



Curso de
**Fundamentos de
Redes Neuronales
con Python y Keras**

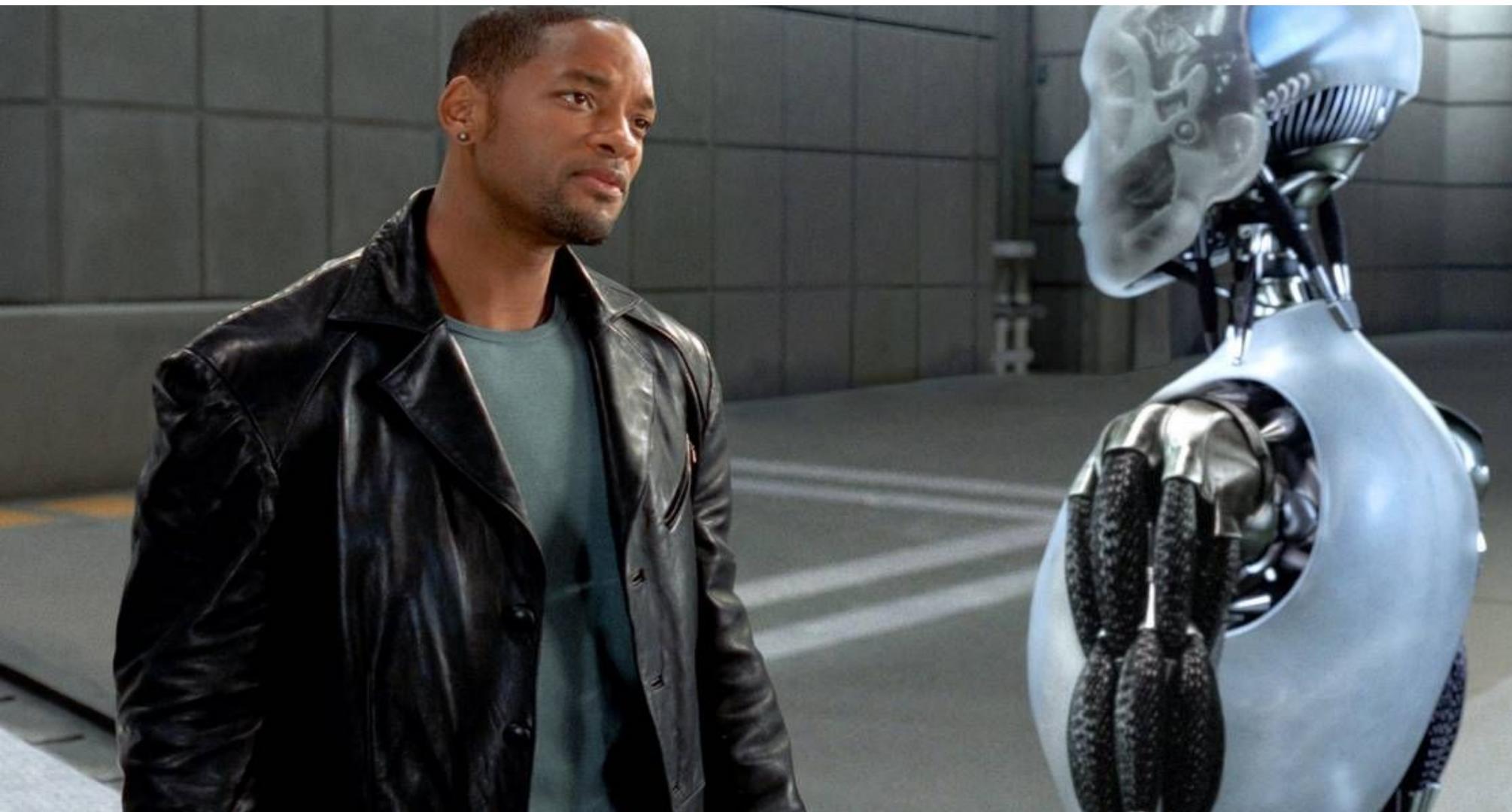
Carlos Alarcón

¿Redes neuronales artificiales?

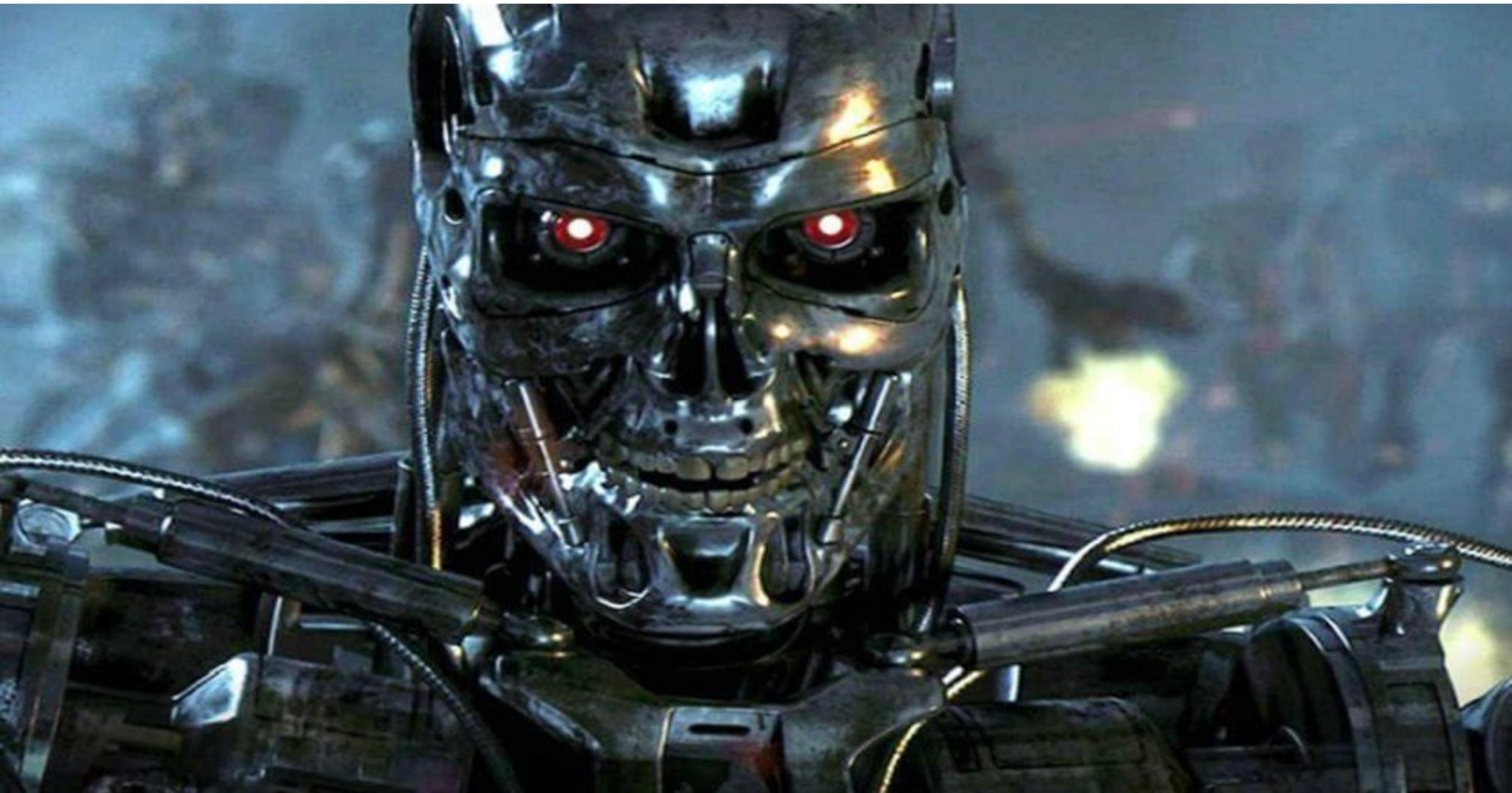
■ ¿Redes neuronales artificiales?



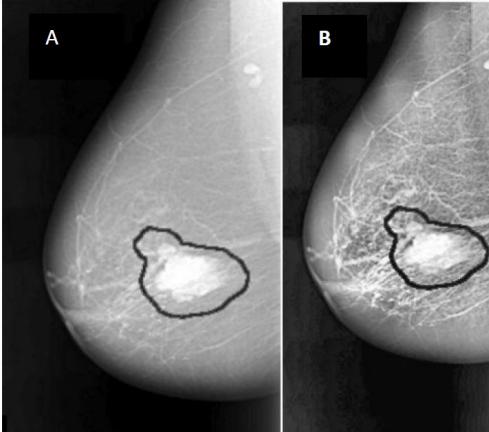
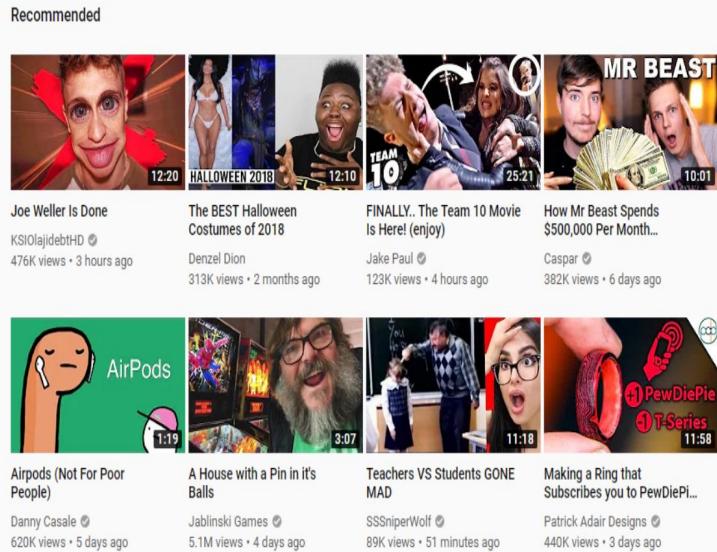
■ ¿Redes neuronales artificiales?



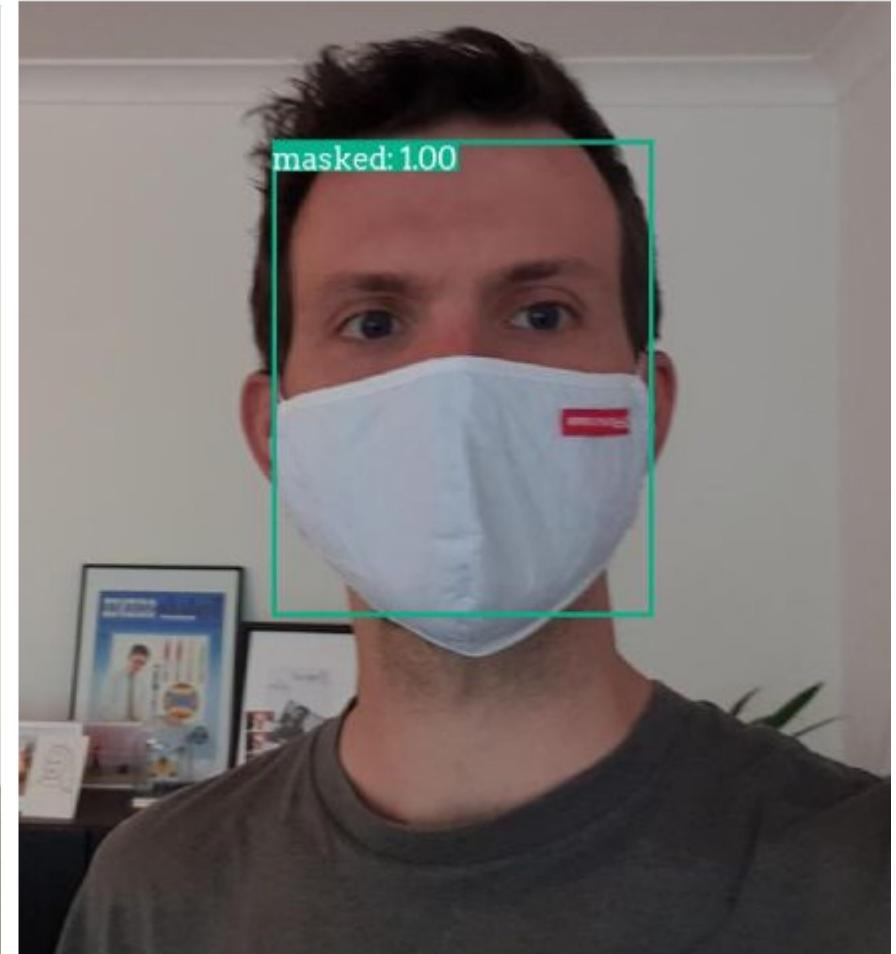
■ ¿Redes neuronales artificiales?



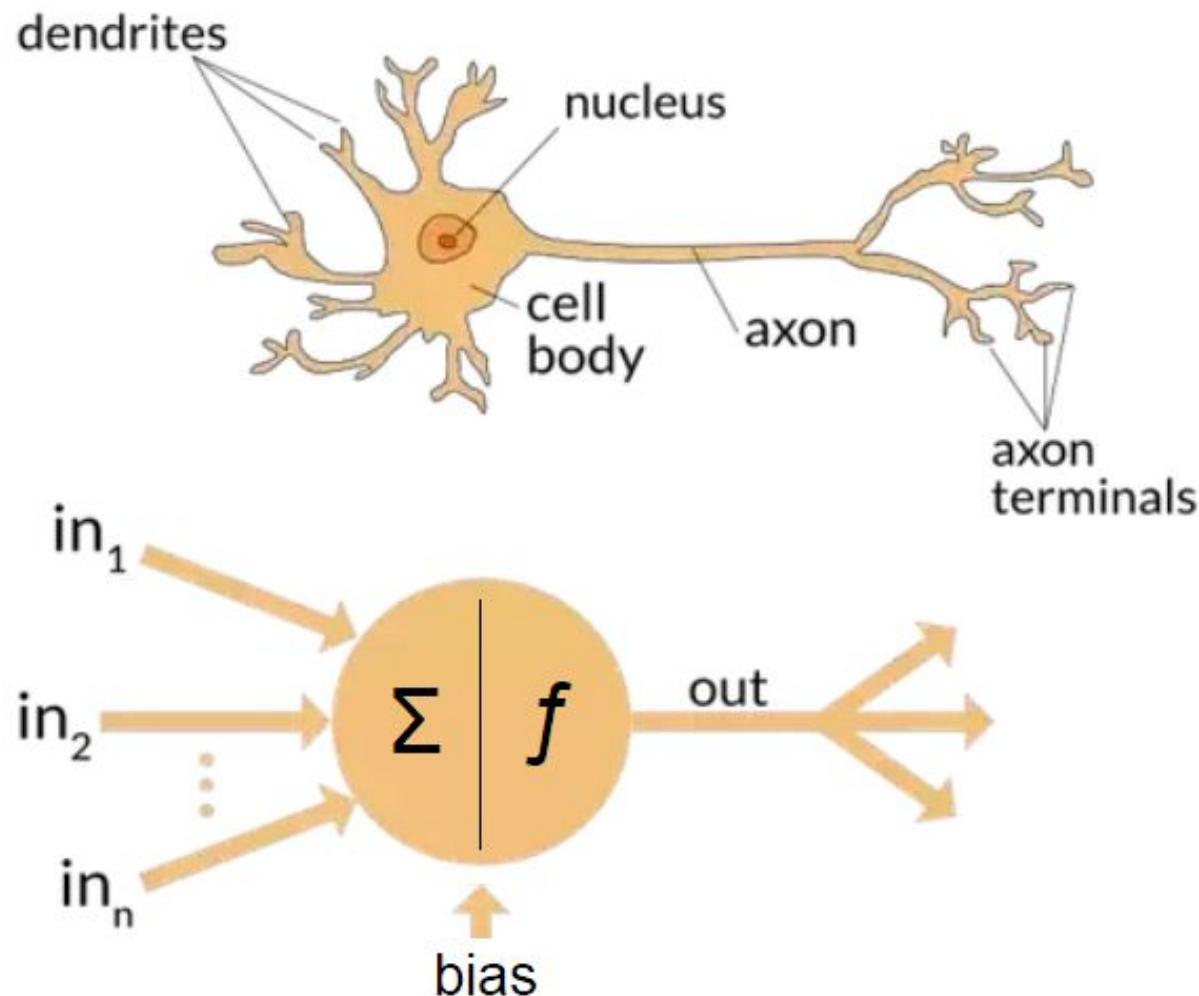
¿Redes neuronales artificiales?



Redes neuronales artificiales



Redes neuronales



¿Conocimiento previo?



**Curso de Matemáticas
para Data Science:
Cálculo**

$$\frac{dx}{dy}$$

¿Qué herramientas
usaremos?

Frameworks



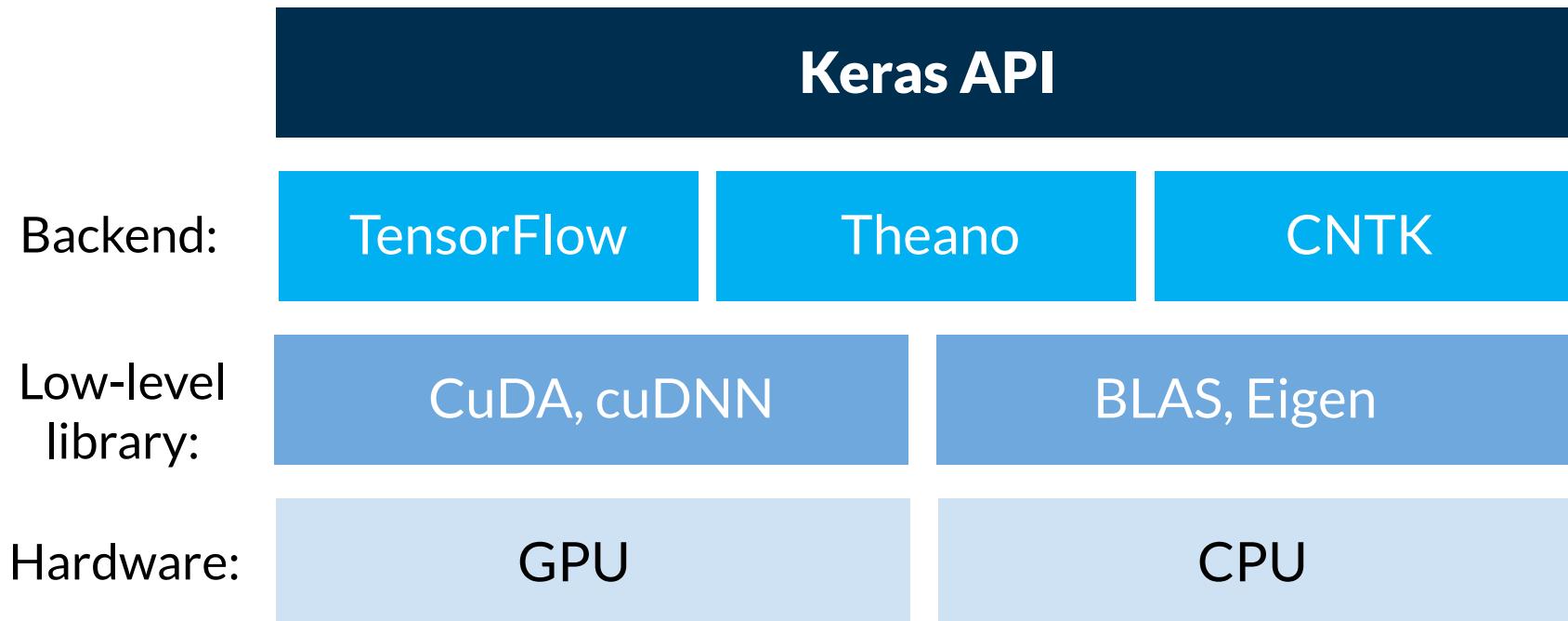
theano

The Keras logo consists of a solid red square containing a white 'K'. To the right of the square, the word "Keras" is written in a large, black, sans-serif font.

The PyTorch logo consists of a red circle with a white dot inside, followed by the word "PyTorch" in a black, sans-serif font.

The torch logo consists of three teal-colored circles connected by lines, forming a triangular shape. To the right of the logo, the word "torch" is written in a black, sans-serif font.

Frameworks

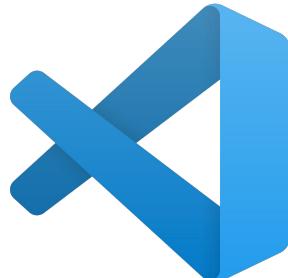


Notebooks



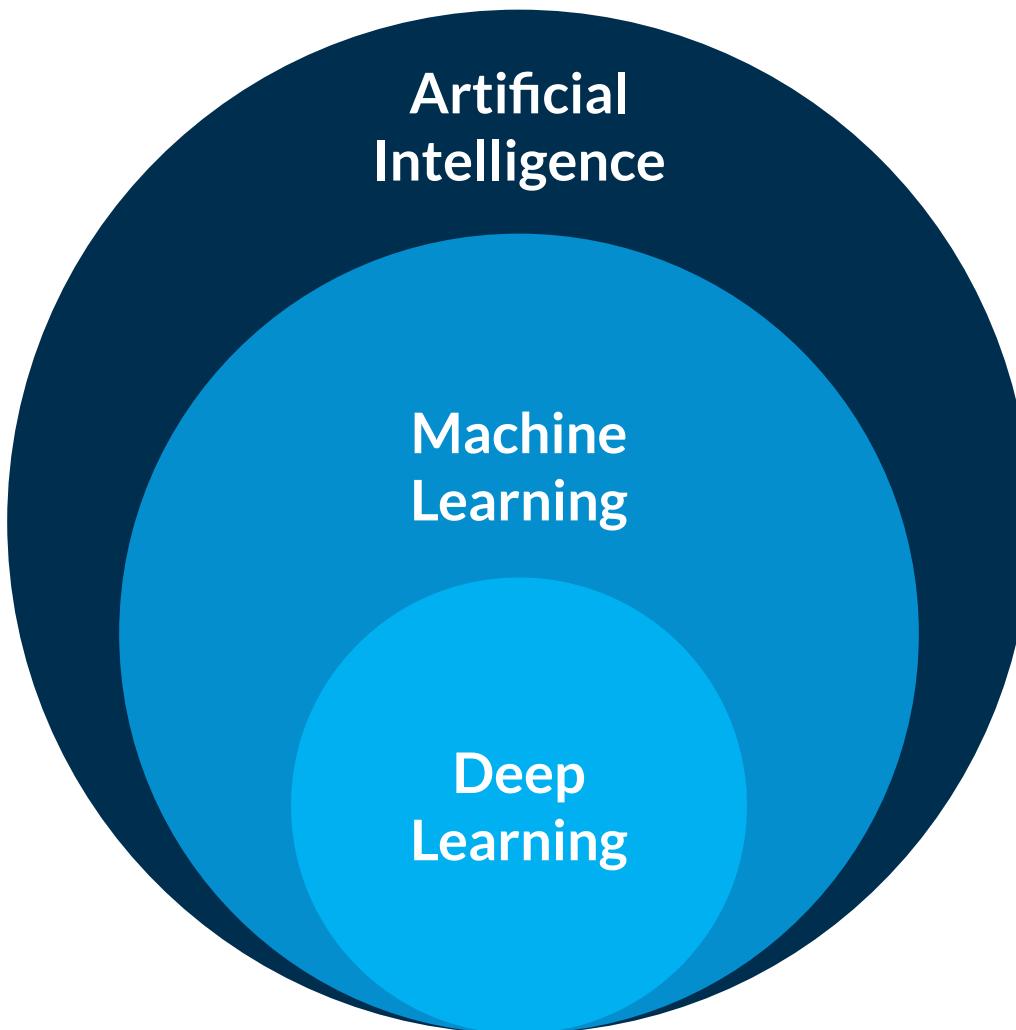
Deepnote

colab



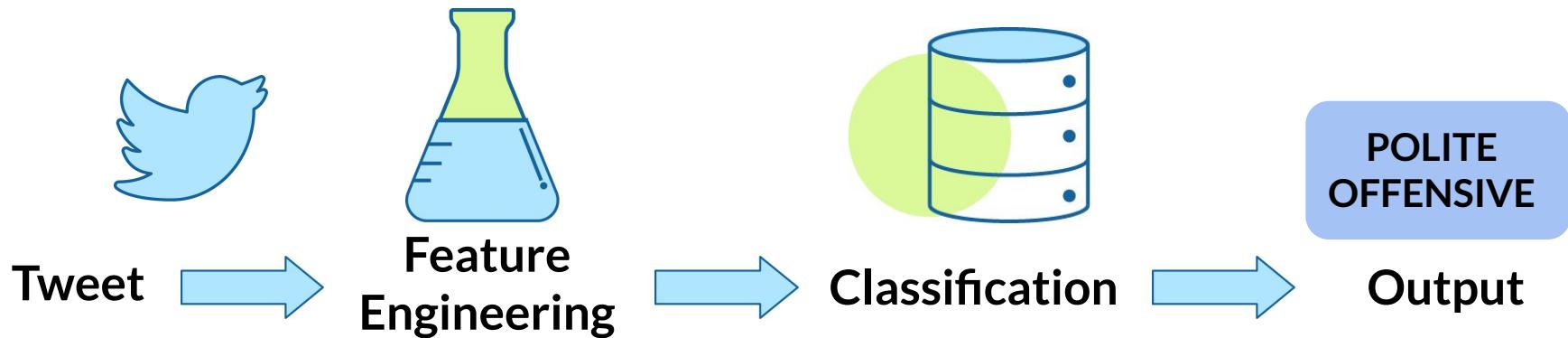
Deep learning

Deep learning

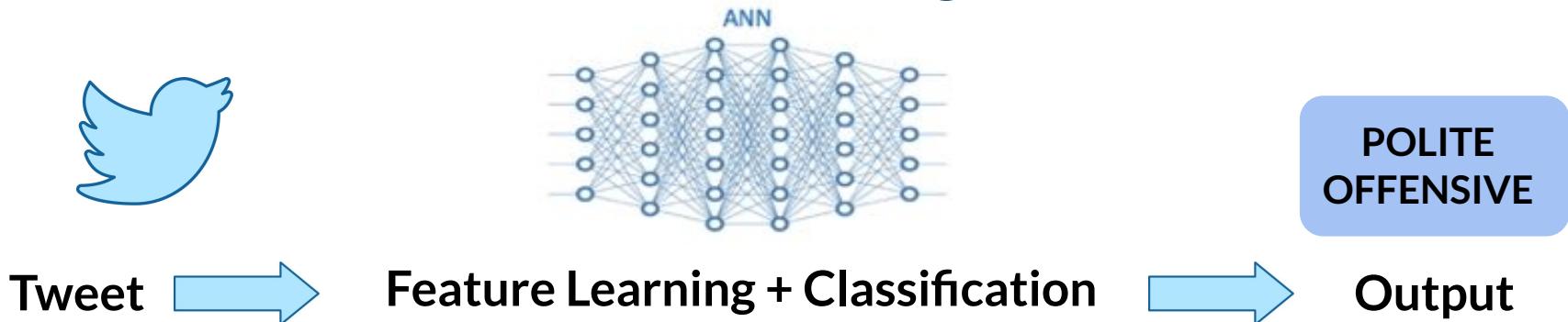


Deep learning

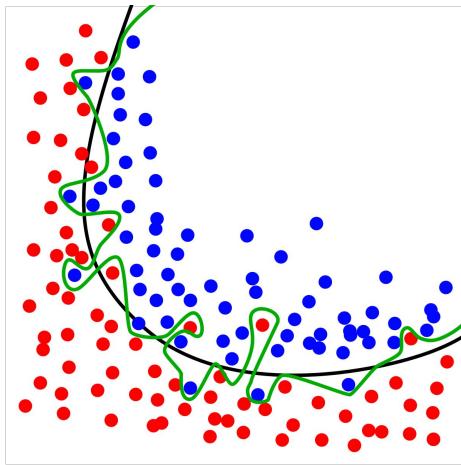
Traditional Machine Learning



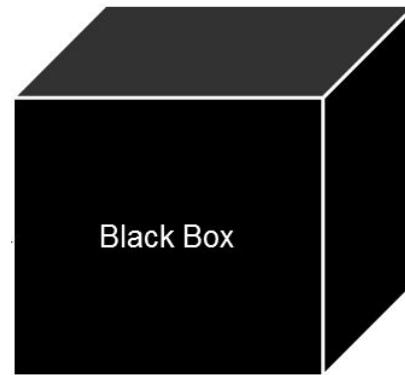
Deep Learning



■ Problemas de deep learning

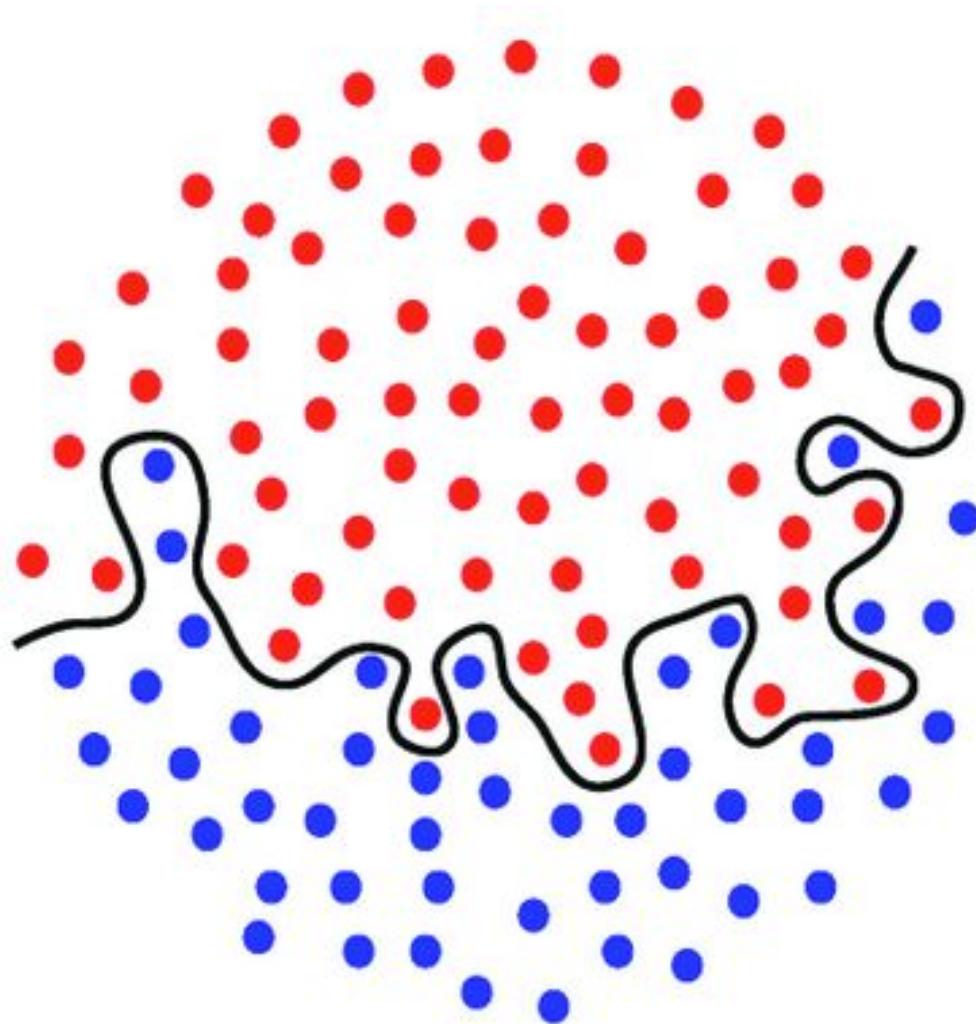


Overfitting
(Problemas de
sobre ajuste)

