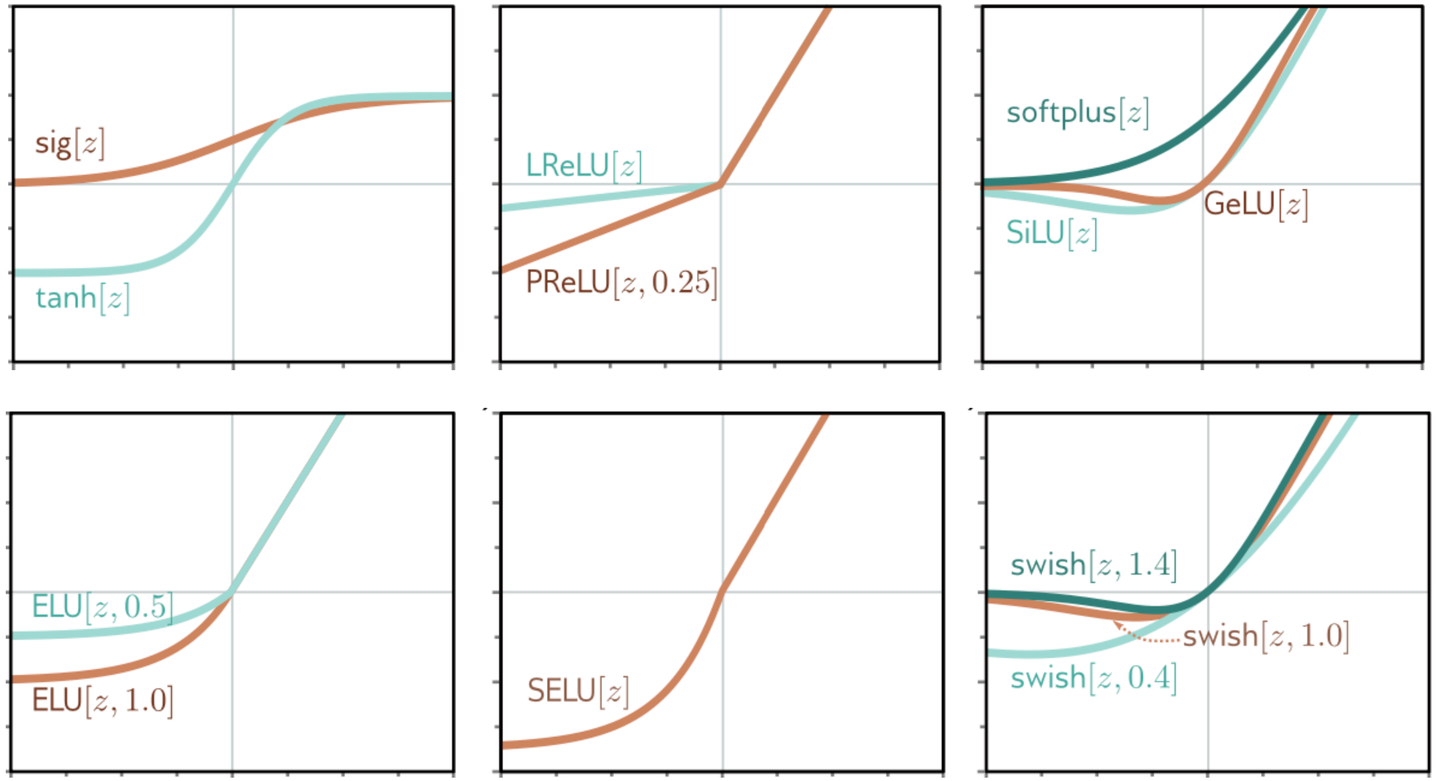

Figura de Understanding Deep Learning

# Mañana

- Descenso del Gradiente
- Estocasticidad para Funciones de Pérdida no Convexas
- Aprendizaje Adaptativo
- Teorema de Aproximación Universal
- Redes Neuronales Profundas

# Redes Neuronales: Función de Activación

## Other Activation Functions




Understanding Deep Learning, S.J.D Prince