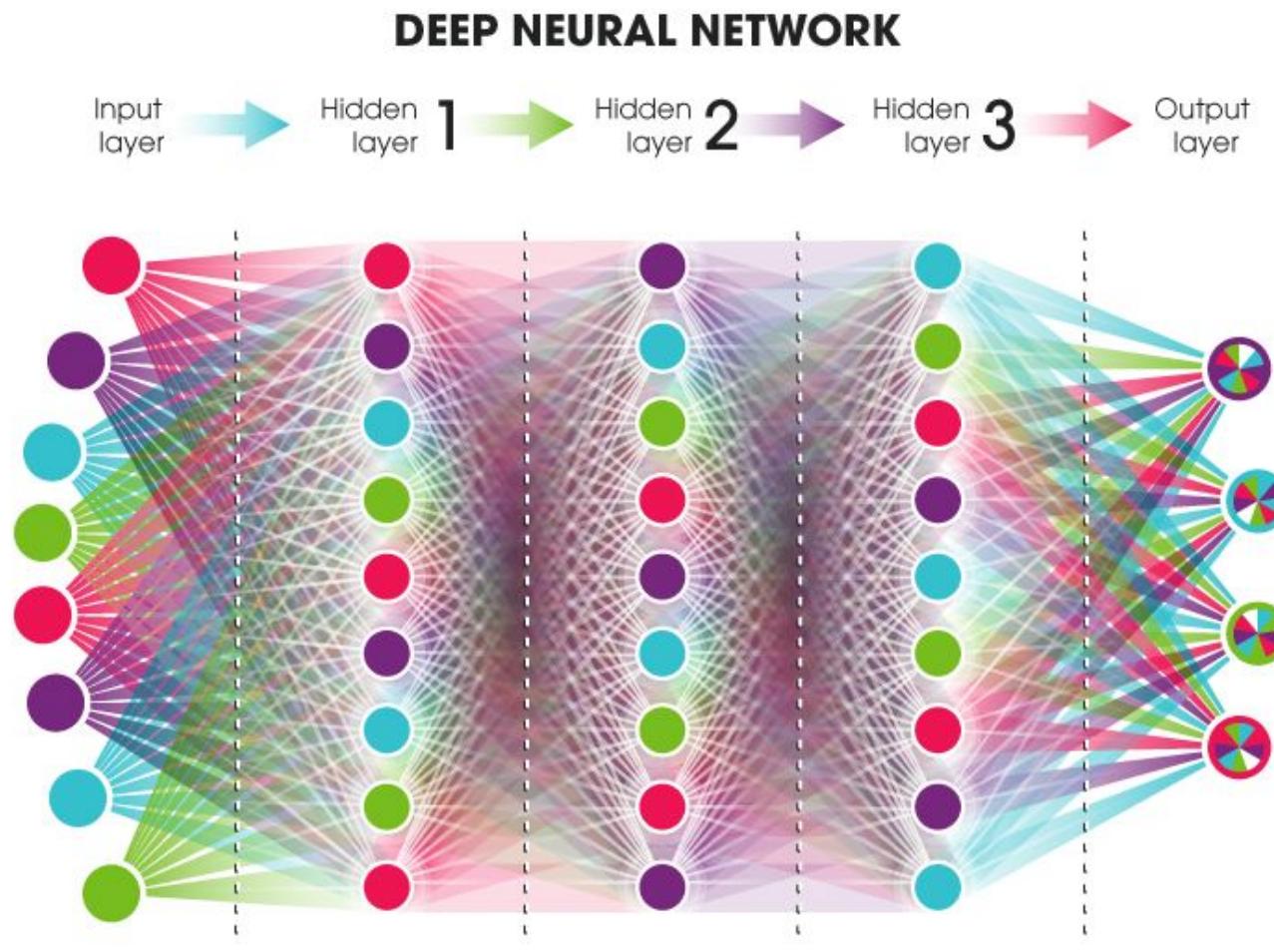


Caja negra
¿Qué está
ocurriendo
adentro?

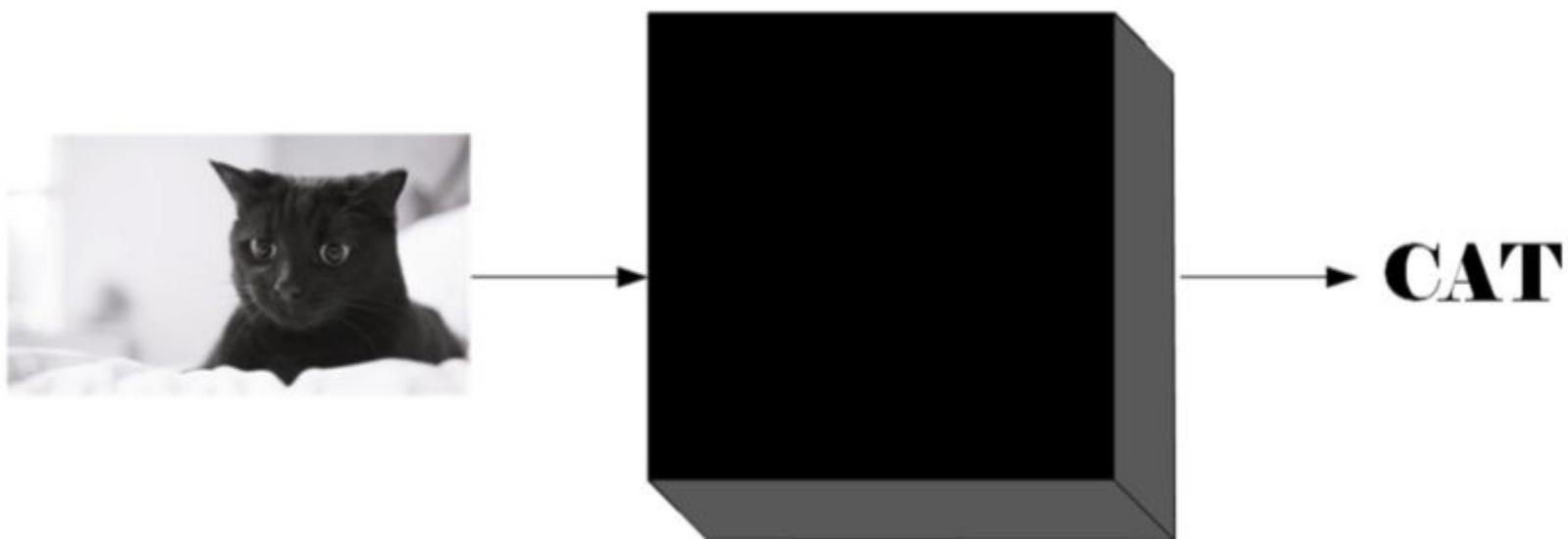
Overfitting



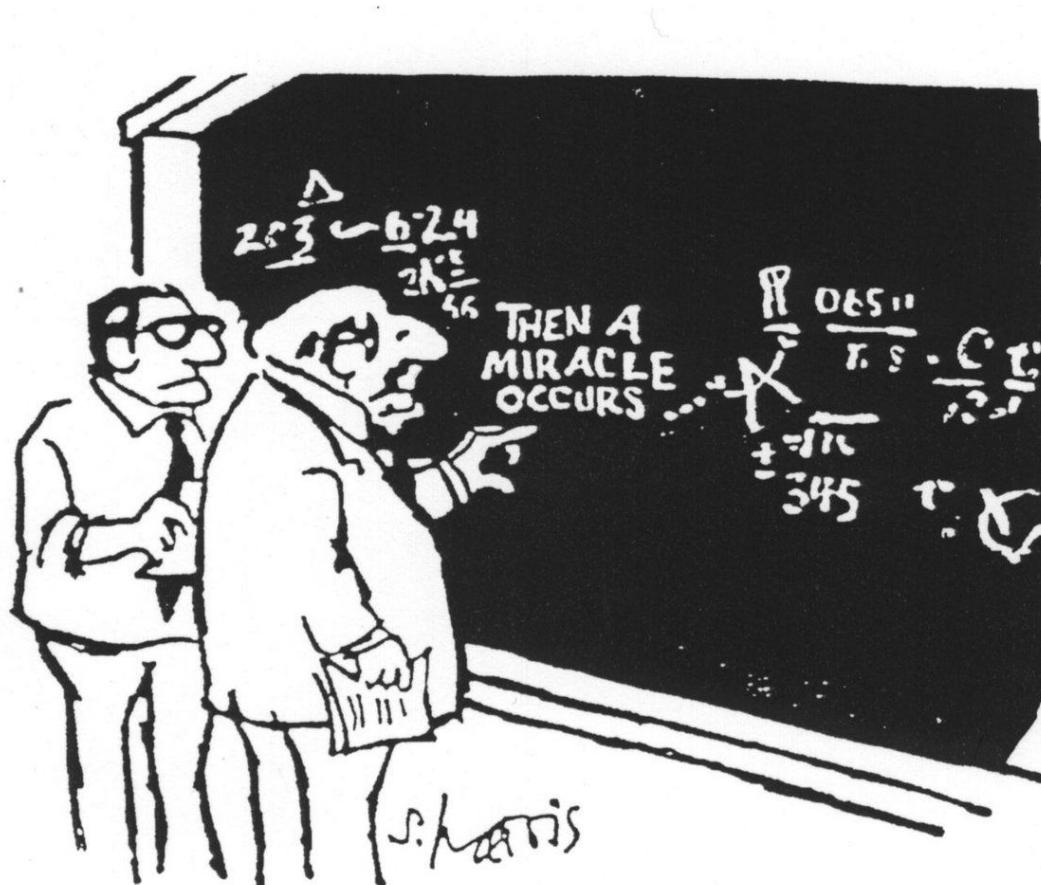
Overfitting



Cajas negras

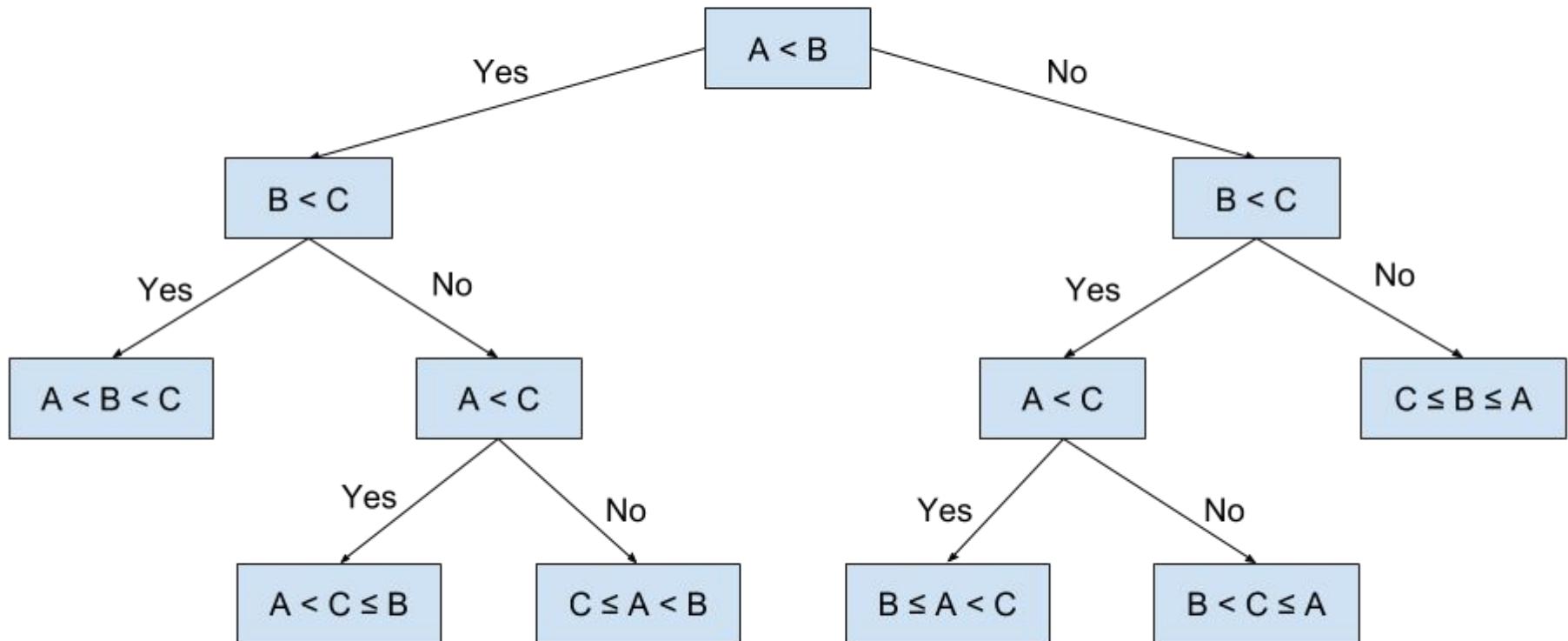


Cajas negras

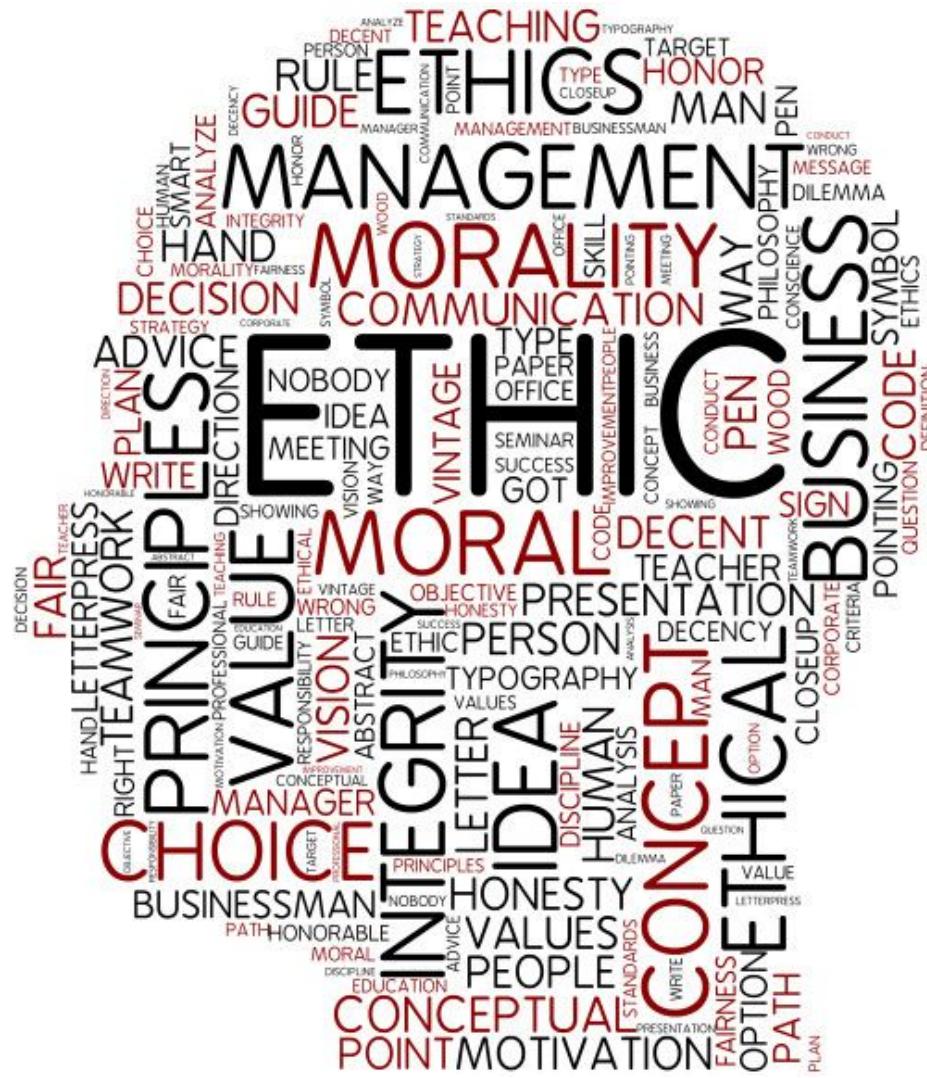


I think you should be a little
more specific, here in Step 2

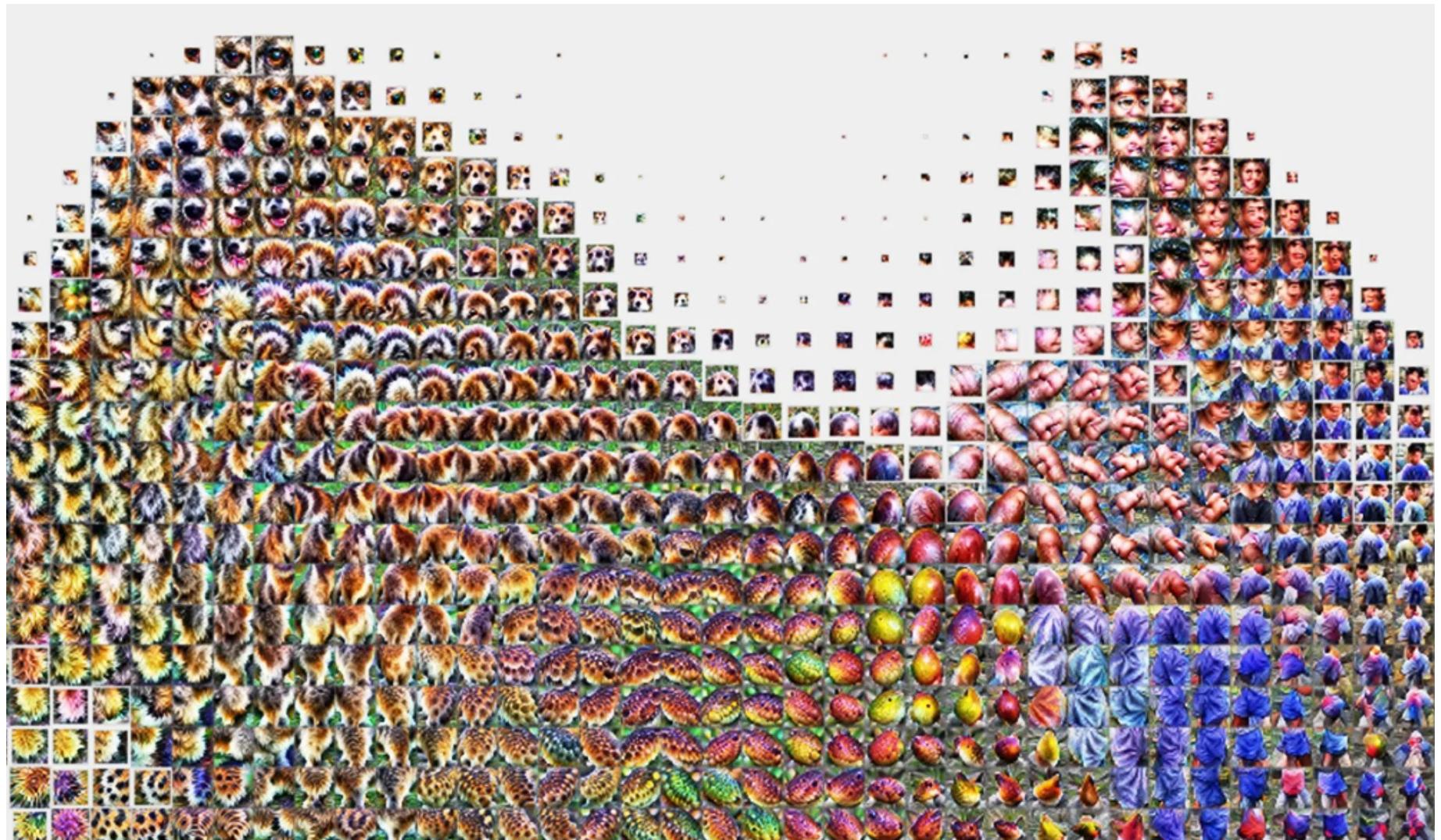
Cajas negras



Cajas negras



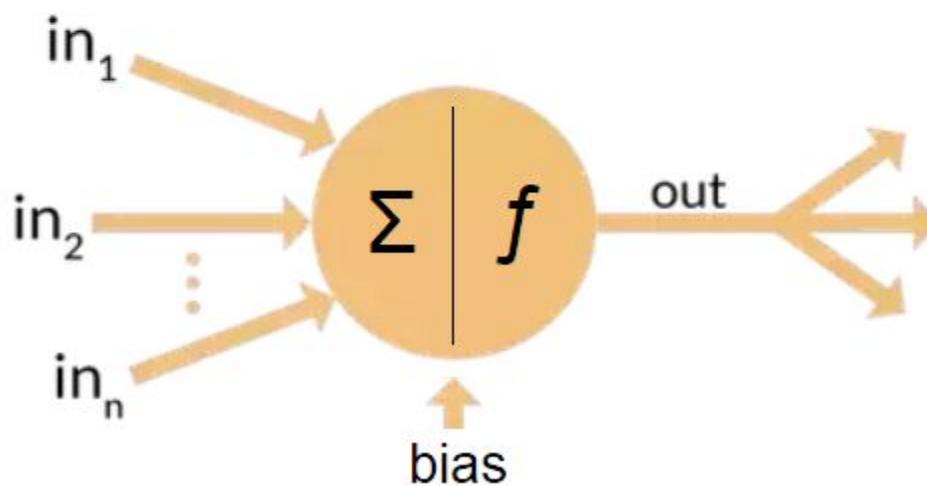
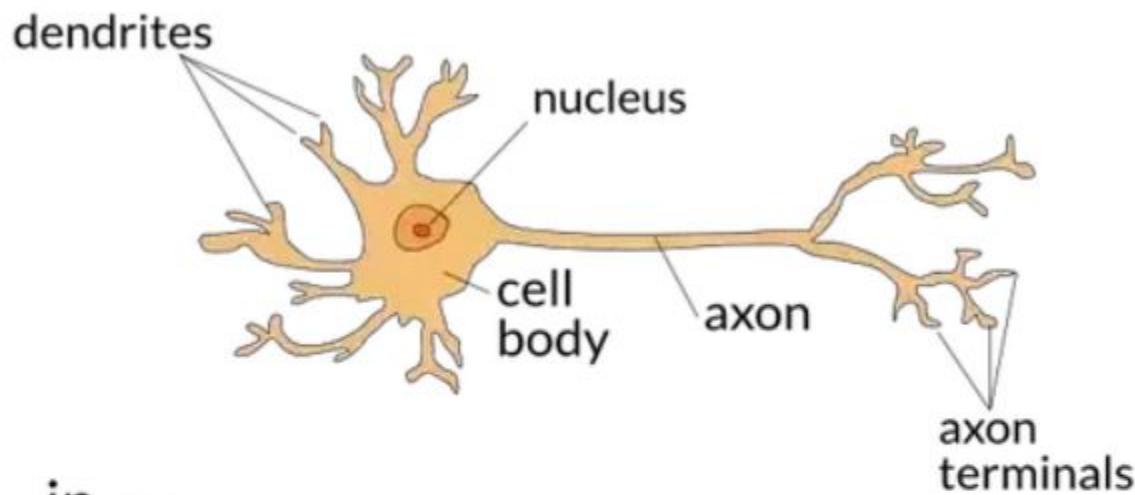
Cajas negras



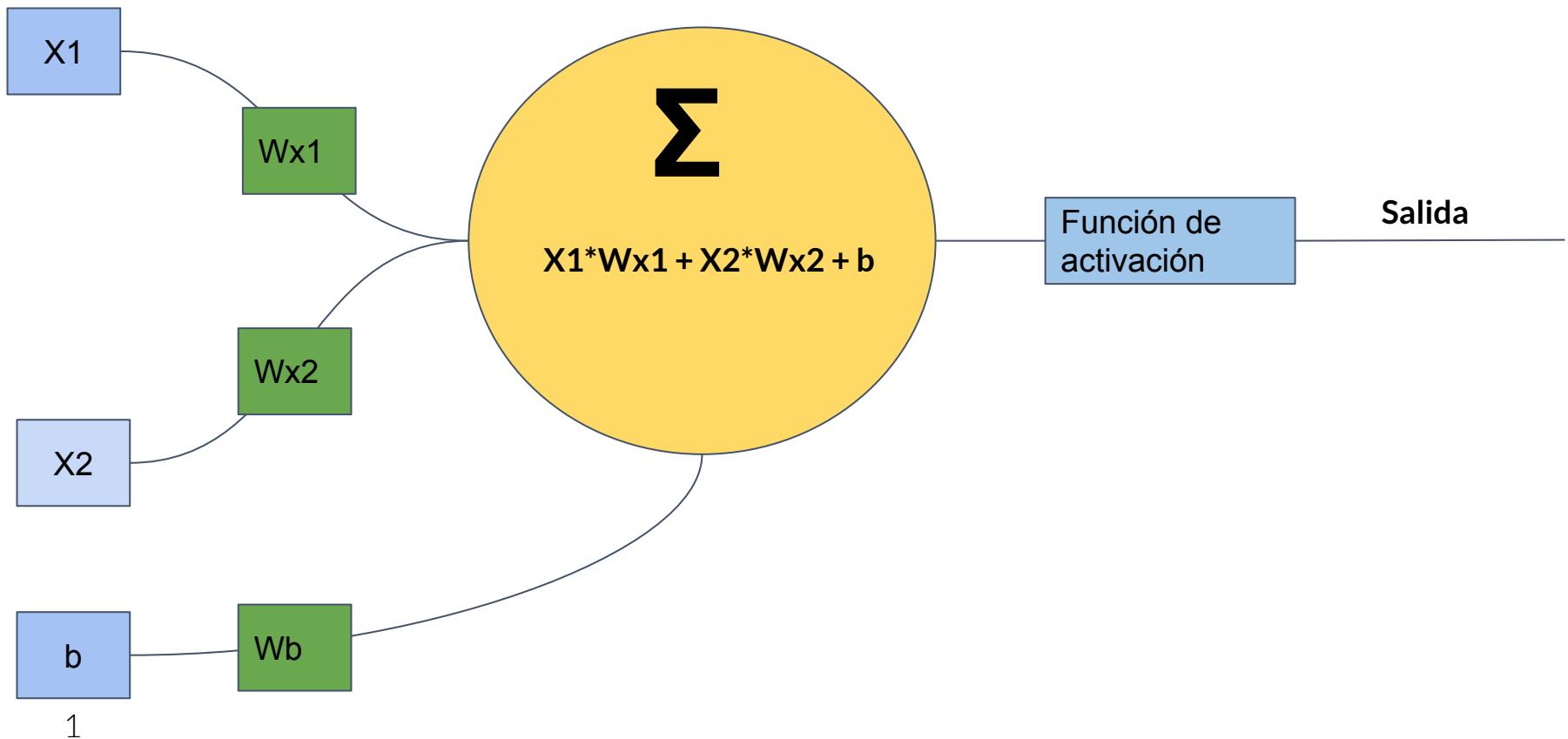
Tu primera red neuronal con Keras

La neurona: una pequeña y poderosa herramienta

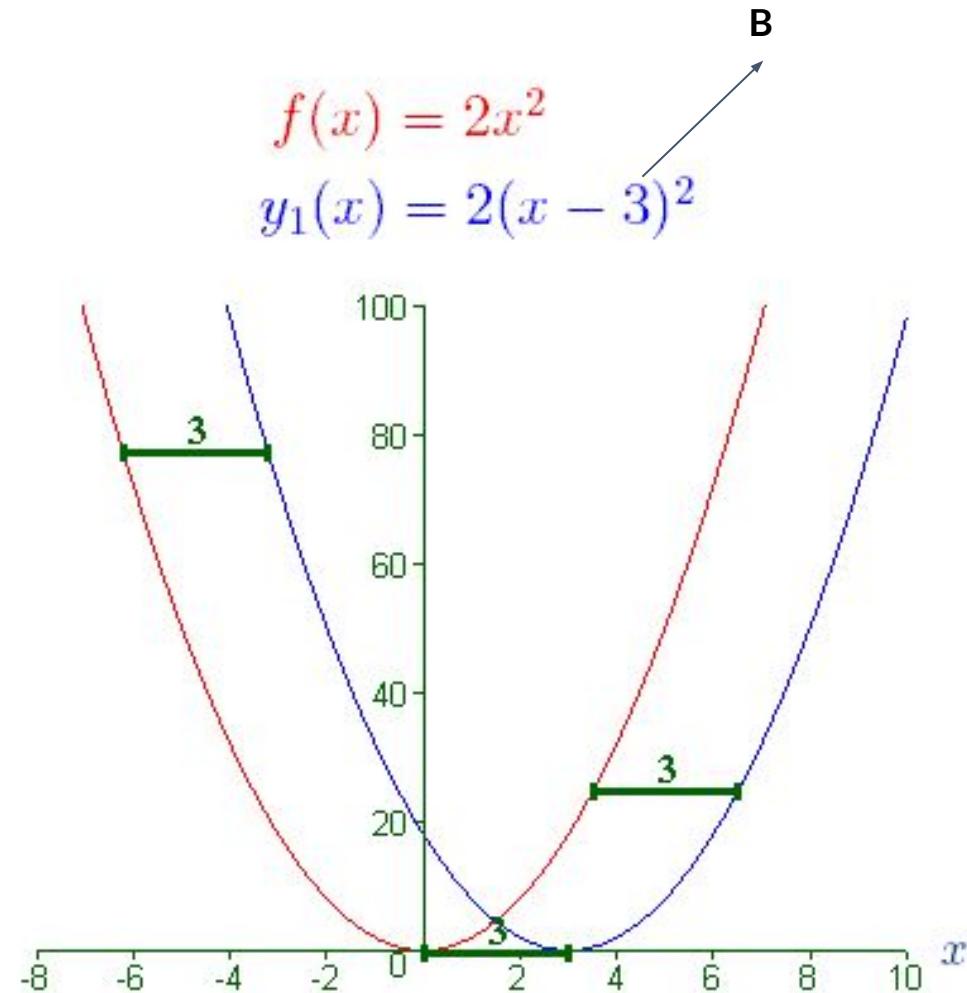
La neurona



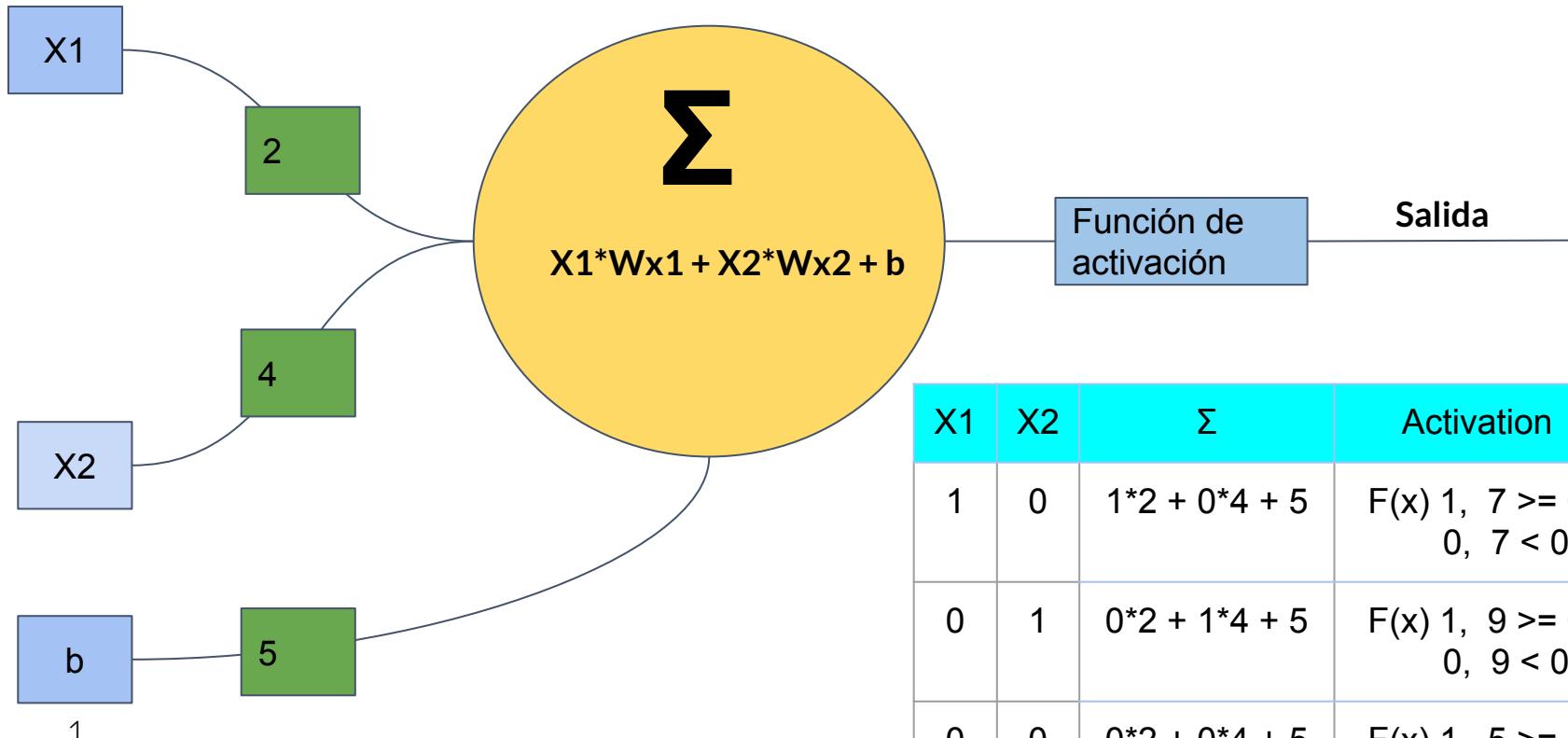
La neurona



La neurona - Bias

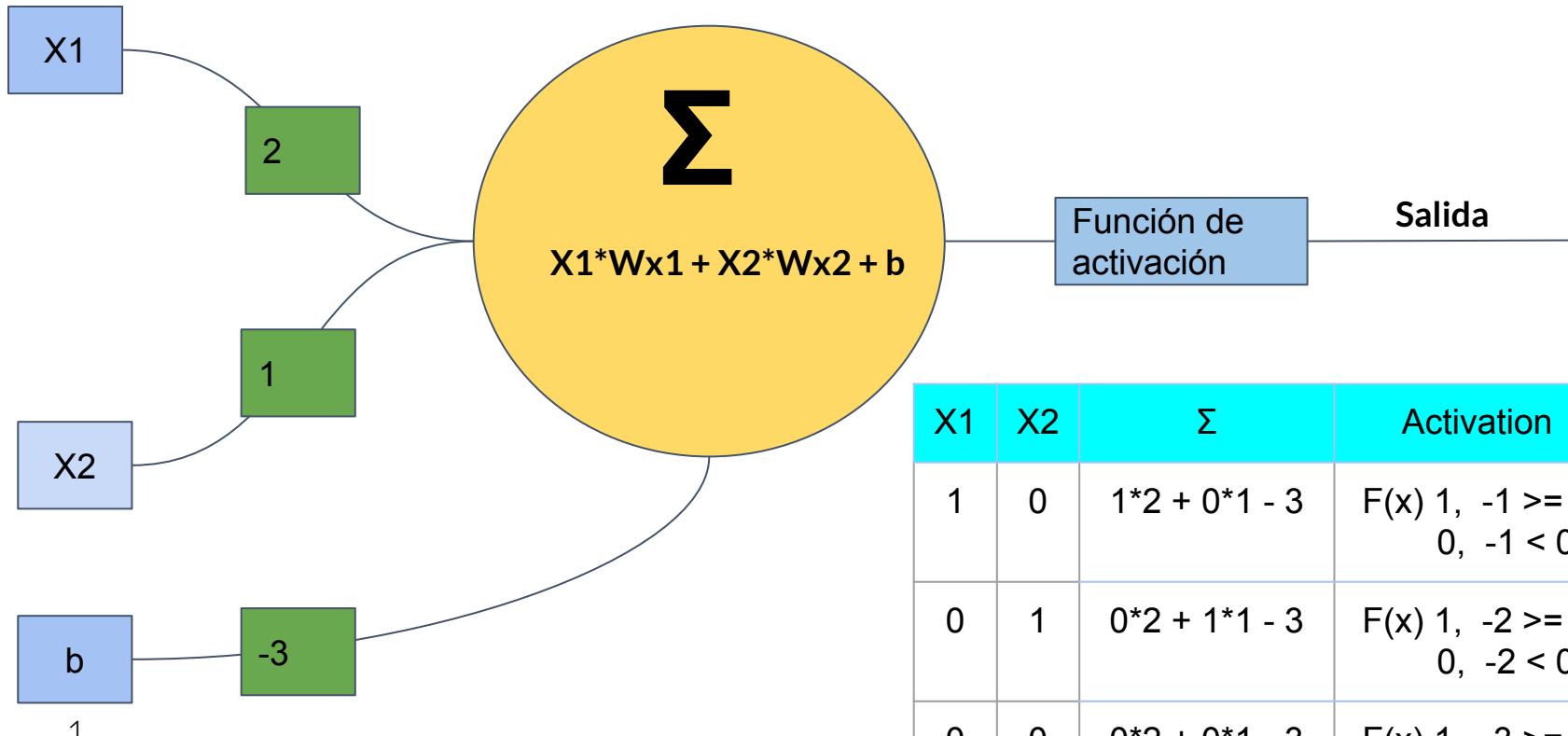


La neurona - AND



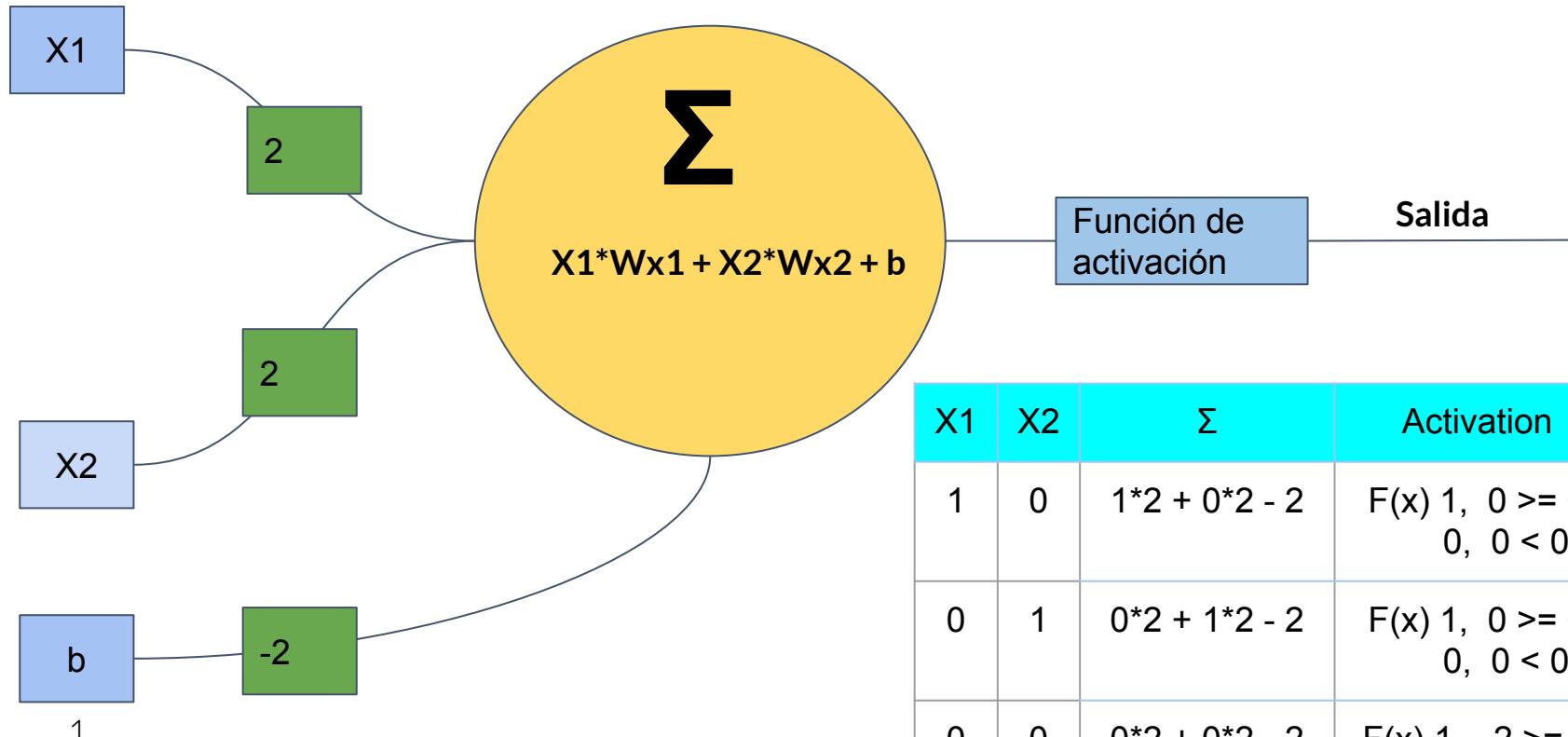
X1	X2	Σ	Activation	Salida
1	0	$1 \cdot 2 + 0 \cdot 4 + 5$	$F(x) \begin{cases} 1, & 7 \geq 0 \\ 0, & 7 < 0 \end{cases}$	1
0	1	$0 \cdot 2 + 1 \cdot 4 + 5$	$F(x) \begin{cases} 1, & 9 \geq 0 \\ 0, & 9 < 0 \end{cases}$	1
0	0	$0 \cdot 2 + 0 \cdot 4 + 5$	$F(x) \begin{cases} 1, & 5 \geq 0 \\ 0, & 5 < 0 \end{cases}$	1
1	1	$1 \cdot 2 + 1 \cdot 4 + 5$	$F(x) \begin{cases} 1, & 11 \geq 0 \\ 0, & 11 < 0 \end{cases}$	1

La neurona - AND



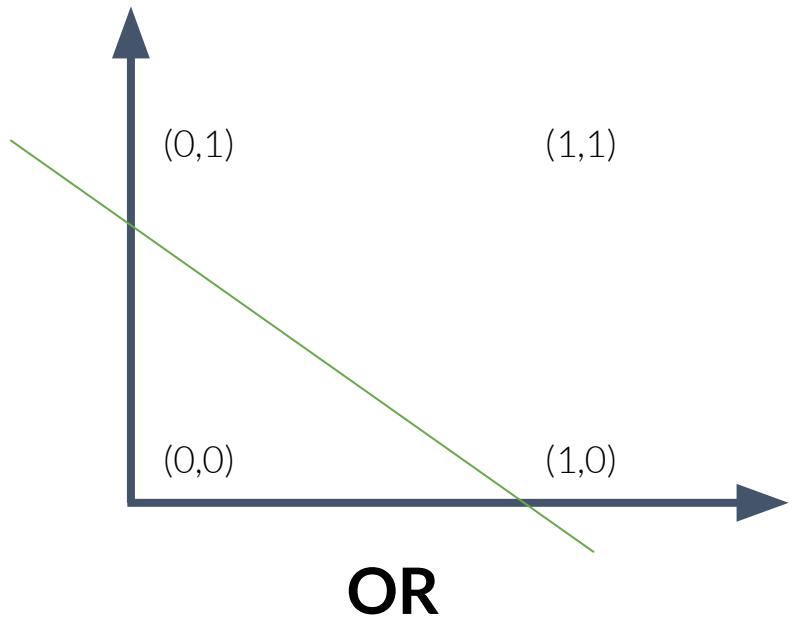
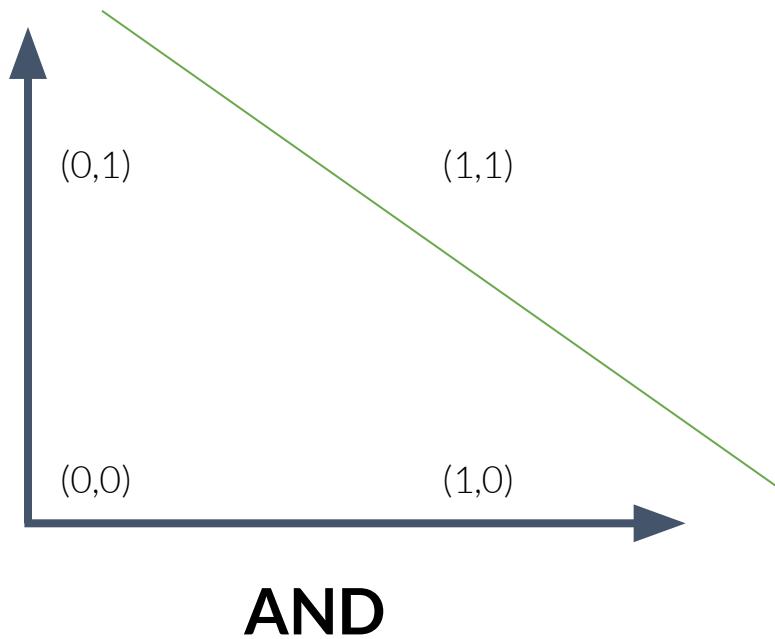
X_1	X_2	Σ	Activation	Salida
1	0	$1 \cdot 2 + 0 \cdot 1 - 3$	$F(x) \begin{cases} 1, & -1 \geq 0 \\ 0, & -1 < 0 \end{cases}$	0
0	1	$0 \cdot 2 + 1 \cdot 1 - 3$	$F(x) \begin{cases} 1, & -2 \geq 0 \\ 0, & -2 < 0 \end{cases}$	0
0	0	$0 \cdot 2 + 0 \cdot 1 - 3$	$F(x) \begin{cases} 1, & -3 \geq 0 \\ 0, & -3 < 0 \end{cases}$	0
1	1	$1 \cdot 2 + 1 \cdot 1 - 3$	$F(x) \begin{cases} 1, & 0 \geq 0 \\ 0, & 0 < 0 \end{cases}$	1

La neurona - OR

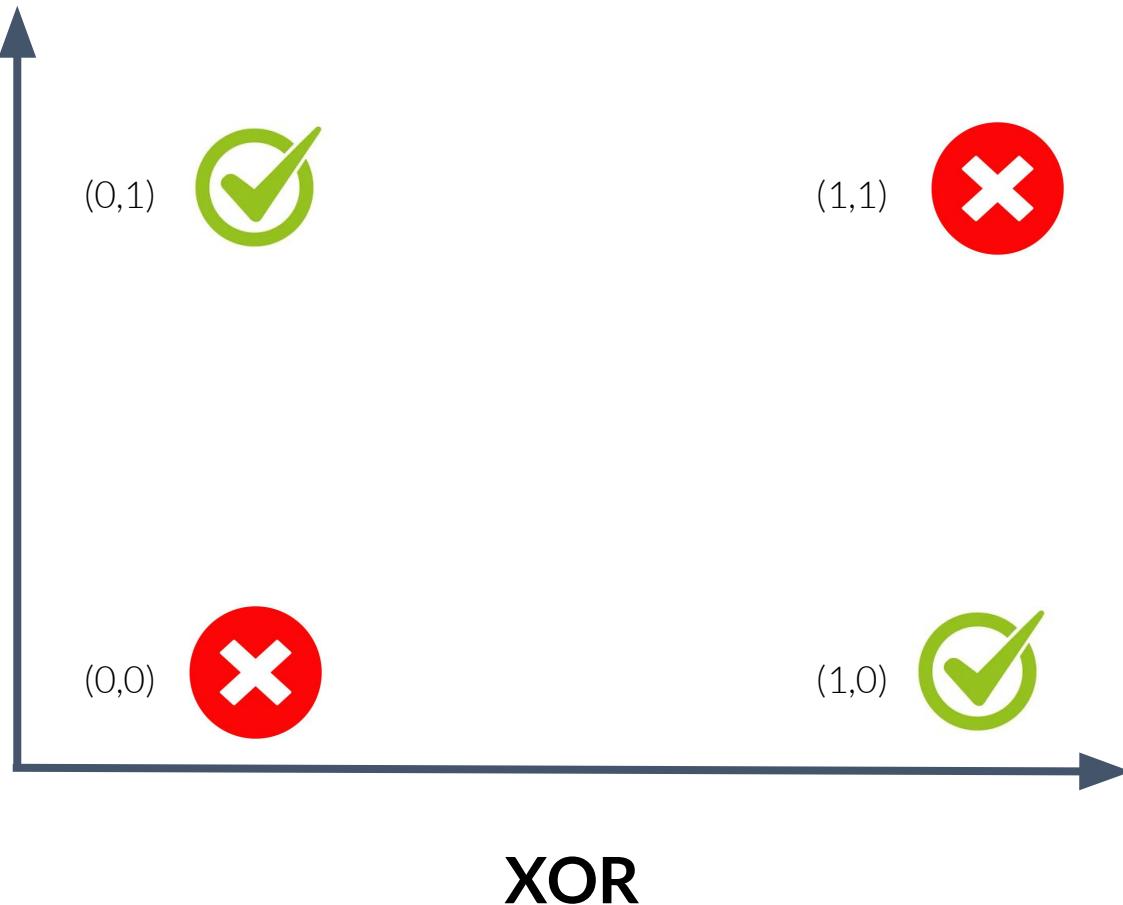


X1	X2	Σ	Activation	Salida
1	0	$1 \cdot 2 + 0 \cdot 2 - 2$	$F(x) \begin{cases} 1, & 0 \geq 0 \\ 0, & 0 < 0 \end{cases}$	1
0	1	$0 \cdot 2 + 1 \cdot 2 - 2$	$F(x) \begin{cases} 1, & 0 \geq 0 \\ 0, & 0 < 0 \end{cases}$	1
0	0	$0 \cdot 2 + 0 \cdot 2 - 2$	$F(x) \begin{cases} 1, & -2 \geq 0 \\ 0, & -2 < 0 \end{cases}$	0
1	1	$1 \cdot 2 + 1 \cdot 2 - 2$	$F(x) \begin{cases} 1, & 2 \geq 0 \\ 0, & 2 < 0 \end{cases}$	1

La neurona



La neurona - XOR



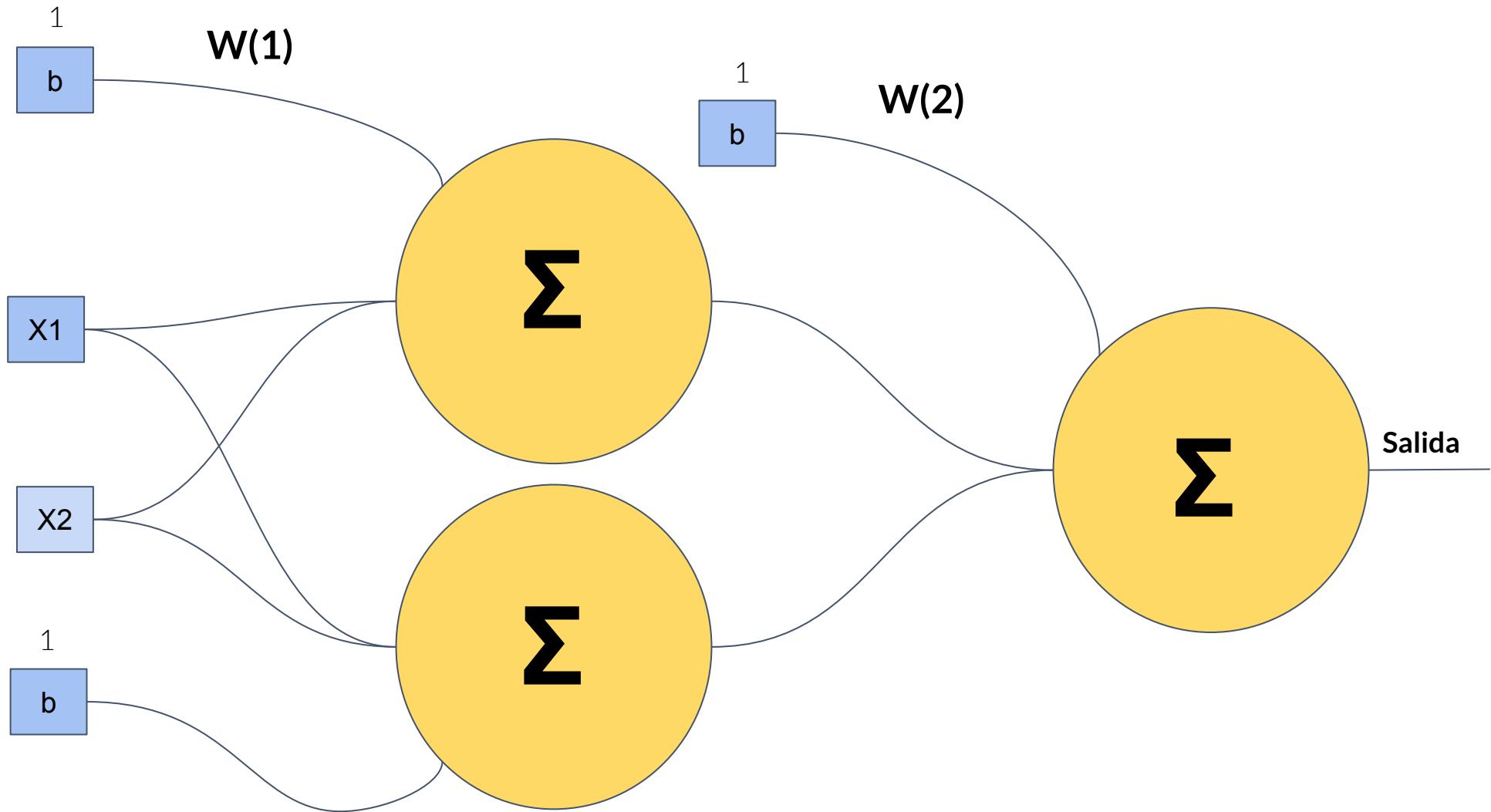
La neurona - XOR



La neurona - XOR

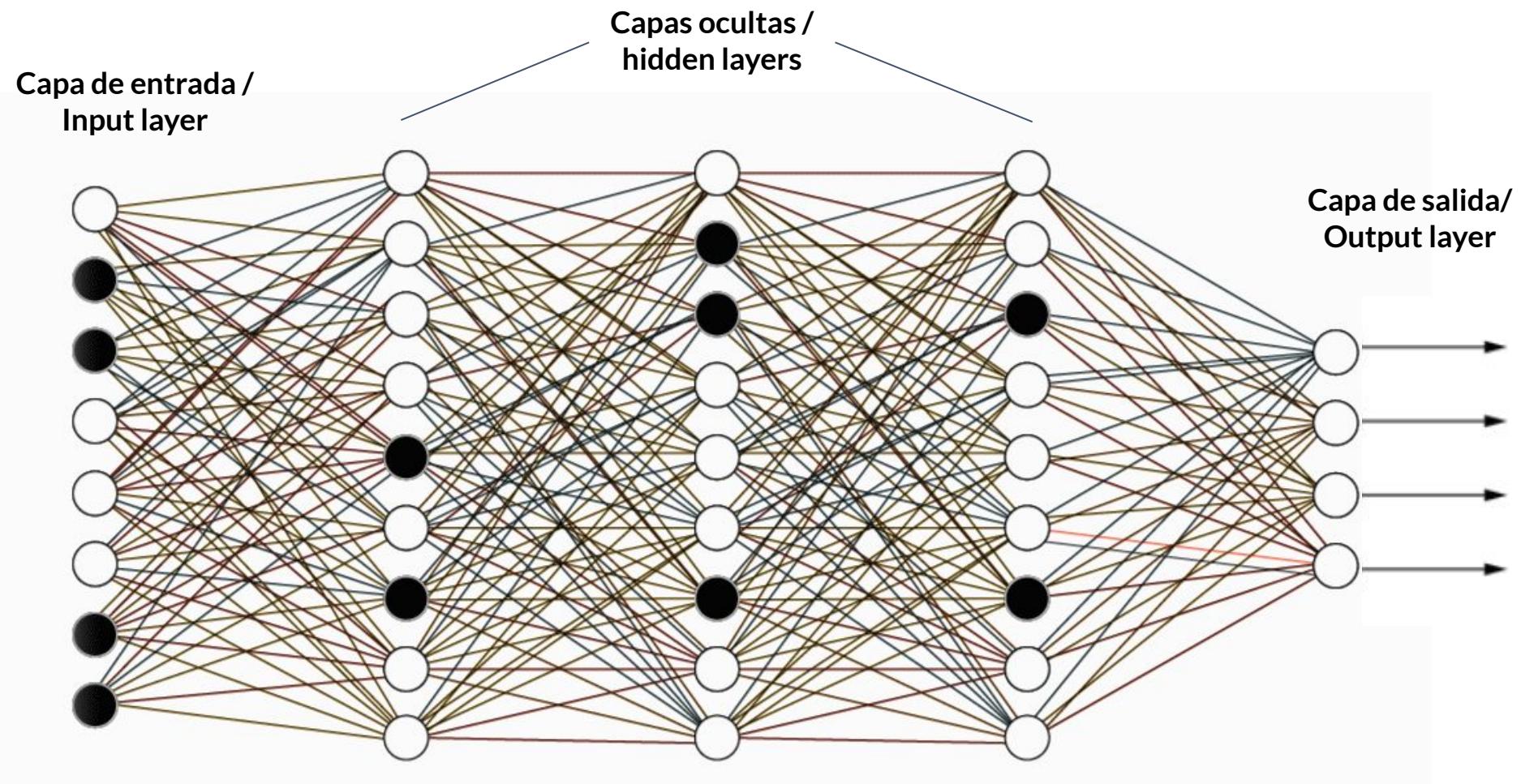


La neurona - XOR

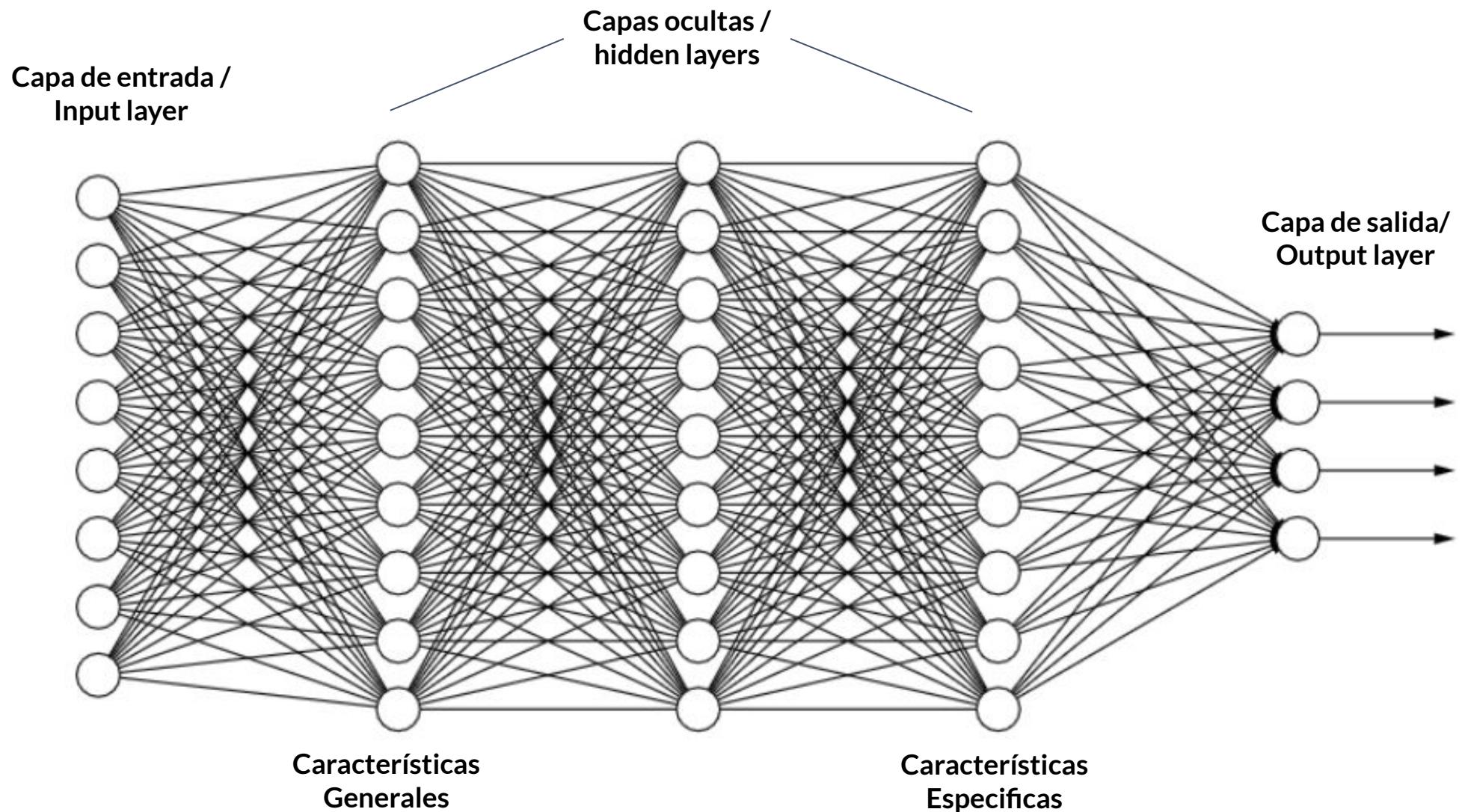


Arquitectura de una red neuronal

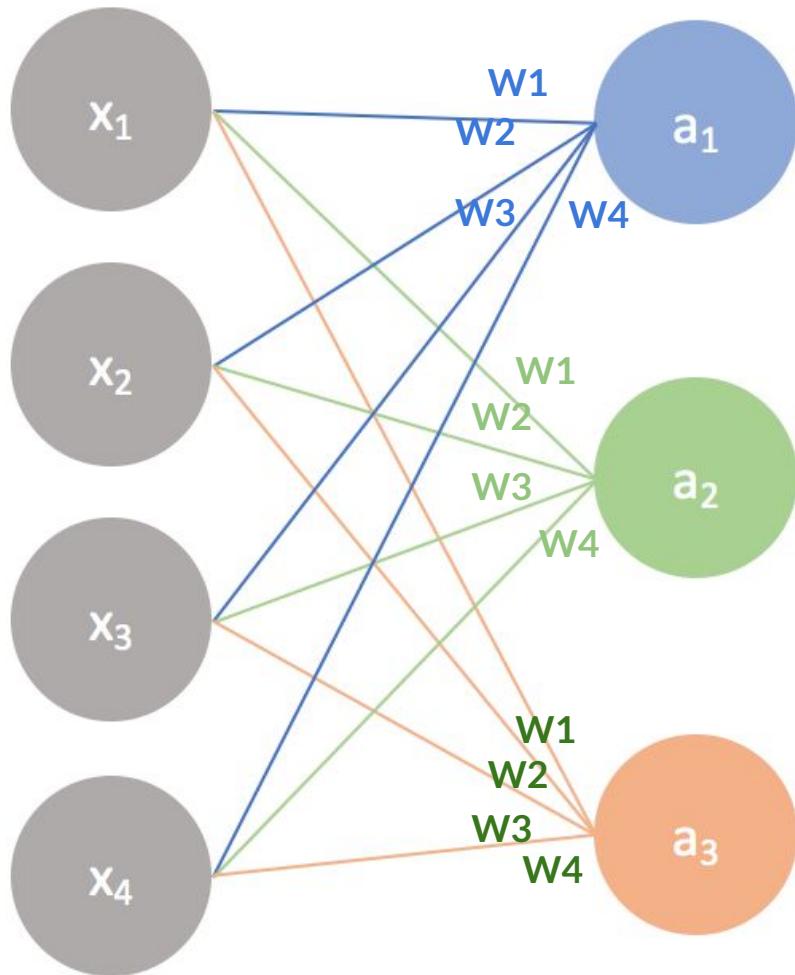
Redes neuronales - Capas



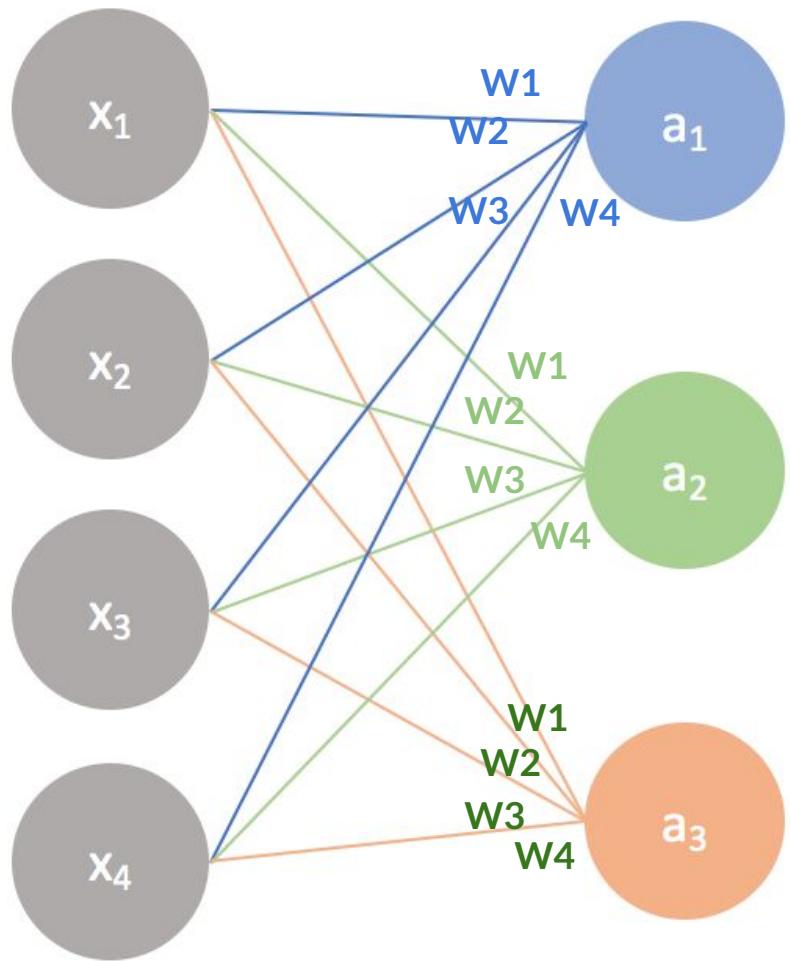
Redes neuronales - Capas



Redes neuronales - Vectores



■ Redes neuronales - Vectores



$$\begin{bmatrix} w_1, w_2, w_3, w_4 \\ w_1, w_2, w_3, w_4 \\ w_1, w_2, w_3, w_4 \end{bmatrix} \cdot \begin{bmatrix} x_1, \\ x_2, \\ x_3, \\ x_4 \end{bmatrix}$$

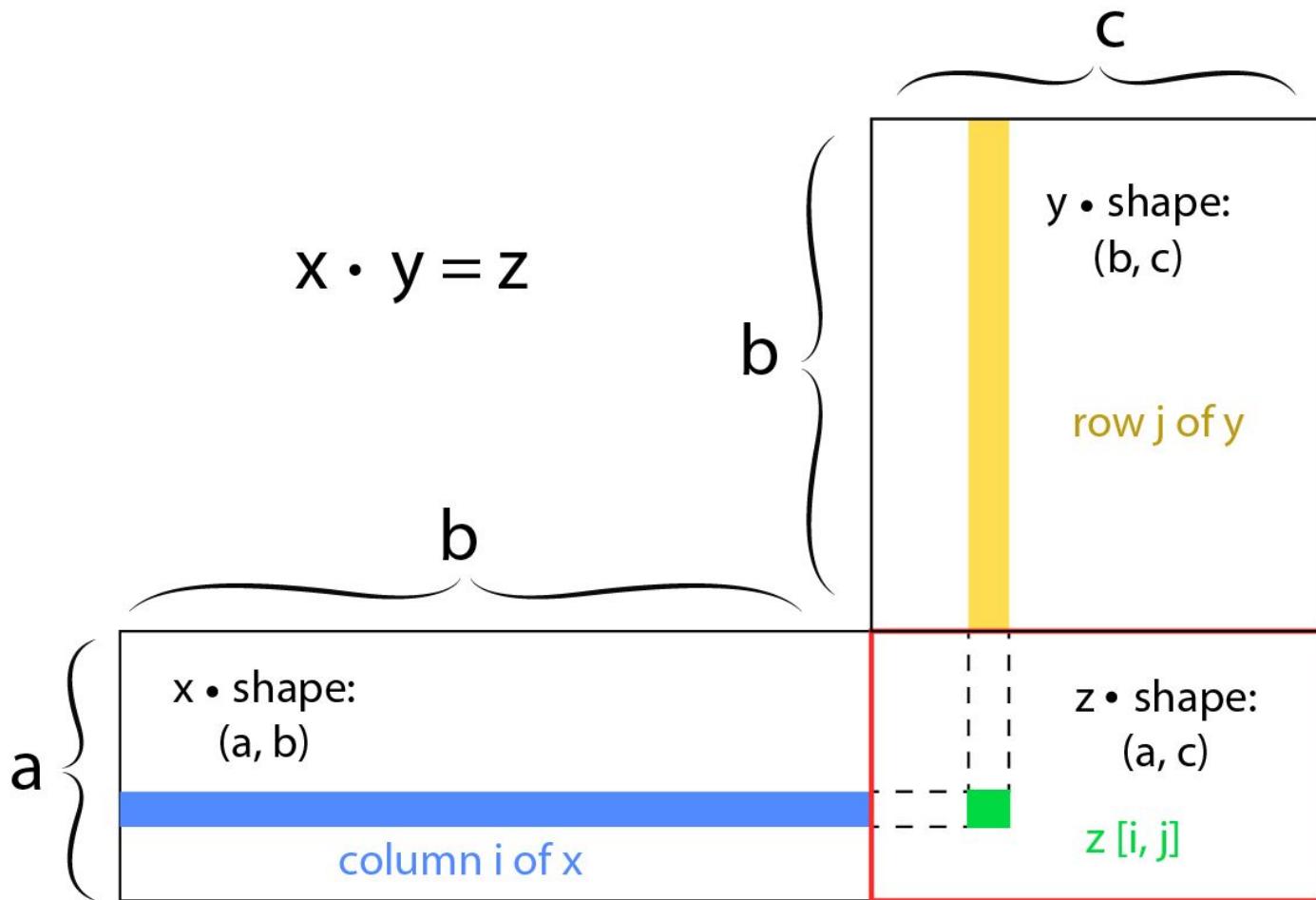
■ Redes neuronales - Vectores

$$\begin{bmatrix} w_1, w_2, w_3, w_4 \\ w_1, w_2, w_3, w_4 \\ w_1, w_2, w_3, w_4 \end{bmatrix} \bullet \begin{bmatrix} x_1, \\ x_2, \\ x_3, \\ x_4 \end{bmatrix} = \begin{bmatrix} w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 \end{bmatrix}$$

■ Redes neuronales - Vectores

$$\begin{bmatrix} w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Redes neuronales - Matrices



Redes neuronales - Matrices

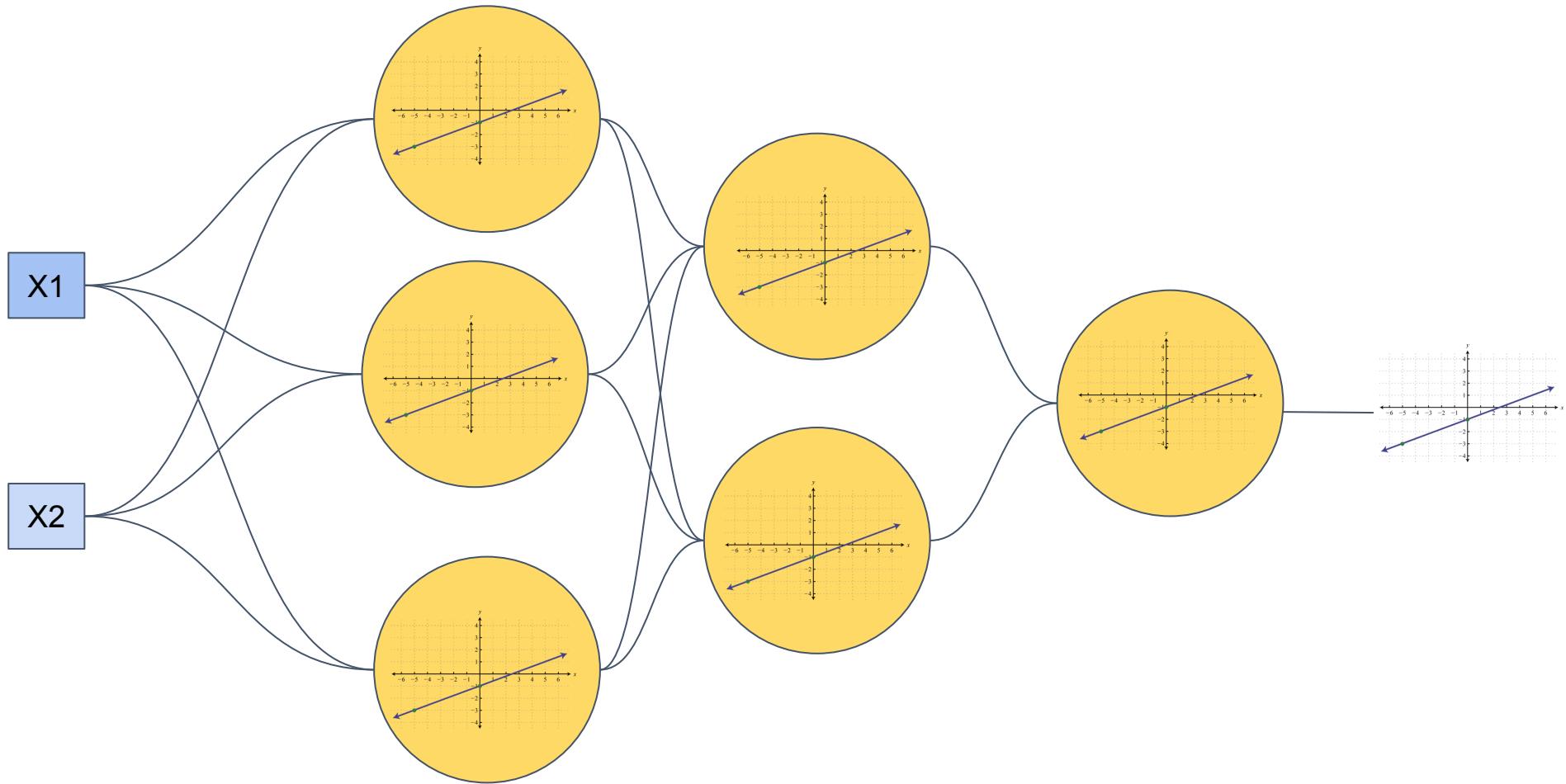
$$\begin{matrix} & (3,4) \\ \left[\begin{matrix} 3,2,4,5 \\ 3,4,1,9 \\ 6,9,2,1 \end{matrix} \right] & \bullet & \left[\begin{matrix} 5,2,6 \\ 7,3,1 \\ 8,5,1 \\ 7,2,8 \end{matrix} \right] & (4,3) \\ & = \left[\begin{matrix} 96,42,64 \\ 114,41,95 \\ 116,51,55 \end{matrix} \right] \end{matrix}$$

$$3*5 + 2*7 + 4*8 + 5*7 = 96$$

$$3*2 + 2*3 + 4*5 + 5*2 = 42$$

$$3*6 + 2*1 + 4*1 + 5*8 = 64$$

Redes neuronales

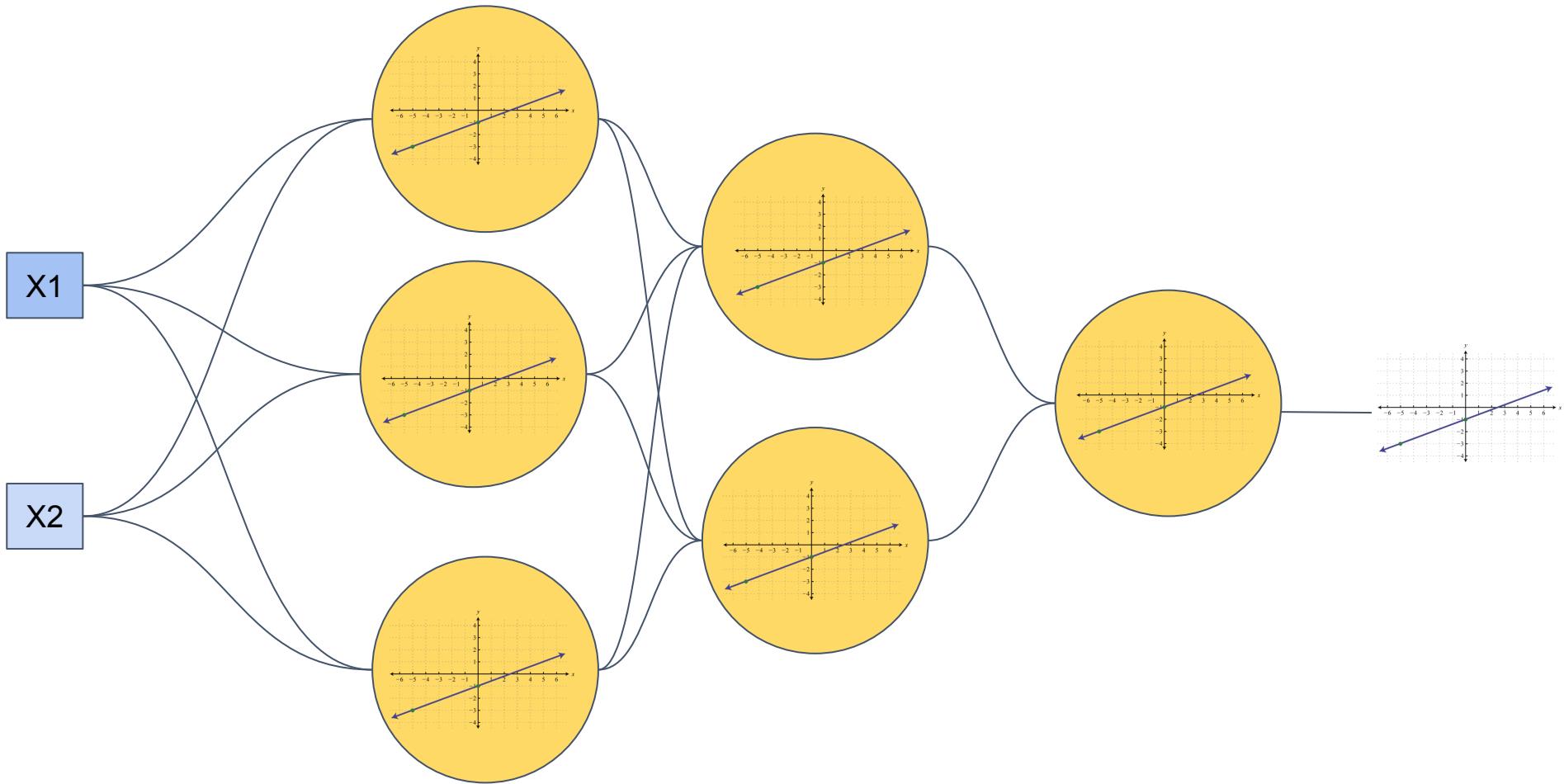


Redes neuronales

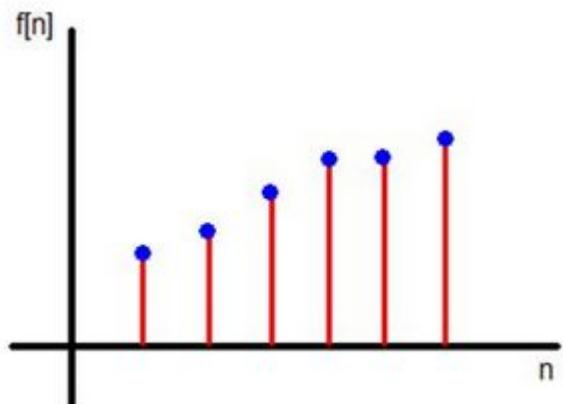


Funciones de activación

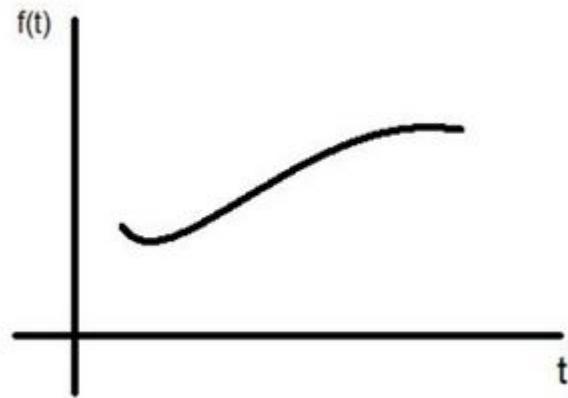
Redes neuronales



Funciones de activación

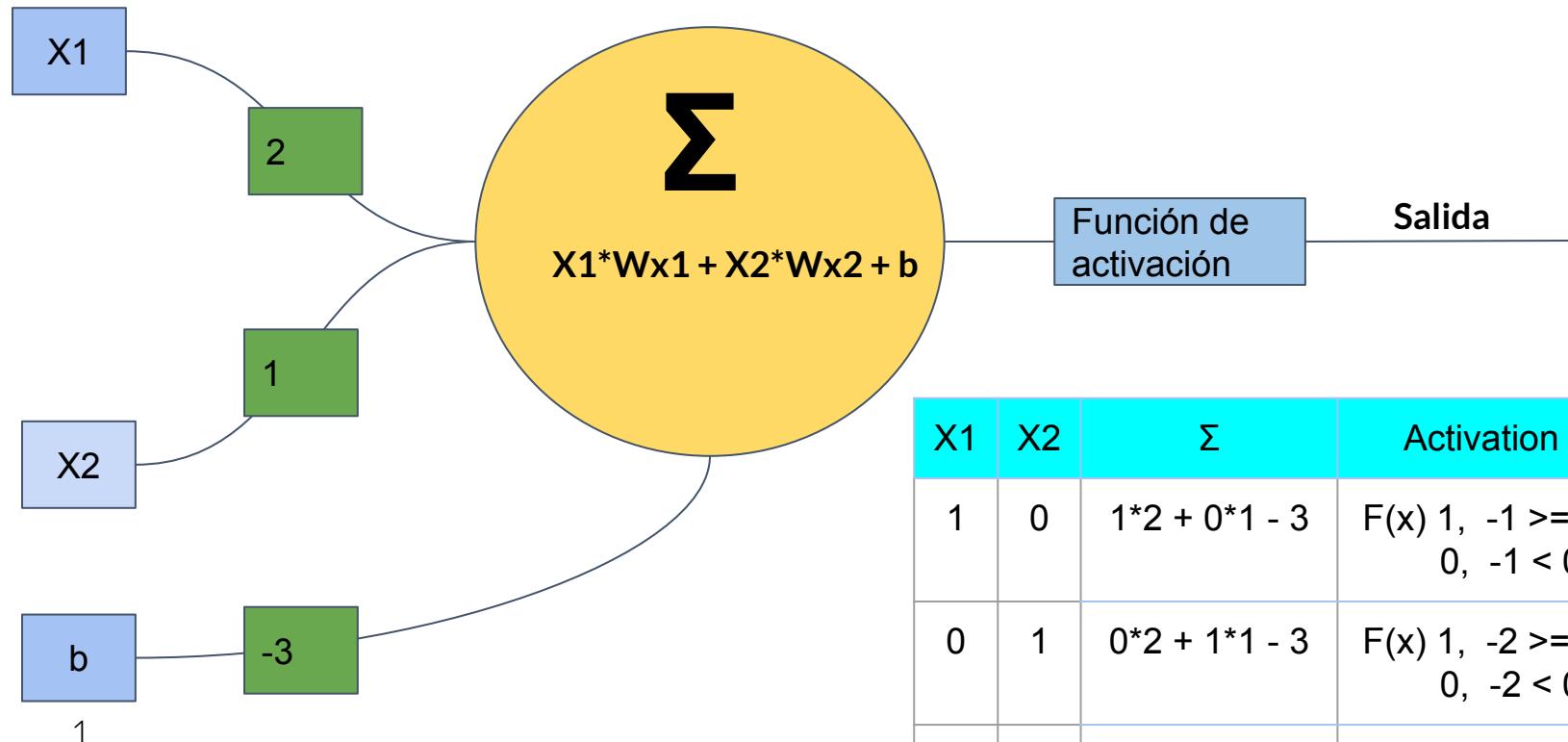


Discretas
Conjunto finito de
valores



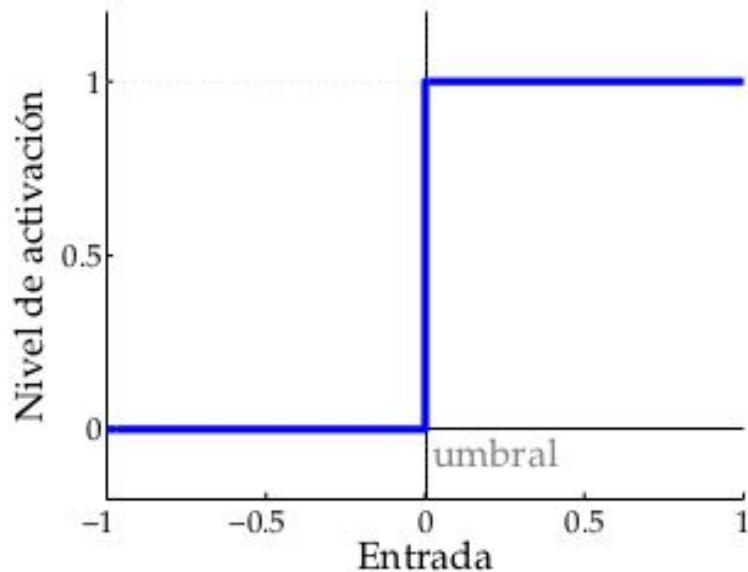
Continuas
Valores dentro de
algun intervalo

La neurona - AND



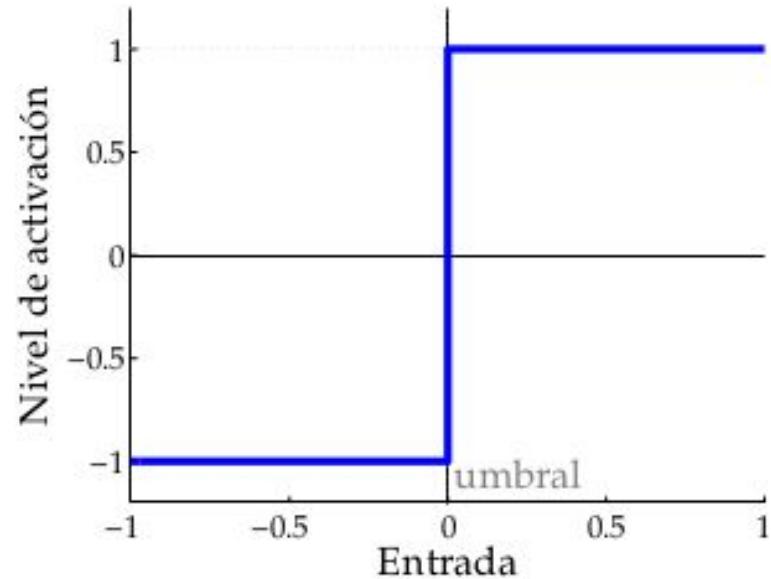
X_1	X_2	Σ	Activation	Salida
1	0	$1 \cdot 2 + 0 \cdot 1 - 3$	$F(x) \begin{cases} 1, & -1 \geq 0 \\ 0, & -1 < 0 \end{cases}$	0
0	1	$0 \cdot 2 + 1 \cdot 1 - 3$	$F(x) \begin{cases} 1, & -2 \geq 0 \\ 0, & -2 < 0 \end{cases}$	0
0	0	$0 \cdot 2 + 0 \cdot 1 - 3$	$F(x) \begin{cases} 1, & -3 \geq 0 \\ 0, & -3 < 0 \end{cases}$	0
1	1	$1 \cdot 2 + 1 \cdot 1 - 3$	$F(x) \begin{cases} 1, & 0 \geq 0 \\ 0, & 0 < 0 \end{cases}$	1

Función escalón/escalonada/threshold



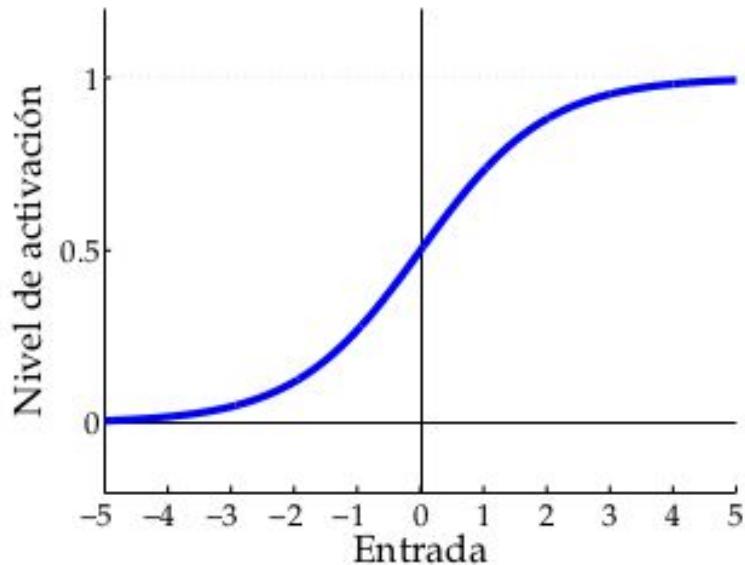
$$y = f_{tlu}(z) = u(z) = \mathbf{1}_{z \geq 0} = \begin{cases} 1 & \text{si } z \geq 0 \\ 0 & \text{si } z < 0 \end{cases}$$

Función signo/signum



$$y = f_{sgn}(z) = sgn(z) = \begin{cases} 1 & \text{si } z \geq 0 \\ -1 & \text{si } z < 0 \end{cases}$$

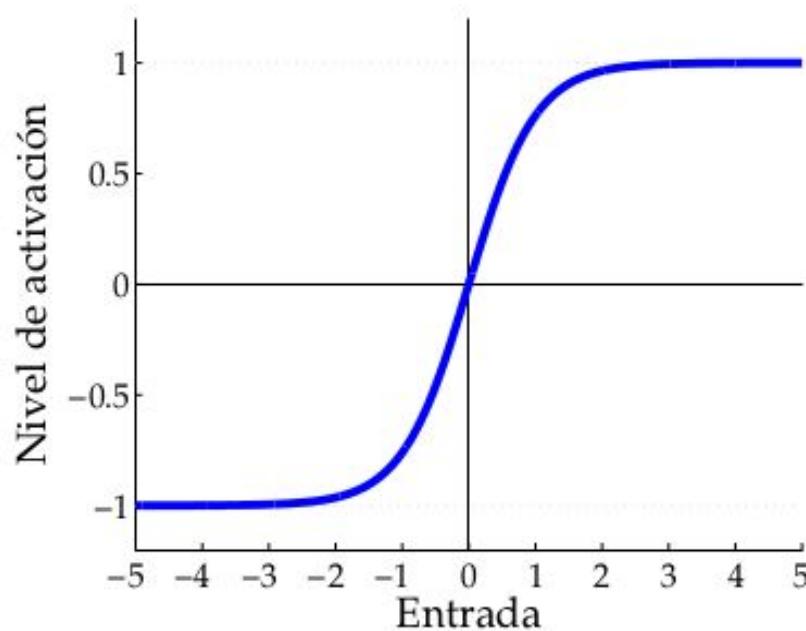
Función sigmoidal/sigmoid



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

$$\frac{d}{dx} \left(\frac{1}{1+\exp(-x)} \right) = \frac{e^{-x}}{(e^{-x}+1)^2}$$

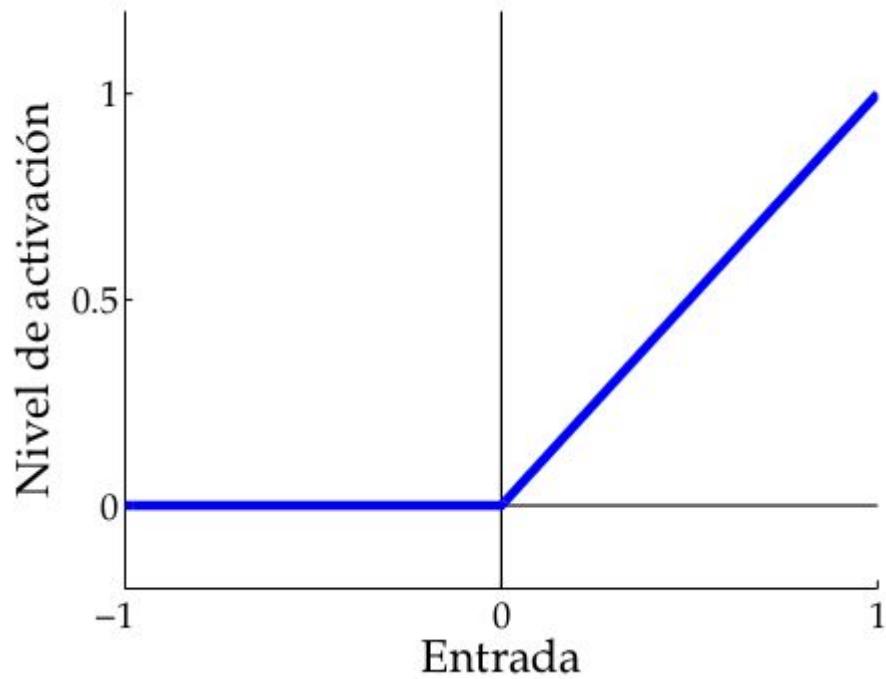
Función tangente hiperbólica/tanh



$$y = f_{\text{logistic}}(z) = \sigma(z) = \frac{1}{1 + e^{-z}}$$

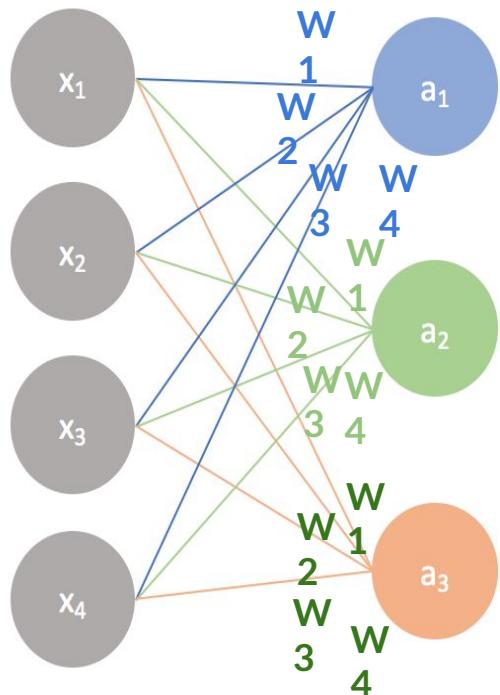
$$\frac{d}{dx} \left(\frac{1}{1+\exp(-x)} \right) = \frac{e^{-x}}{(e^{-x}+1)^2}$$

Función lineal rectificada/ReLU



$$y = f_{relu}(z) = \begin{cases} z & \text{si } z \leq 0 \\ 0 & \text{si } z < 0 \end{cases} \quad \frac{d}{dx} f_{relu}(z) = u(z) = \mathbf{1}_{z \geq 0} = \begin{cases} 1 & \text{si } z \geq 0 \\ 0 & \text{si } z < 0 \end{cases}$$

Softmax

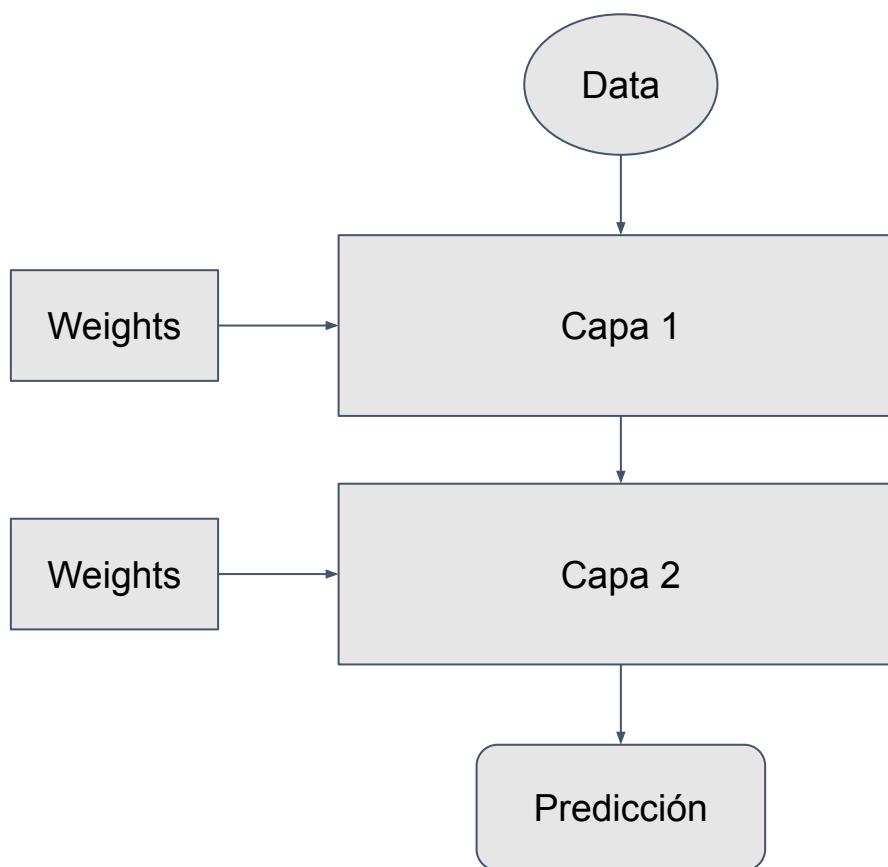


$$\mathbf{z} = \begin{pmatrix} 1.6 \\ 0.55 \\ 0.98 \end{pmatrix} = \begin{pmatrix} 0.51 \\ 0.18 \\ 0.31 \end{pmatrix}$$

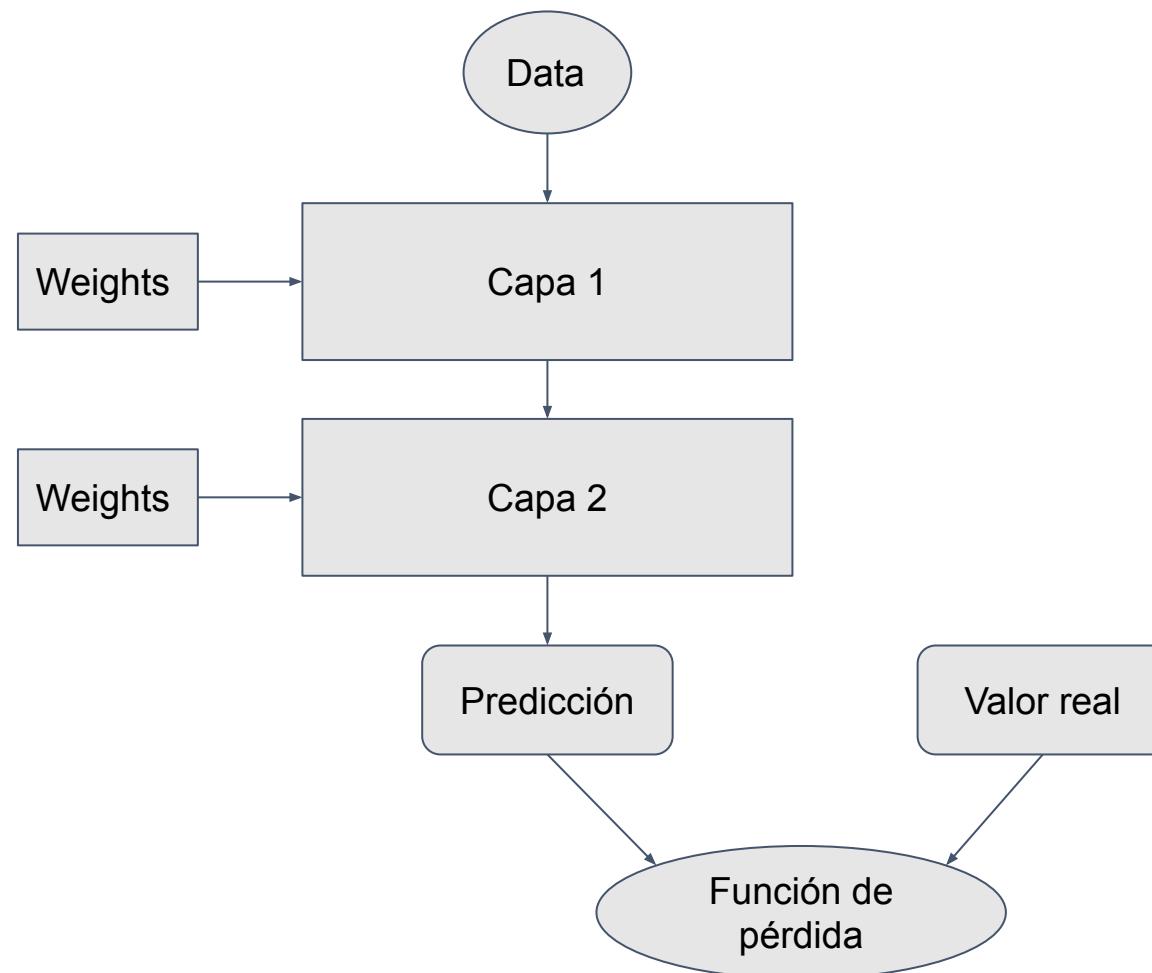
$$y_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}$$

Función de pérdida

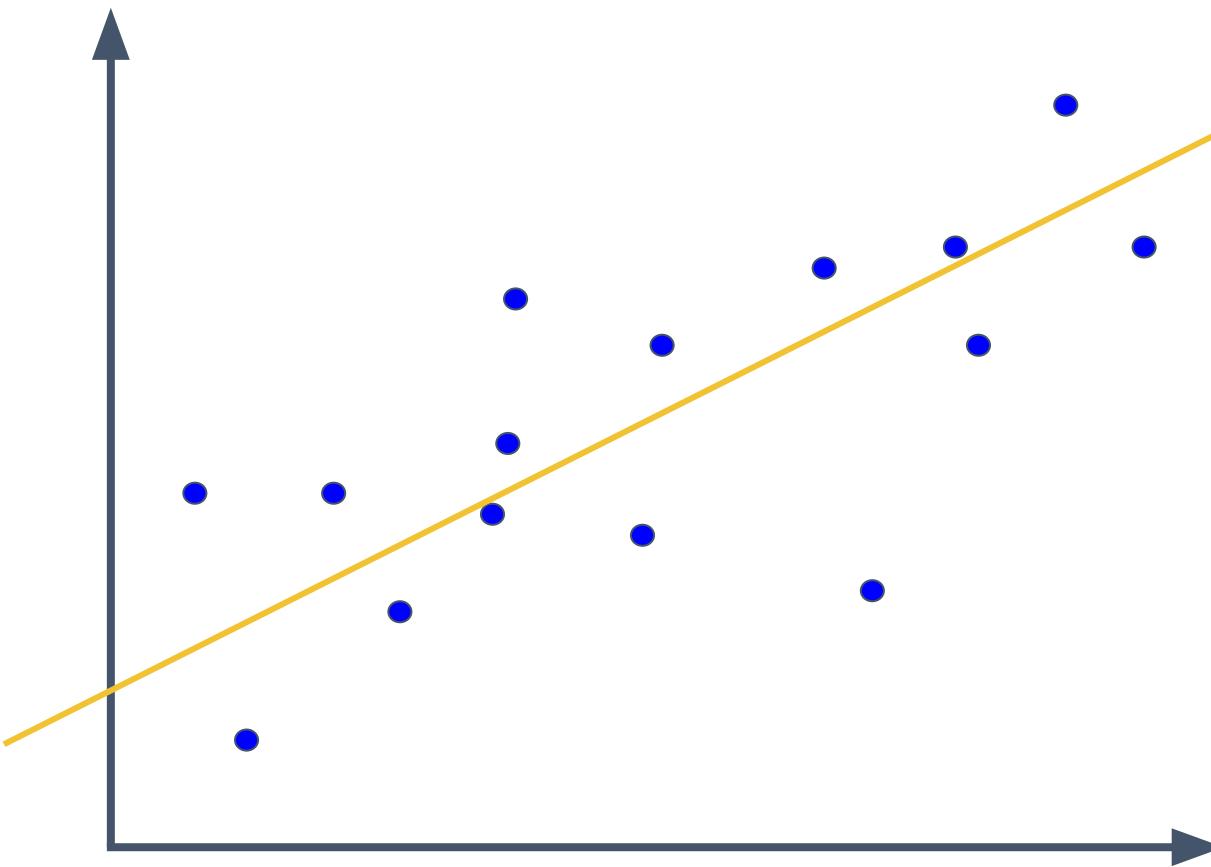
Función de pérdida



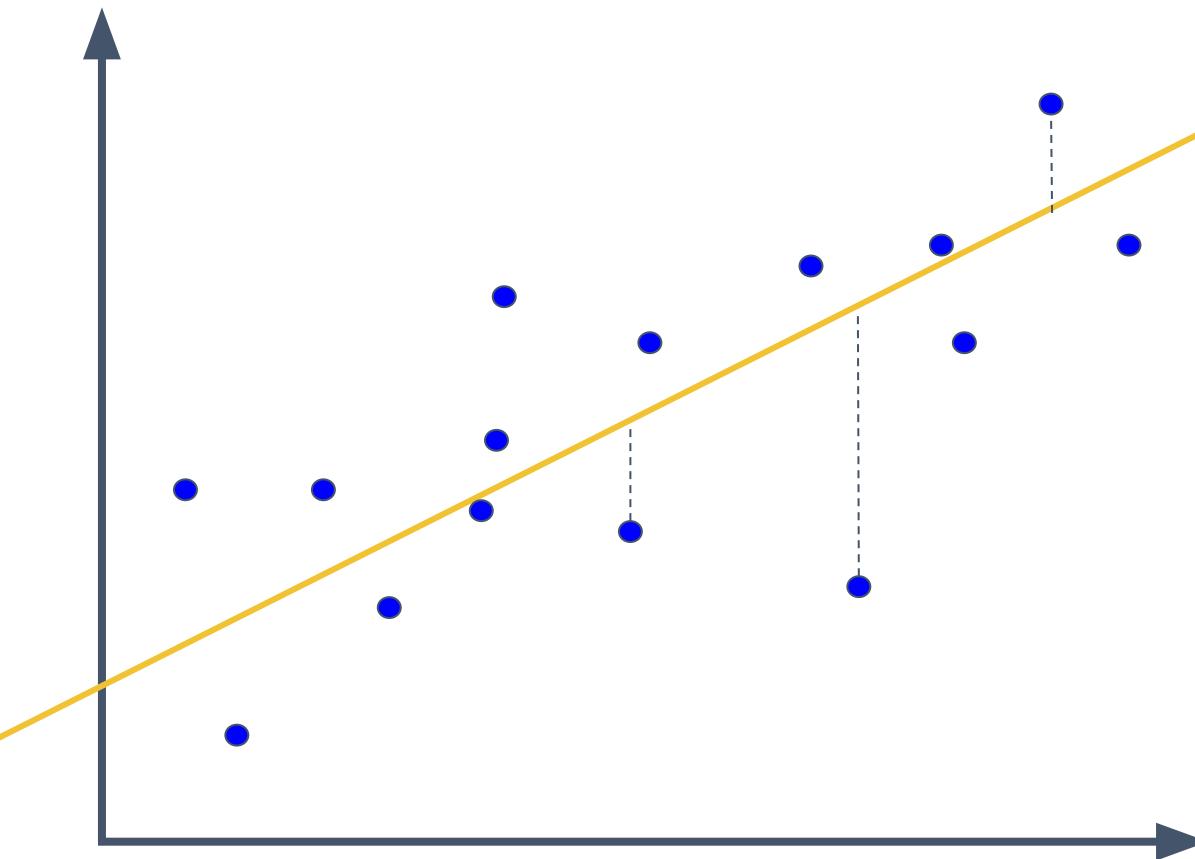
Función de pérdida



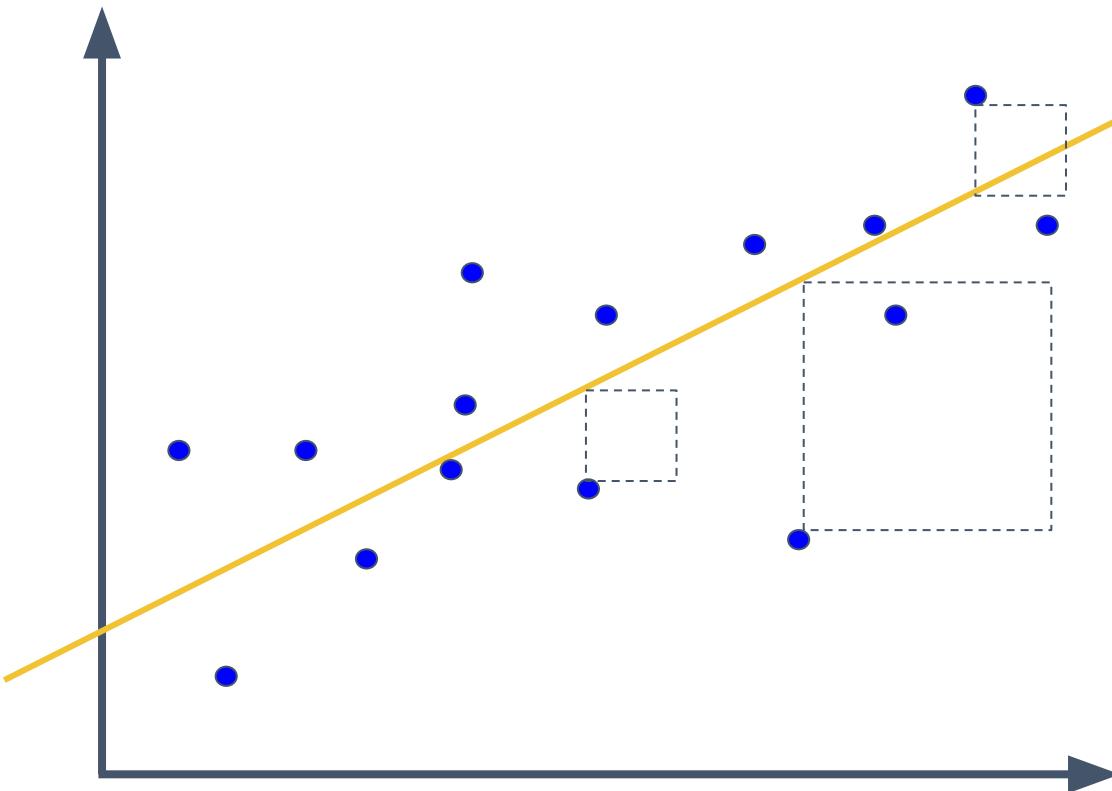
MSE - Mean squared error



MSE - Mean squared error



MSE - Mean squared error



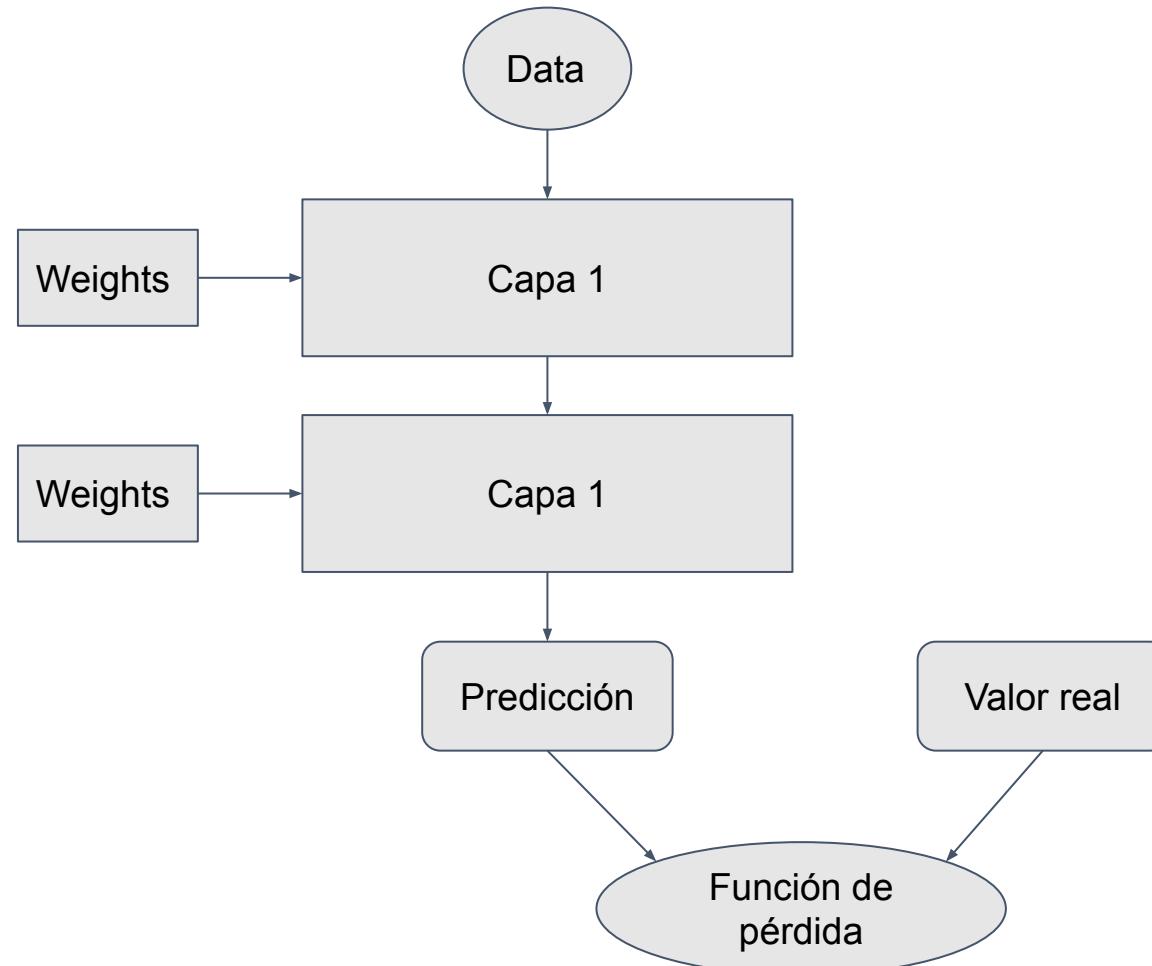
$$MSE = \frac{1}{n} \sum \underbrace{\left(y - \hat{y} \right)^2}_{\text{Error}}$$

Cross Entropy

Verdad	Predicción	
	$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0.5 \\ 0.3 \\ 0.2 \end{pmatrix}$
		\equiv
		$\begin{matrix} \text{Log}(0.5) = -0,301029996 \\ p(x_0) = 1 \end{matrix}$
		\equiv
		0,301029996

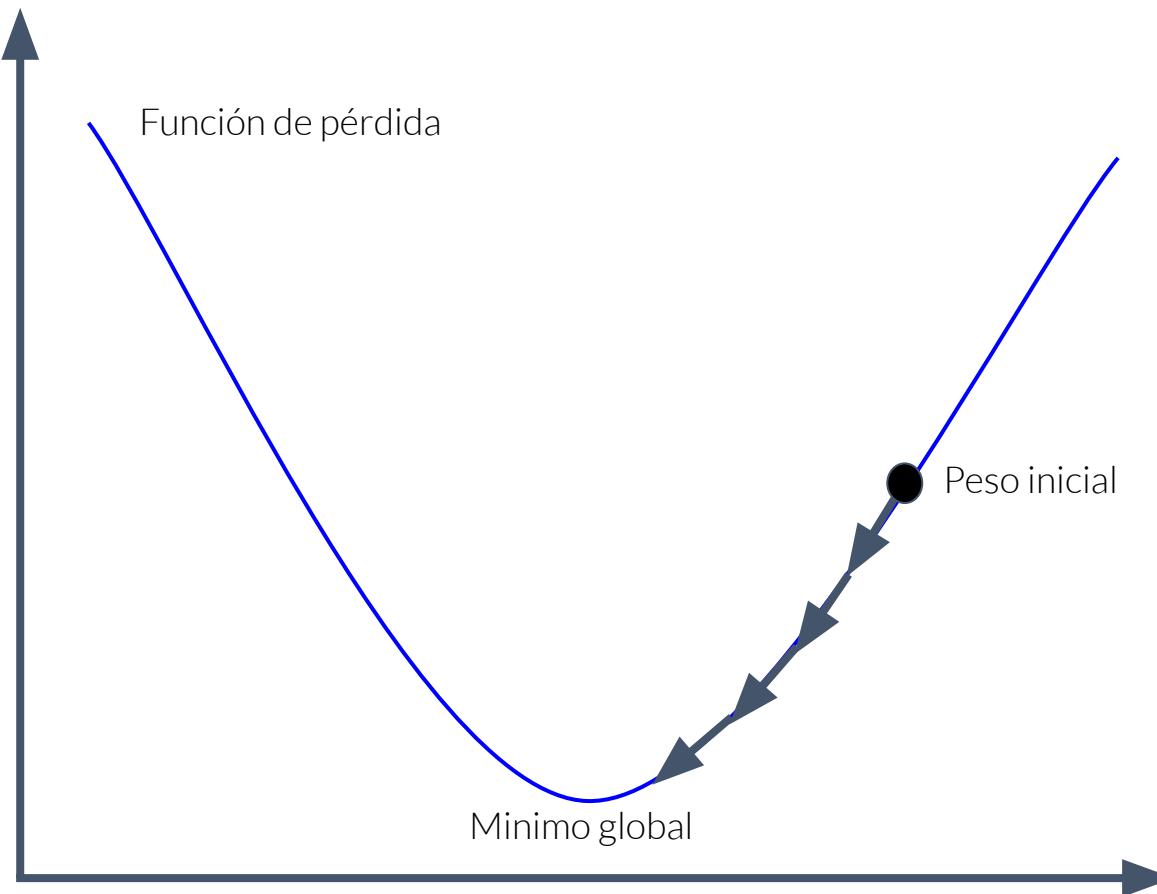
$$-\sum_x p(x) \log q(x).$$

Función de pérdida

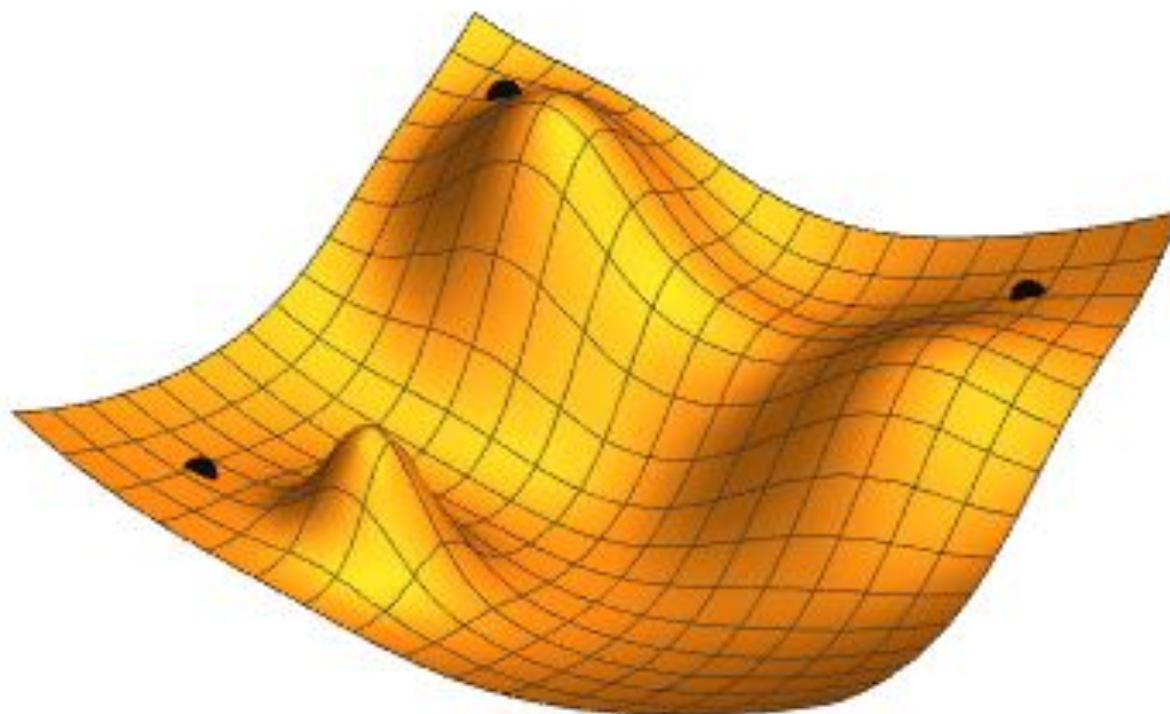


Descenso del gradiente

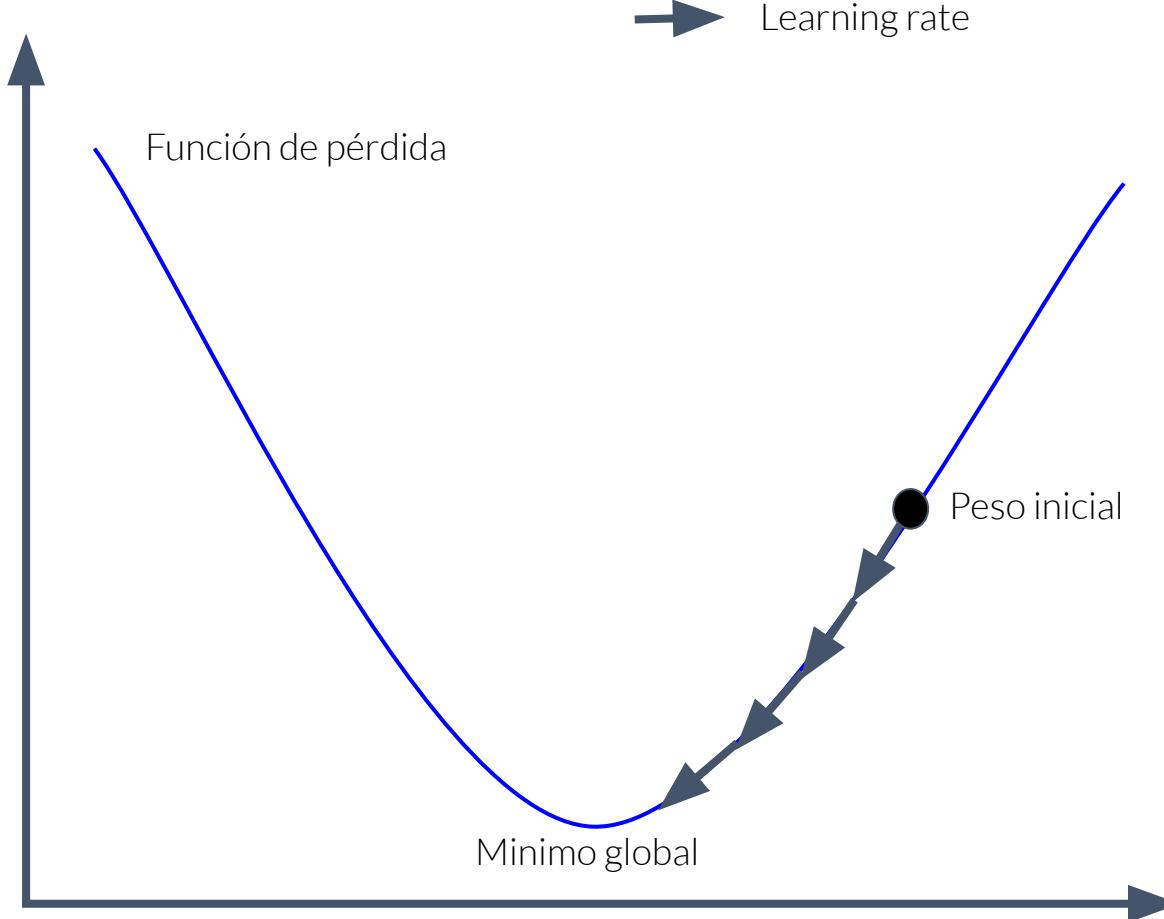
Descenso del gradiente



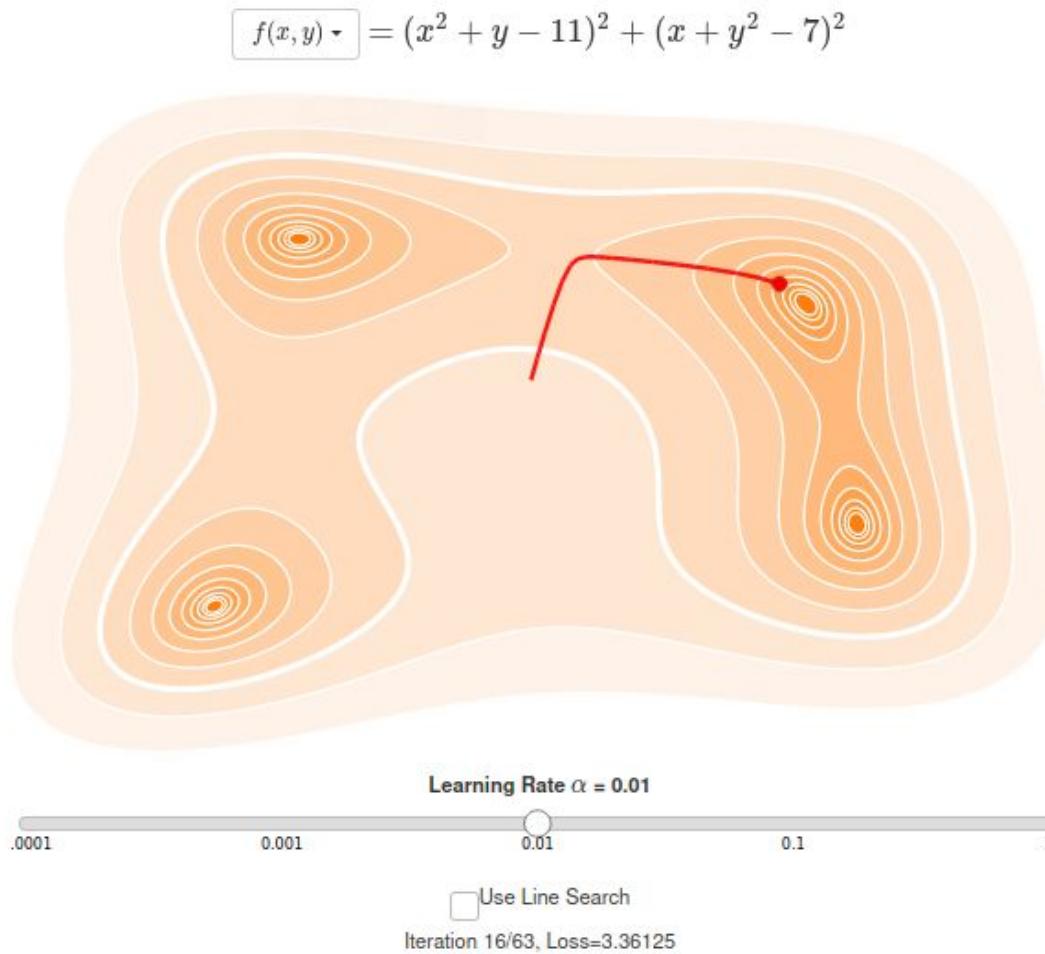
■ Descenso del gradiente



Descenso del gradiente



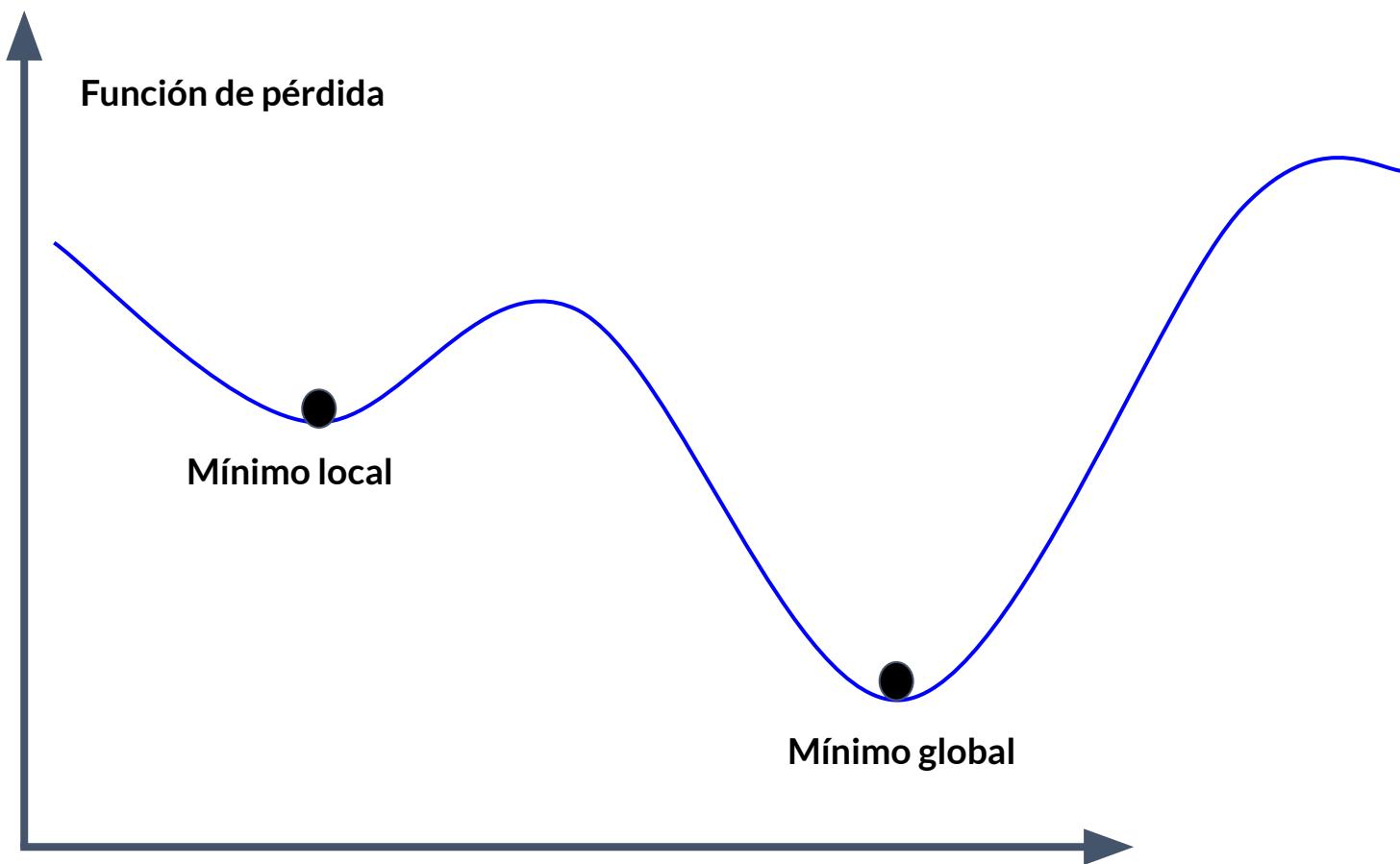
Descenso del gradiente - Learning rate



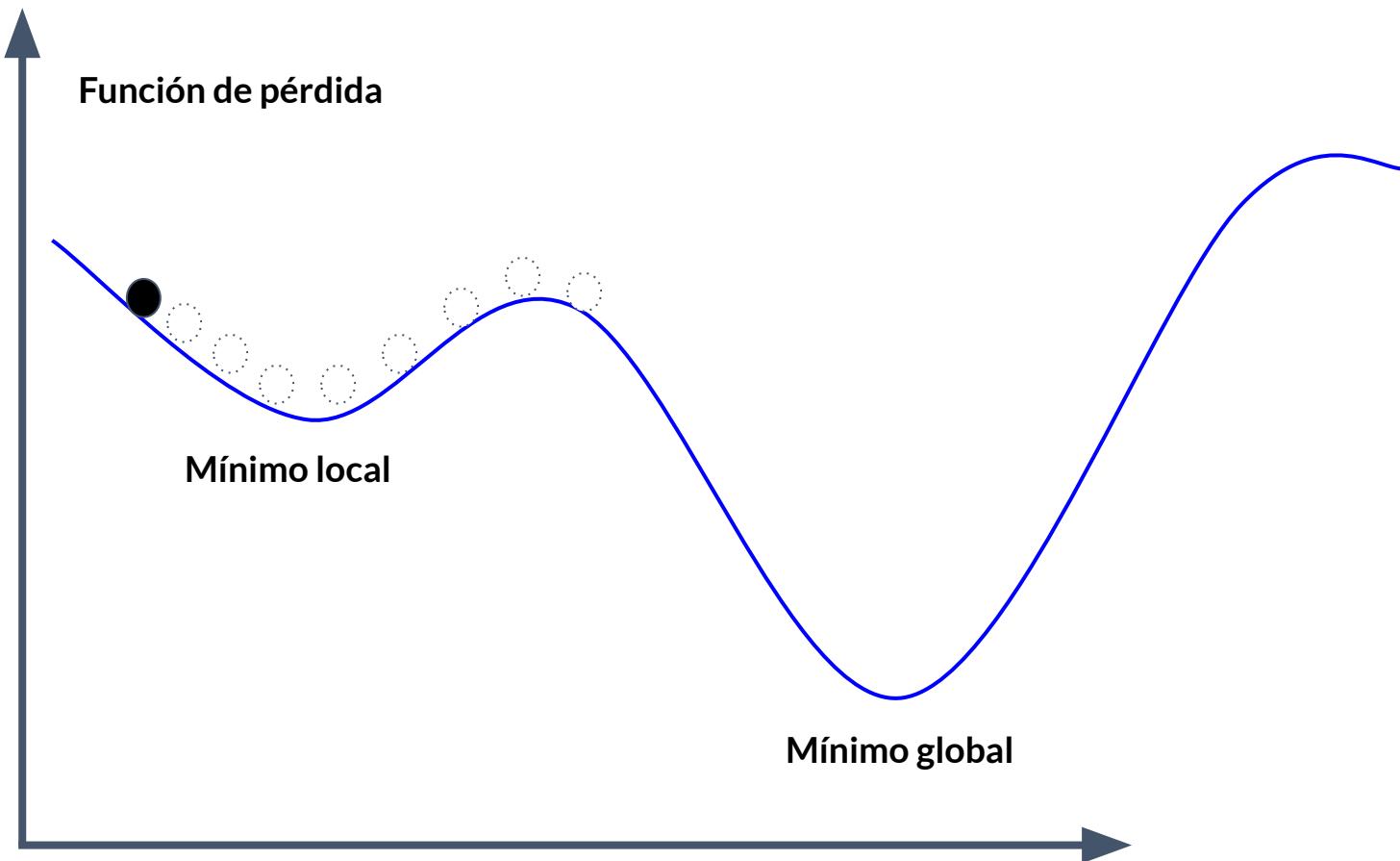
■ Descenso del gradiente

A photograph of a misty forest landscape. In the foreground, dark evergreen trees are visible. Behind them, several layers of mountains are shrouded in thick, white fog. The sky above the mountains is a pale, hazy blue. The overall atmosphere is mysterious and serene.

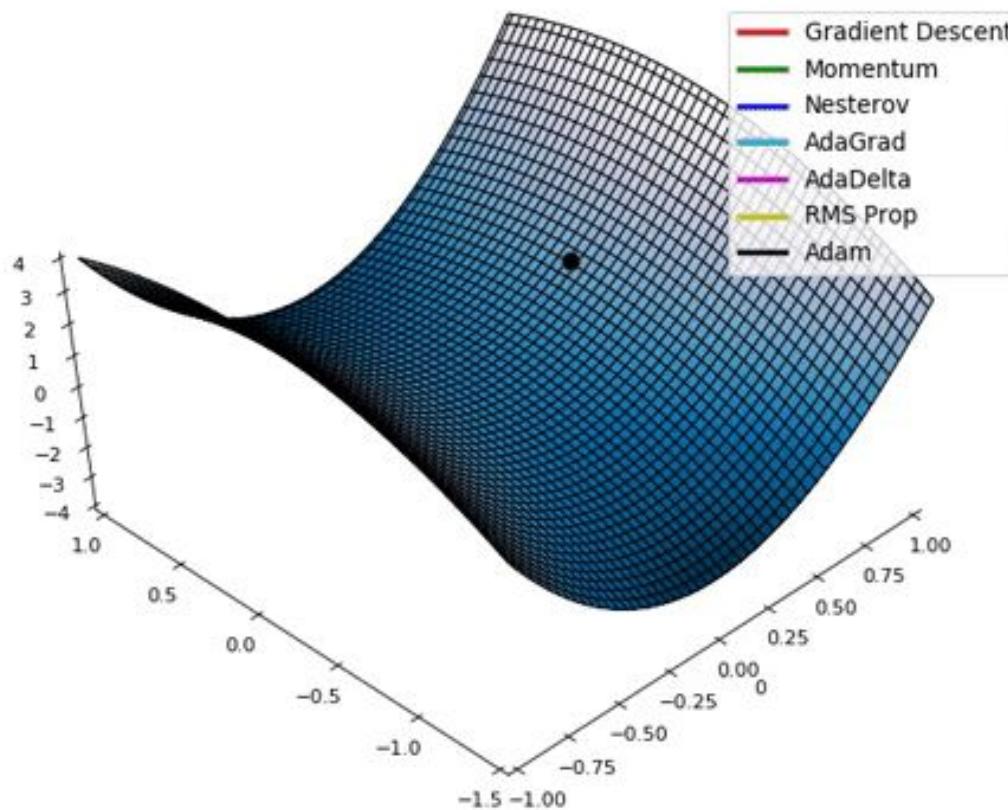
Descenso del gradiente - Mínimos



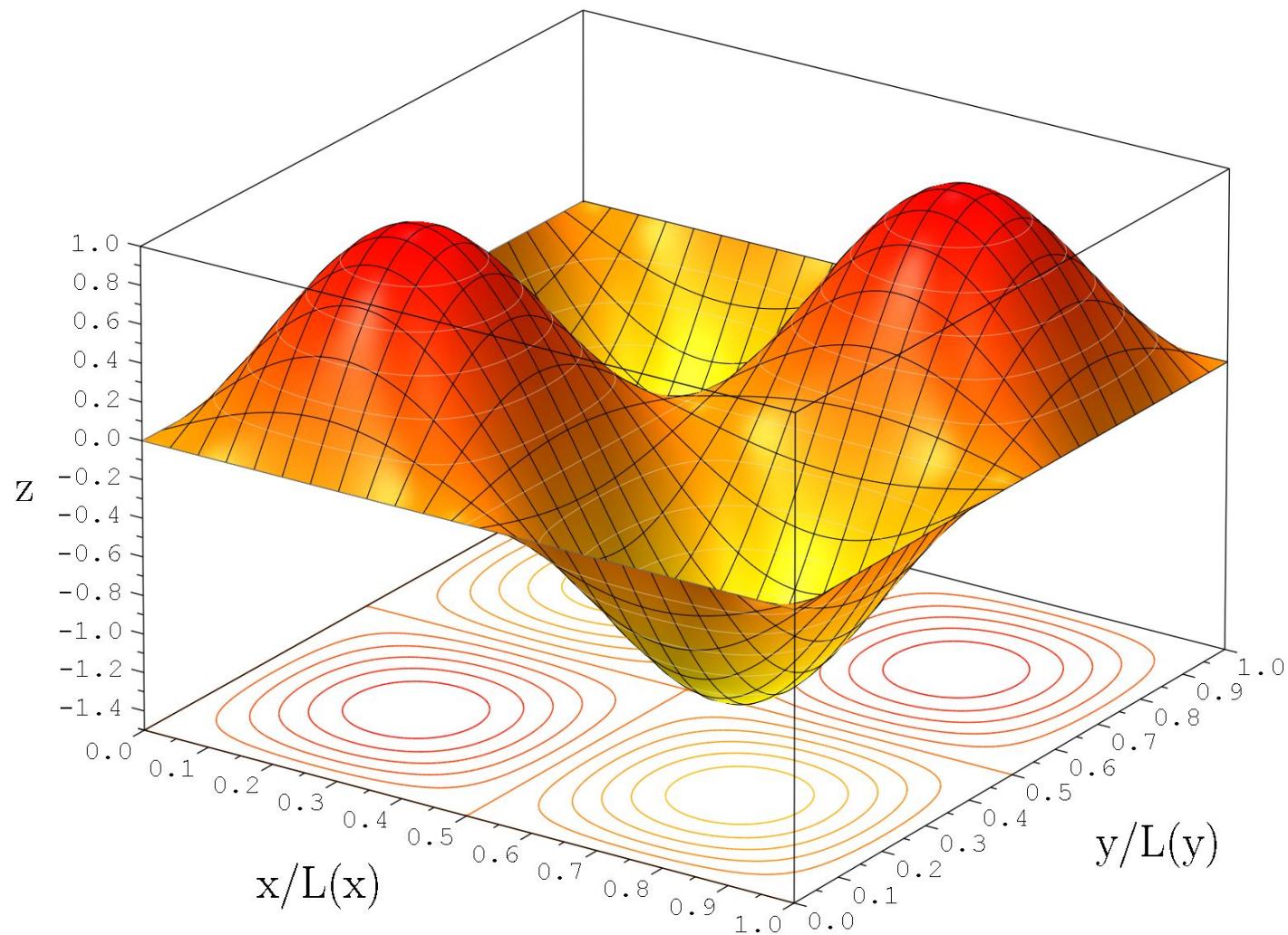
Momentum



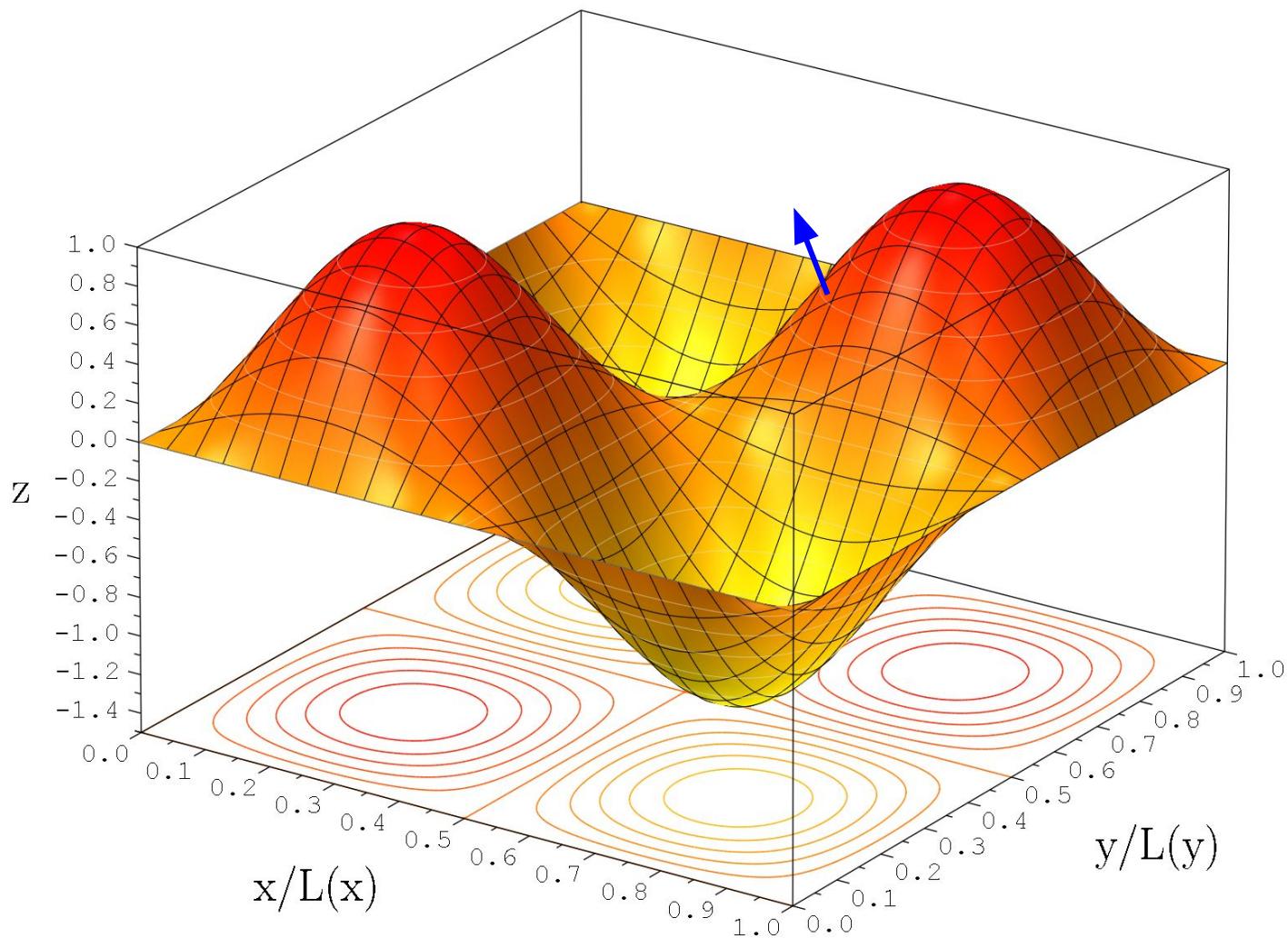
Descenso del gradiente - Optimizadores



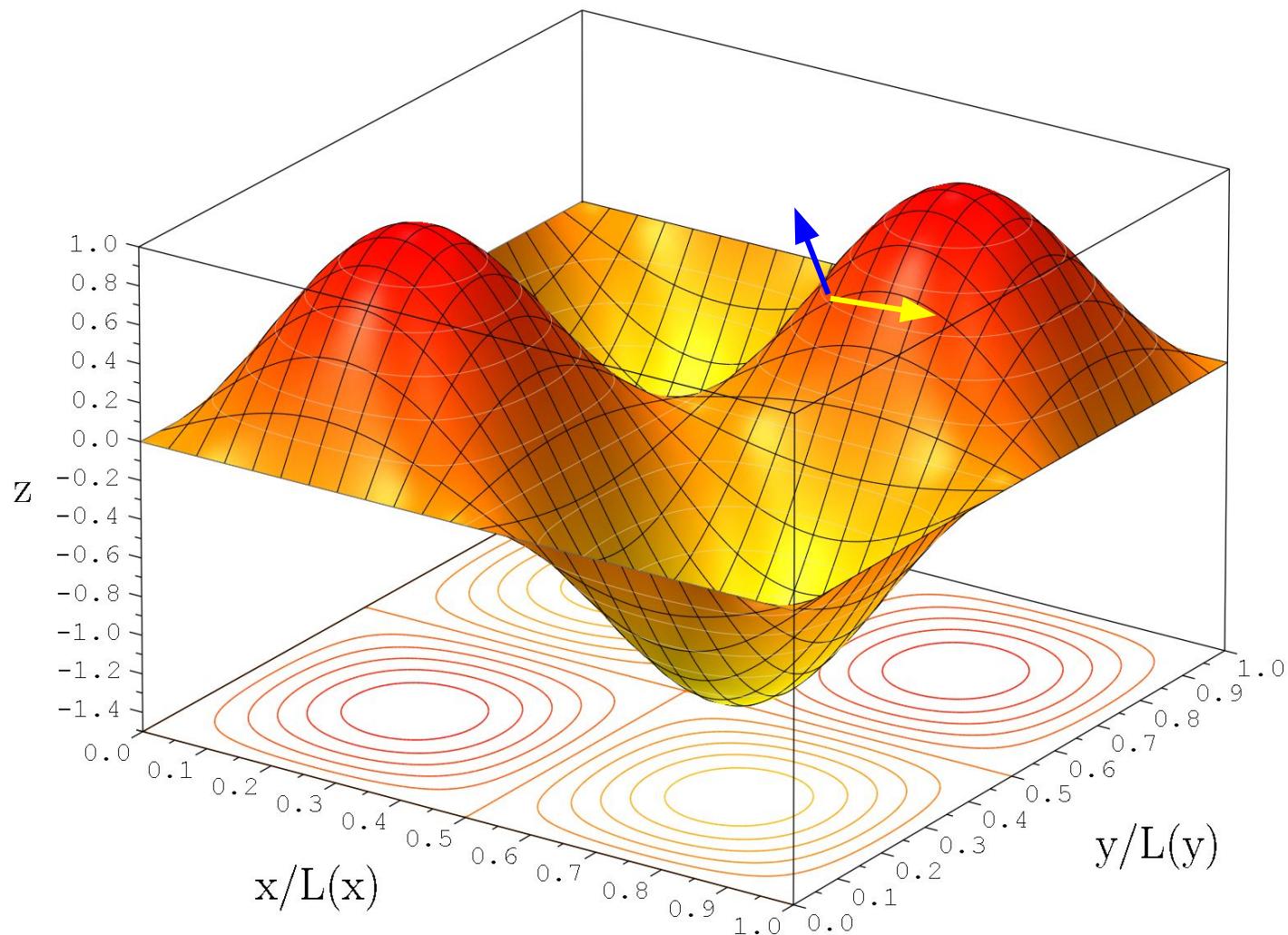
Descenso del gradiente



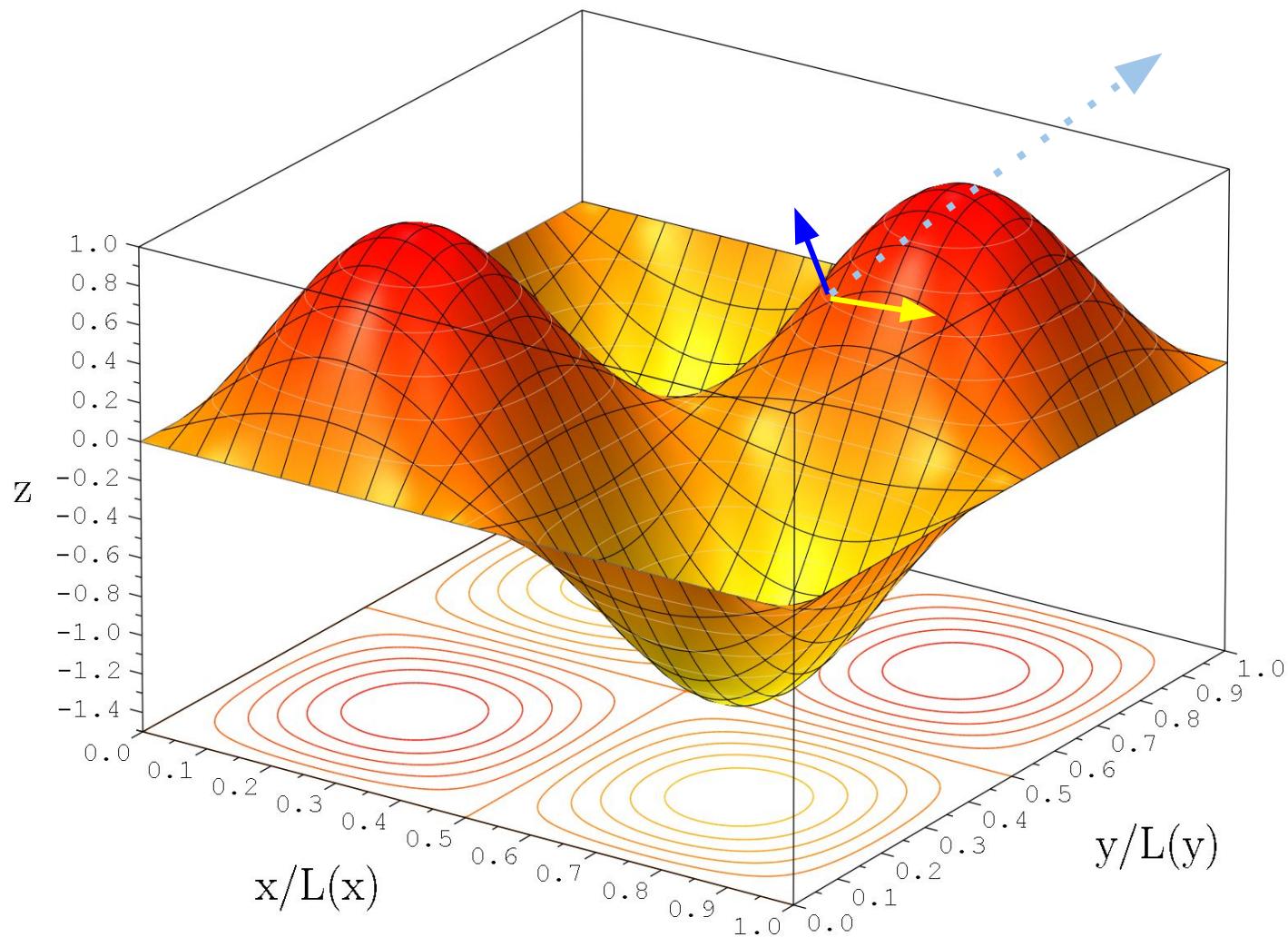
Descenso del gradiente



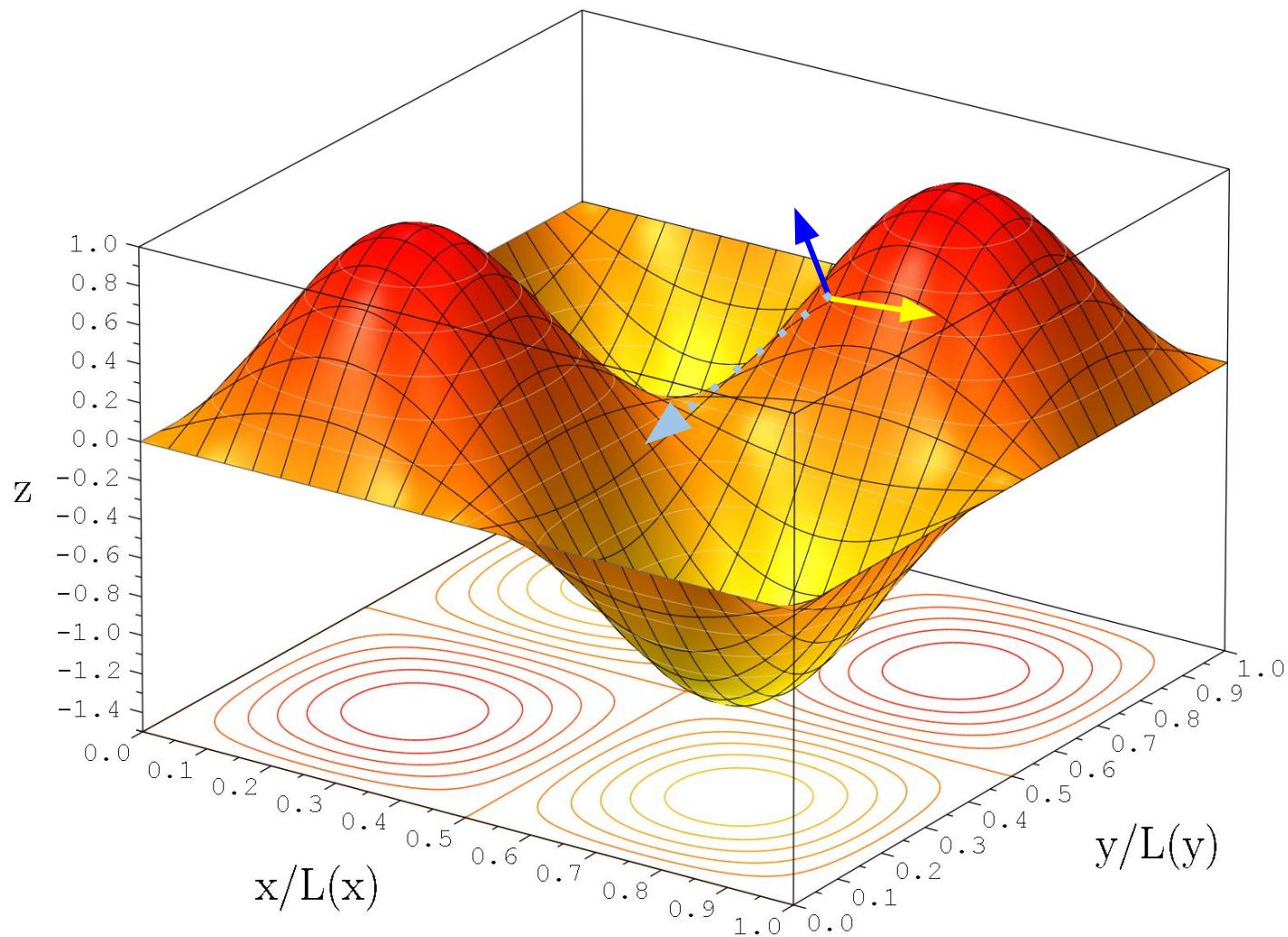
Descenso del gradiente



Descenso del gradiente



Descenso del gradiente



Descenso del gradiente

The screenshot shows a video player interface from Platzi. On the left, there's a sidebar with a navigation bar and a list of chapters numbered 21 through 30. The main content area features a large, semi-transparent dark blue overlay with white text asking, "¿Qué es el descenso del gradiente?". To the right of this overlay is a video frame showing a man with glasses and a beard, wearing a dark hoodie, speaking into a microphone. Below the video frame, the name "Enrique Devars" is displayed. At the bottom of the screen, there are several interactive buttons: "Descripción", "Reportar un problema", a play button, and a link labeled "Graficando nuestra función d...".

Platzi

¿Qué es el descenso del gradiente?

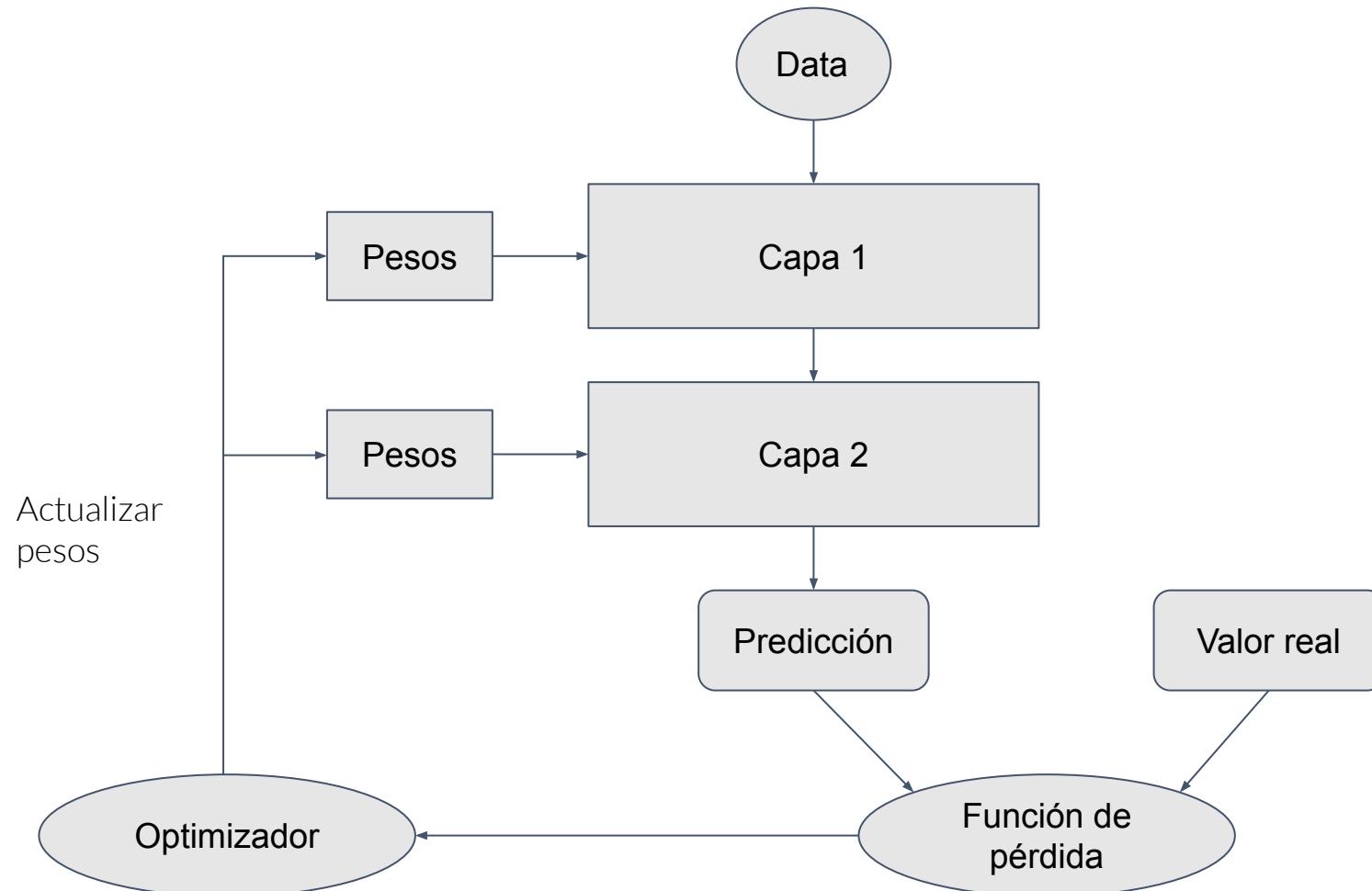
Enrique Devars

Reportar un problema

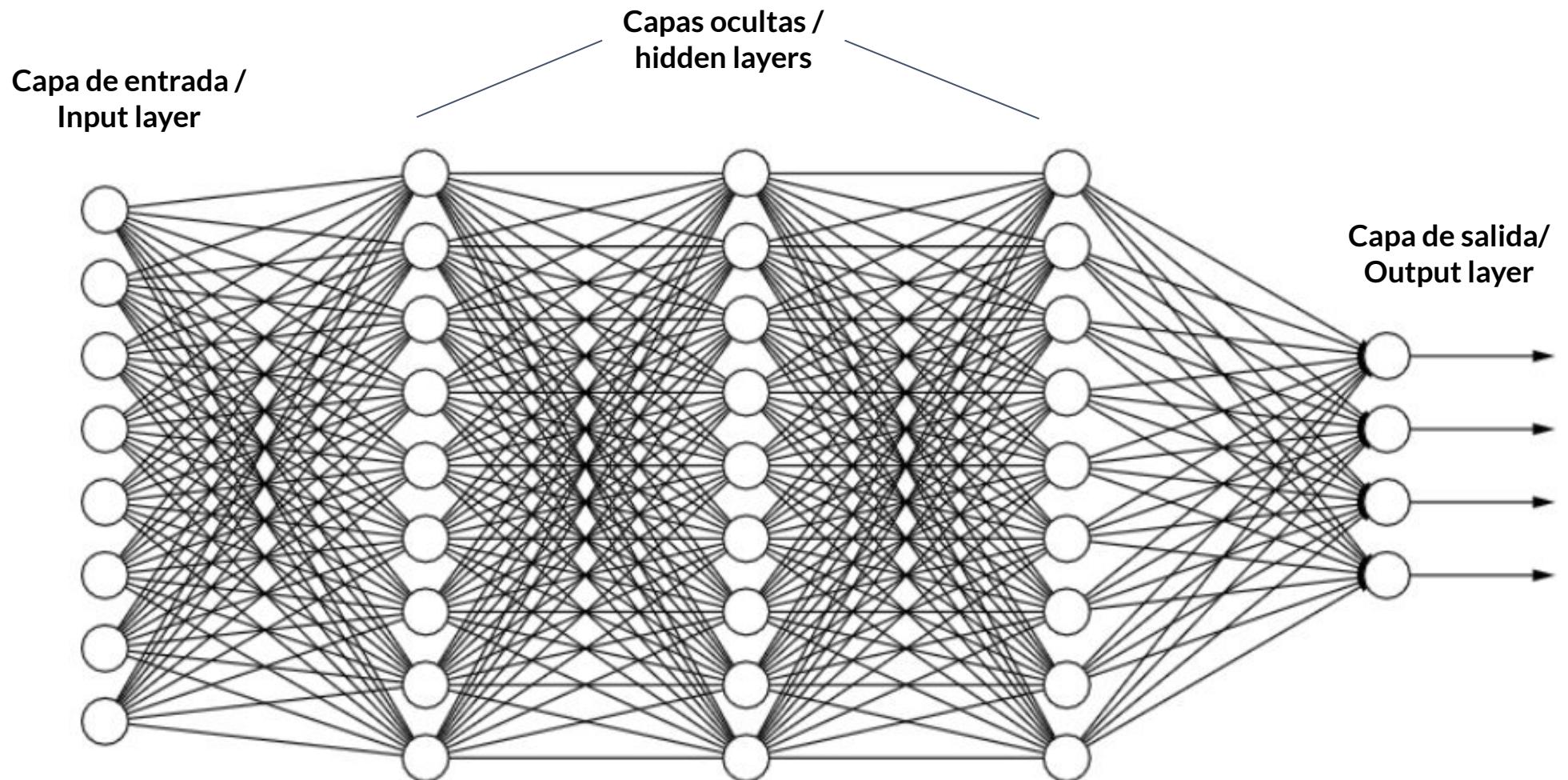
Graficando nuestra función d...

Backpropagation

Backpropagation



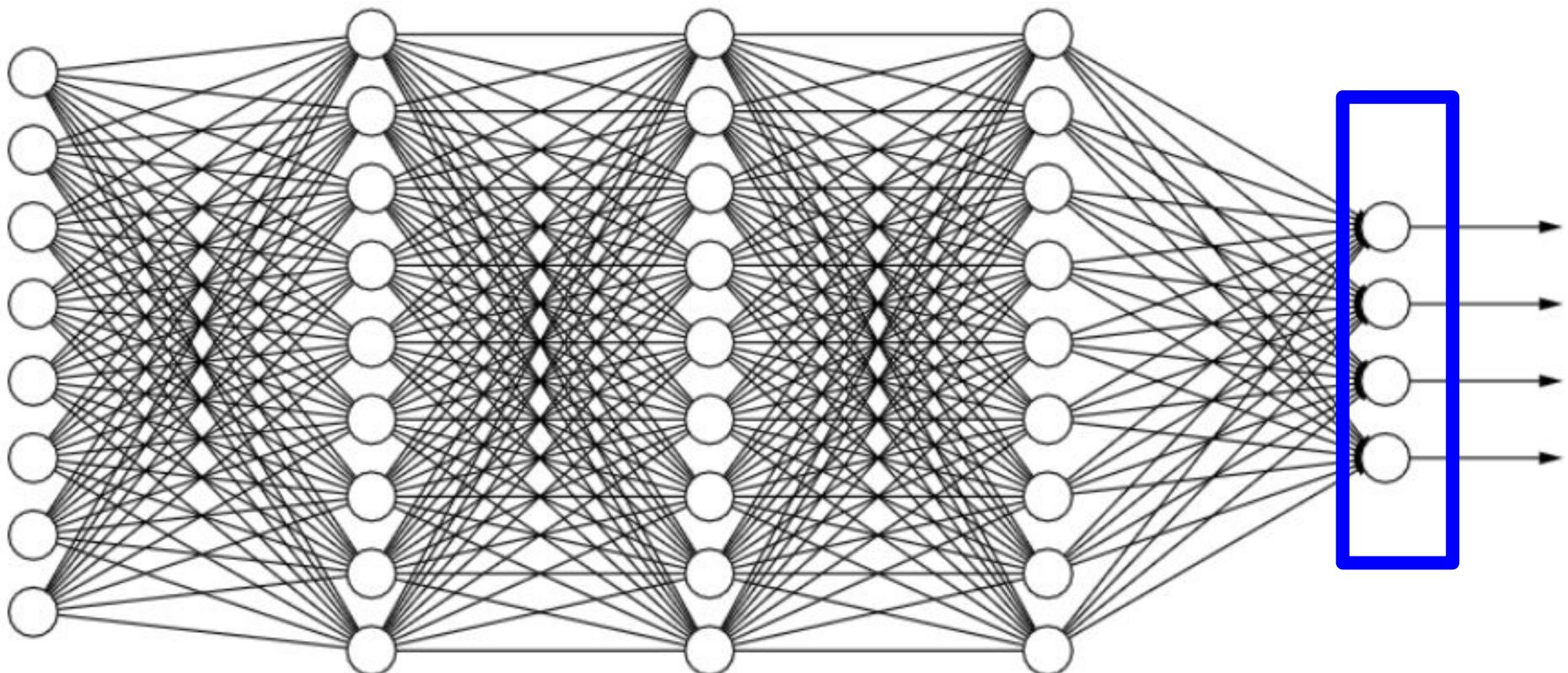
Backpropagation



Backpropagation



Backpropagation

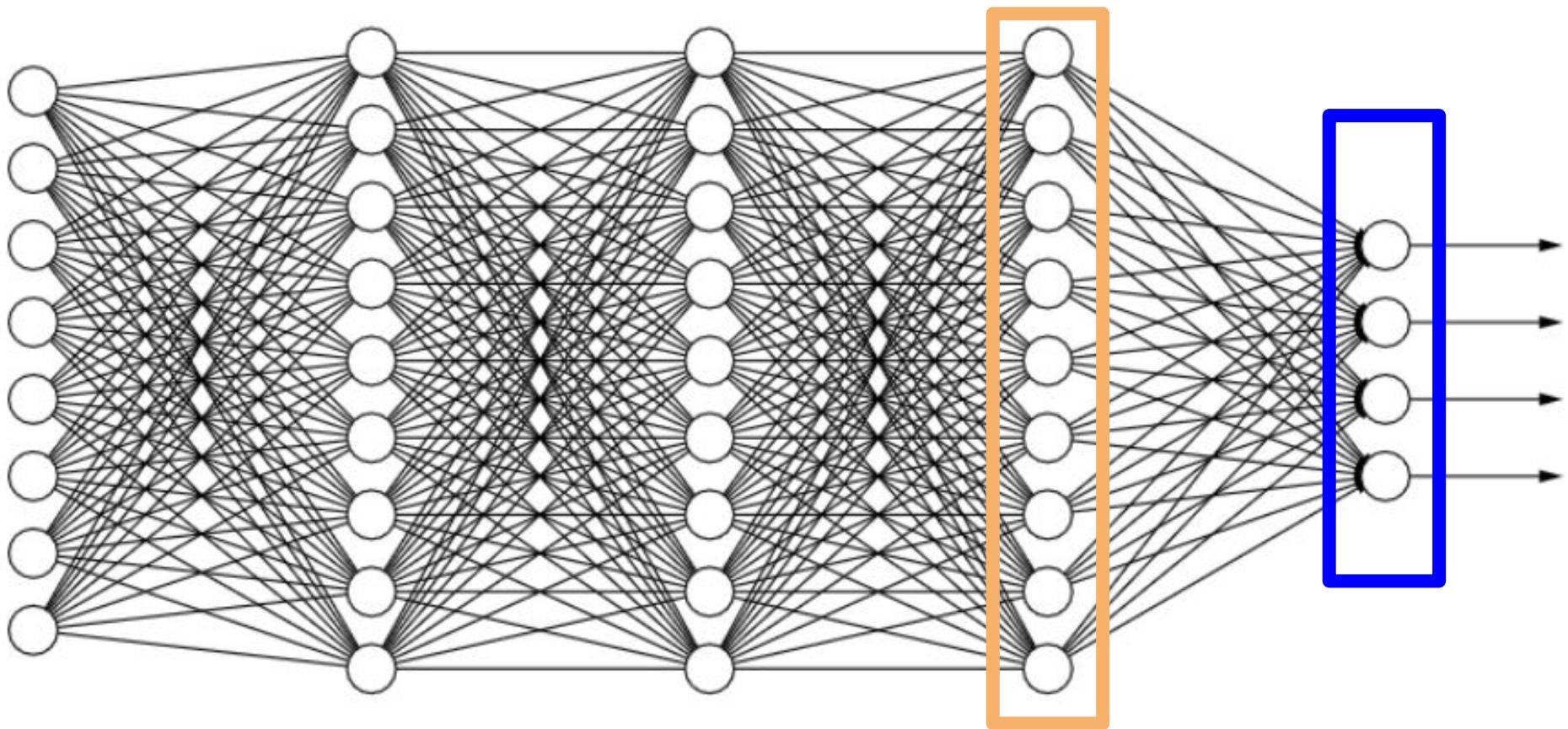


Backpropagation

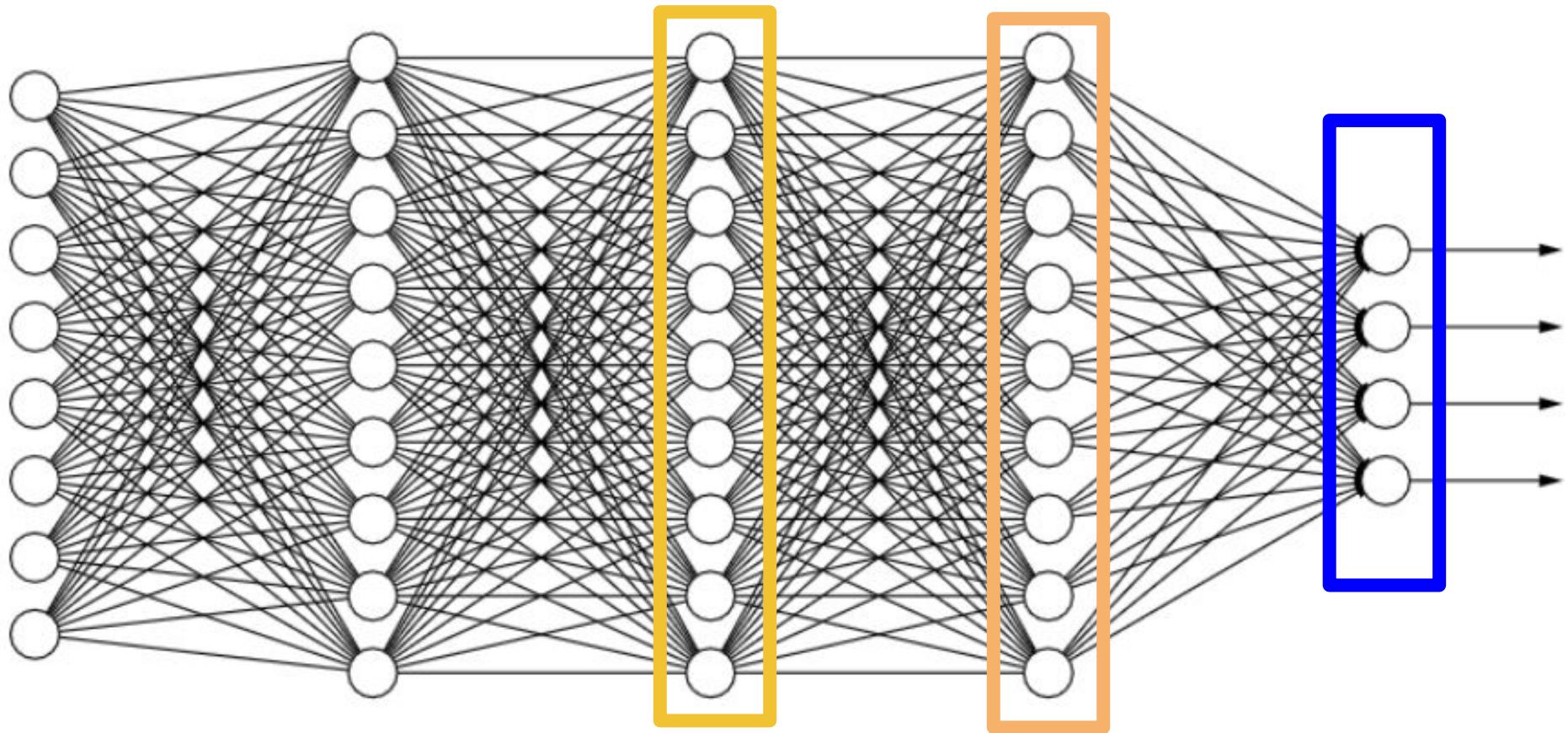


visita: www.crear-meme.com

Backpropagation



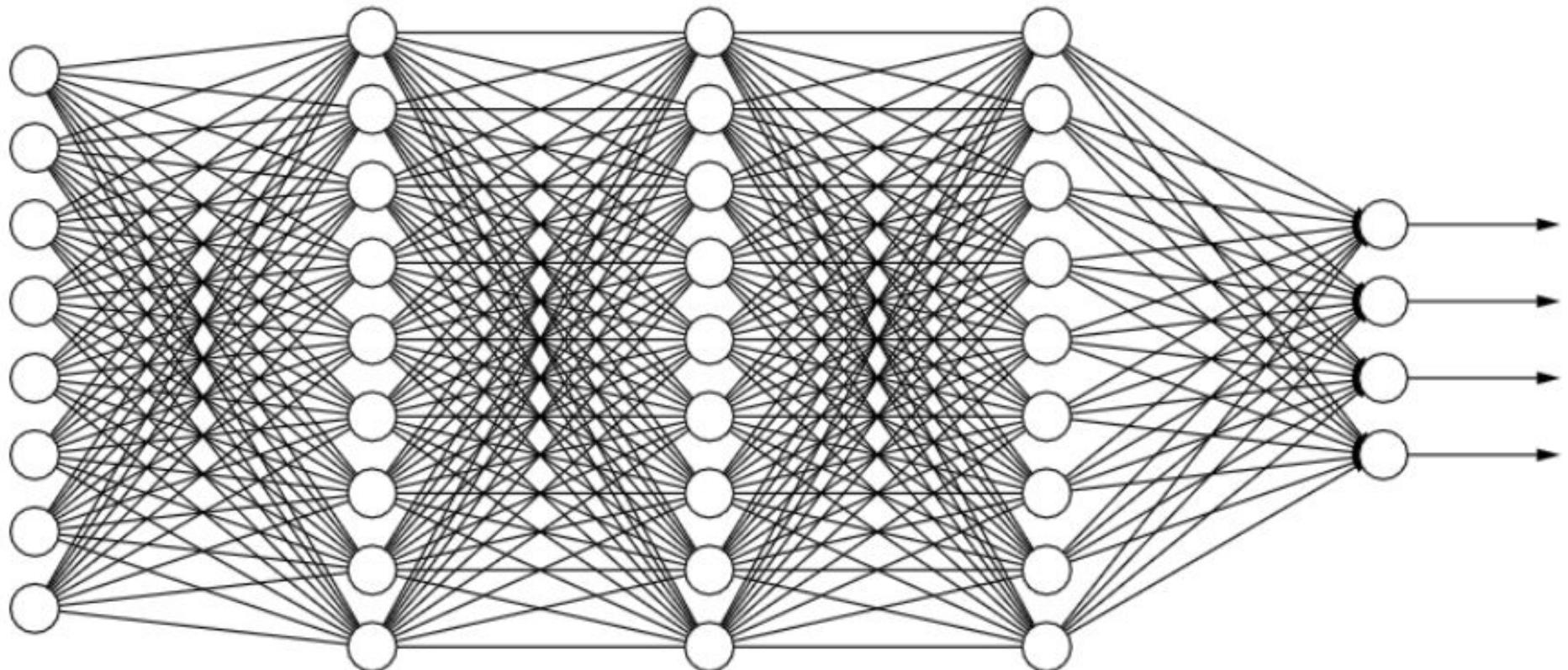
Backpropagation



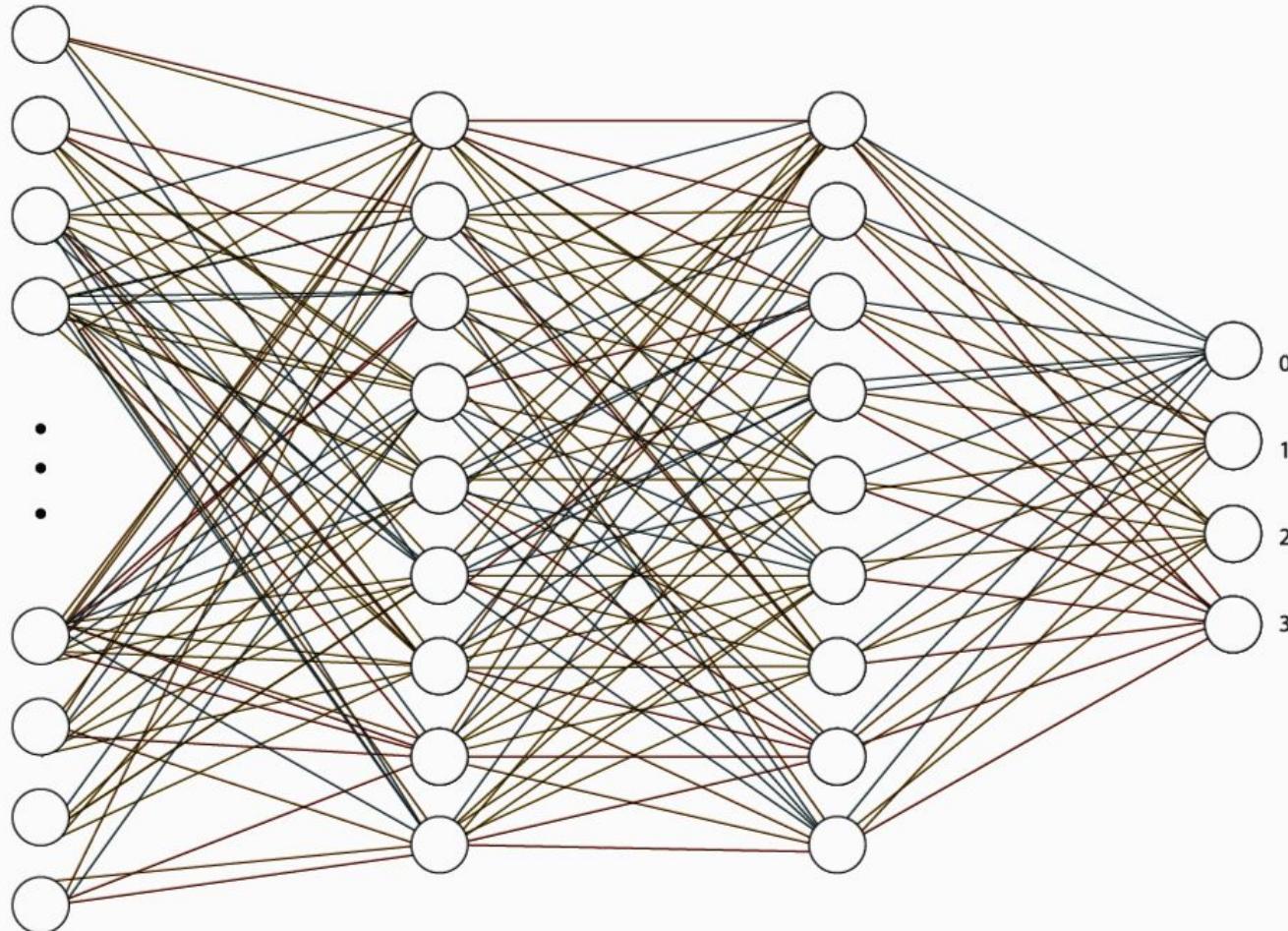
Backpropagation

¿Cómo distribuimos
este error?

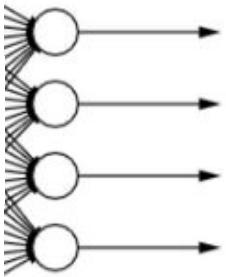
Backpropagation

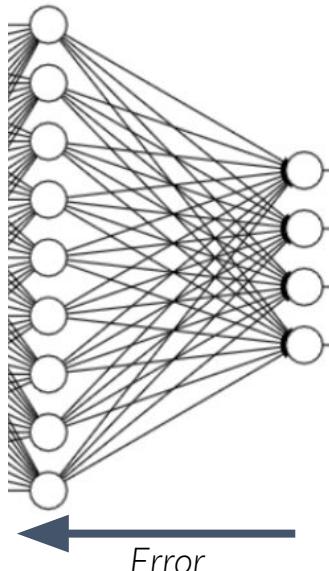


Backpropagation

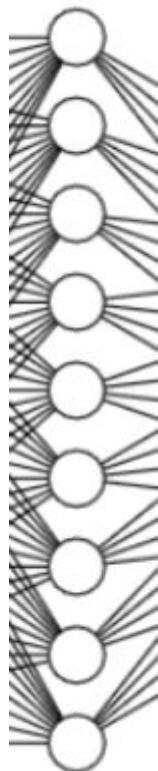


Backpropagation - Derivadas parciales

$$\delta[l] \equiv \text{Costo}(\text{Activación}(z)) \equiv \frac{\partial C}{\partial a[L]} * \frac{\partial a[L]}{\partial z[L]}$$


$$\delta[l-1] \equiv W[l]\delta[l] * \frac{\partial a[l-1]}{\partial z[l-1]}$$


Backpropagation - Derivadas parciales



$$\equiv \delta[l - 1] * a[l - 2]$$

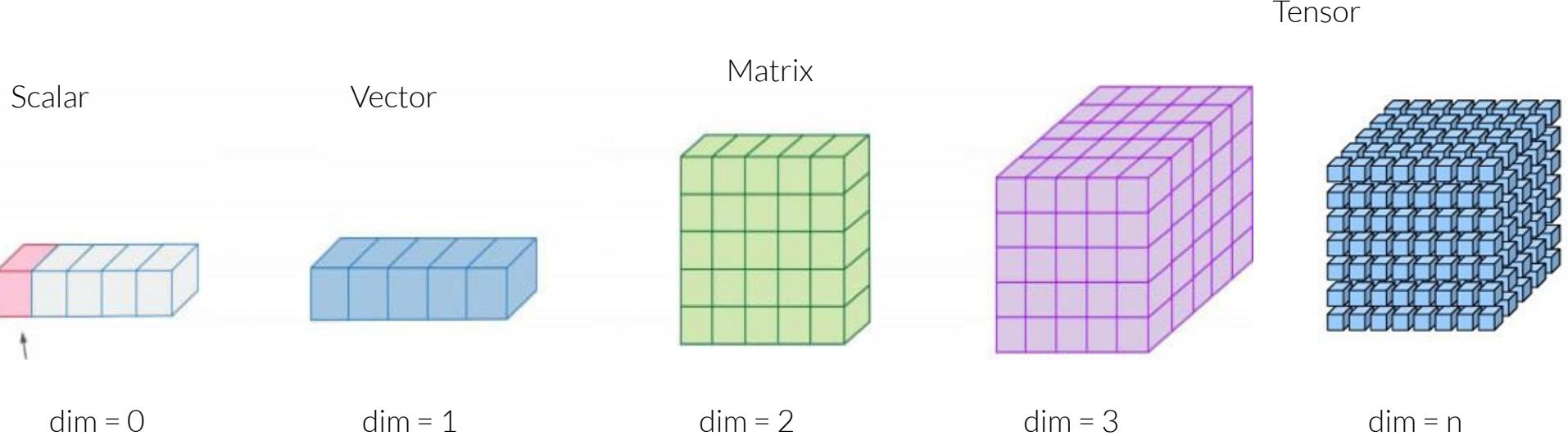
Backpropagation - Derivadas parciales



Playground - Tensorflow

Dimensiones, tensores y reshape

Dimensiones de los datos



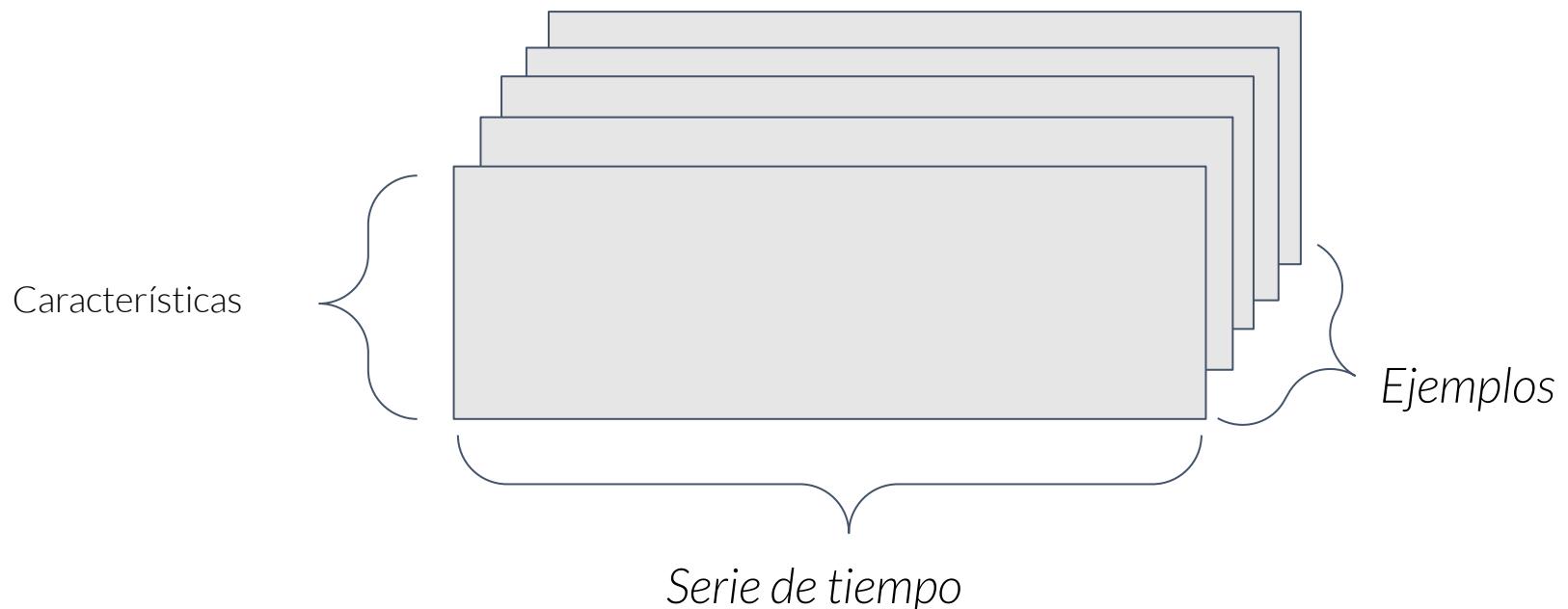
Matrix - 2d

(Ejemplos, Características)

	A	B	C	D	E	F
1	Order ID	Product	Category	Amount	Date	Country
2	1	Carrots	Vegetables	\$4,270	1/6/2012	United States
3	2	Broccoli	Vegetables	\$8,239	1/7/2012	United Kingdom
4	3	Banana	Fruit	\$617	1/8/2012	United States
5	4	Banana	Fruit	\$8,384	1/10/2012	Canada
6	5	Beans	Vegetables	\$2,626	1/10/2012	Germany
7	6	Orange	Fruit	\$3,610	1/11/2012	United States
8	7	Broccoli	Vegetables	\$9,062	1/11/2012	Australia
9	8	Banana	Fruit	\$6,906	1/16/2012	New Zealand
10	9	Apple	Fruit	\$2,417	1/16/2012	France
11	10	Apple	Fruit	\$7,431	1/16/2012	Canada
12	11	Banana	Fruit	\$8,250	1/16/2012	Germany
13	12	Broccoli	Vegetables	\$7,012	1/18/2012	United States
14	13	Carrots	Vegetables	\$1,903	1/20/2012	Germany

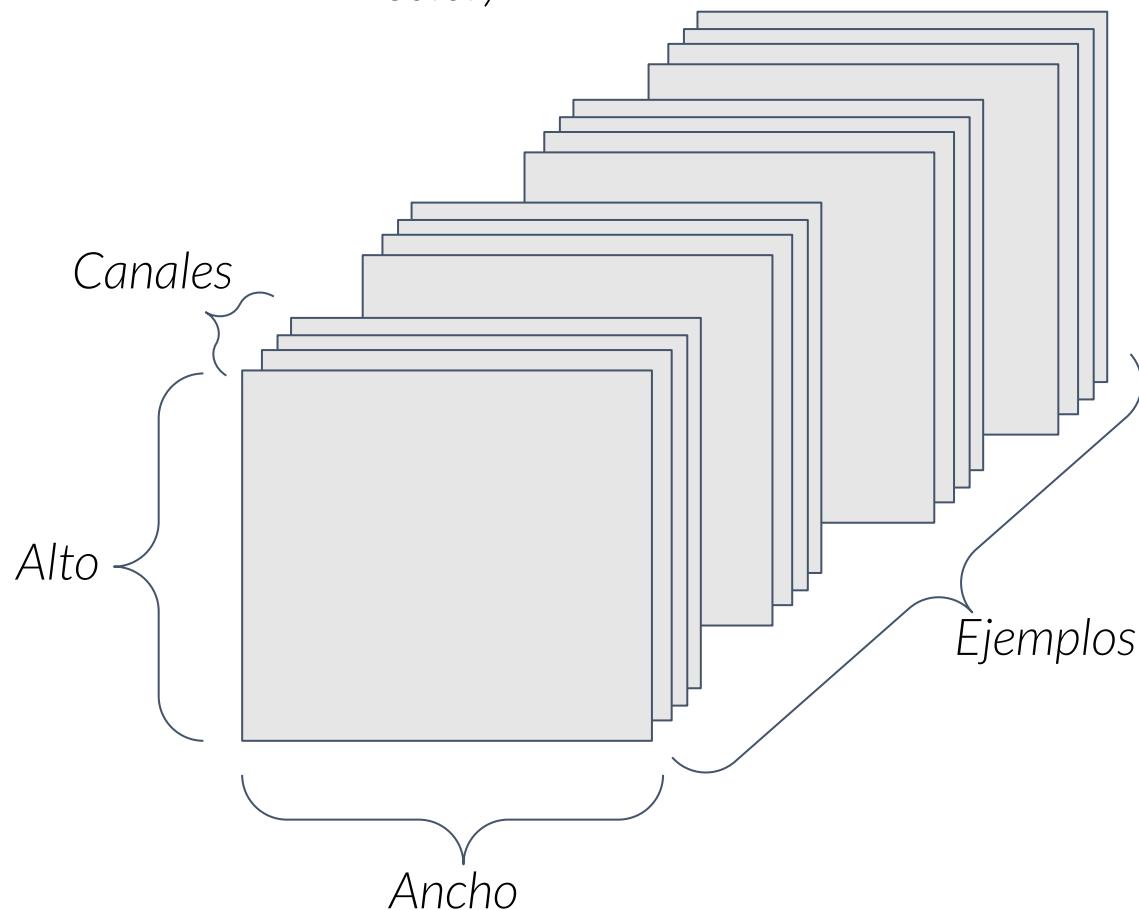
Tensor - 3d

(Ejemplos, Serie de tiempo, Características)



Tensor - 4d - Images

(Ejemplos, Ancho, Alto, Canales de color)



Creando nuestra red neuronal usando numpy y matemáticas

Manejo de datos

Train, validation y test

Manejo de los datos

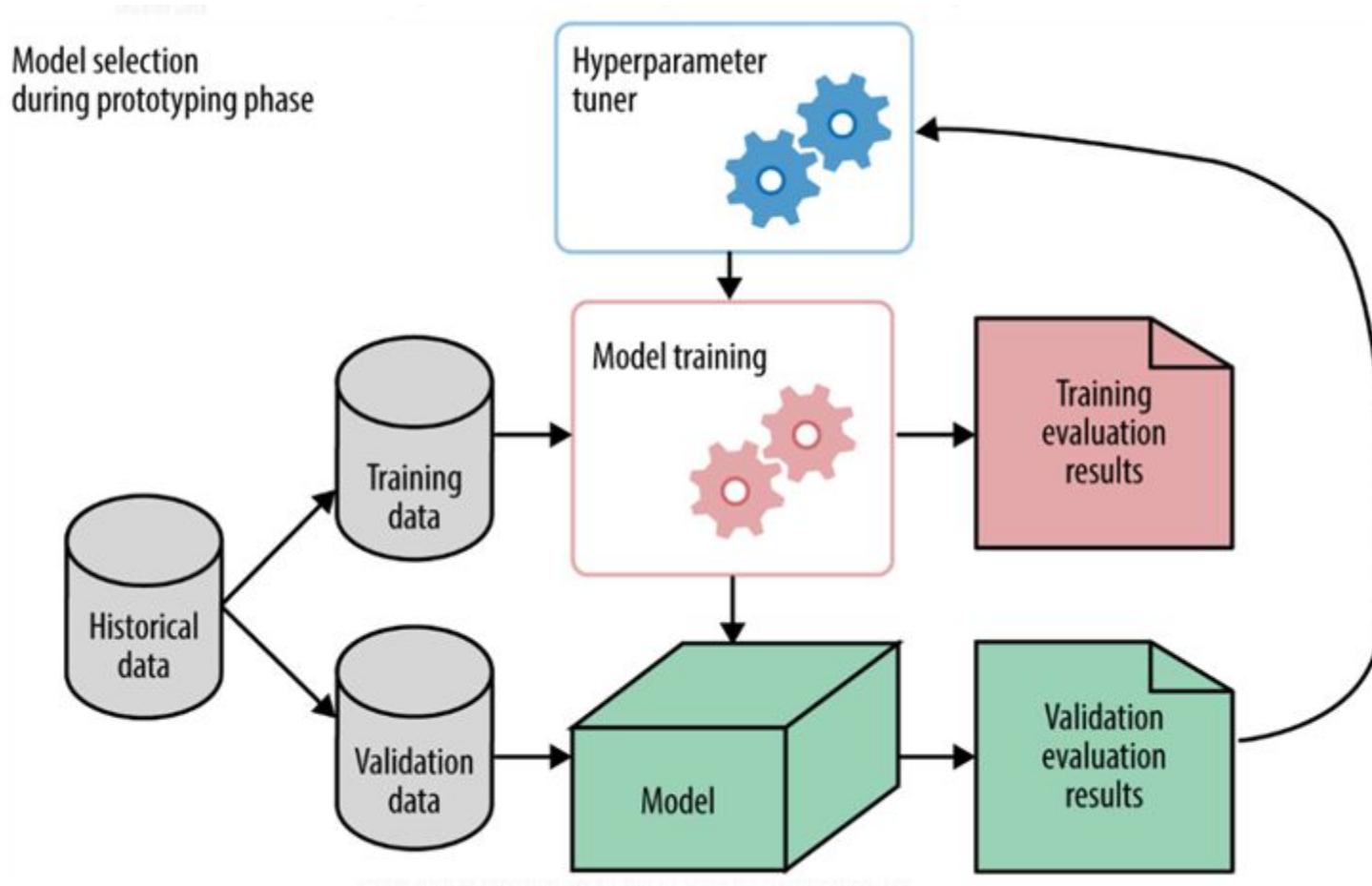
ENTRENAMIENTO

TEST

Manejo de los datos

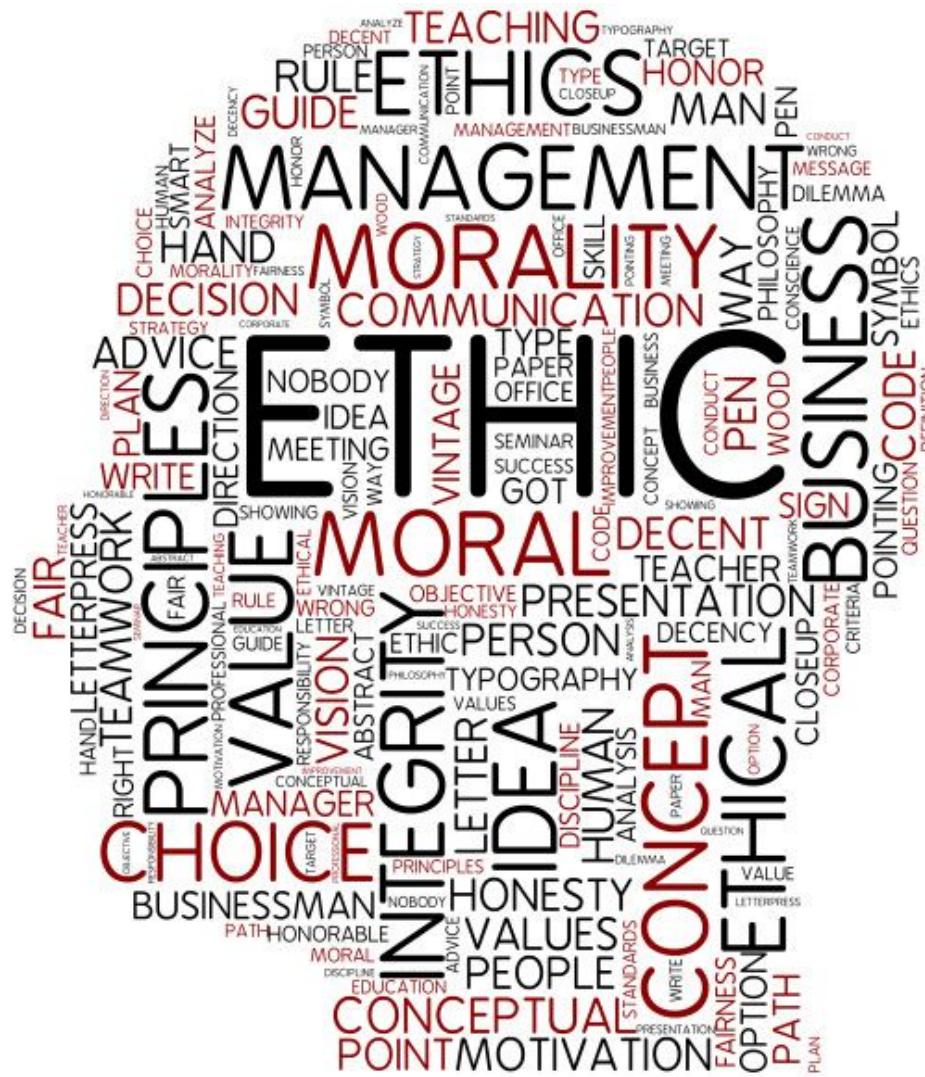


Set de datos de validación



Set de datos de test

TEST

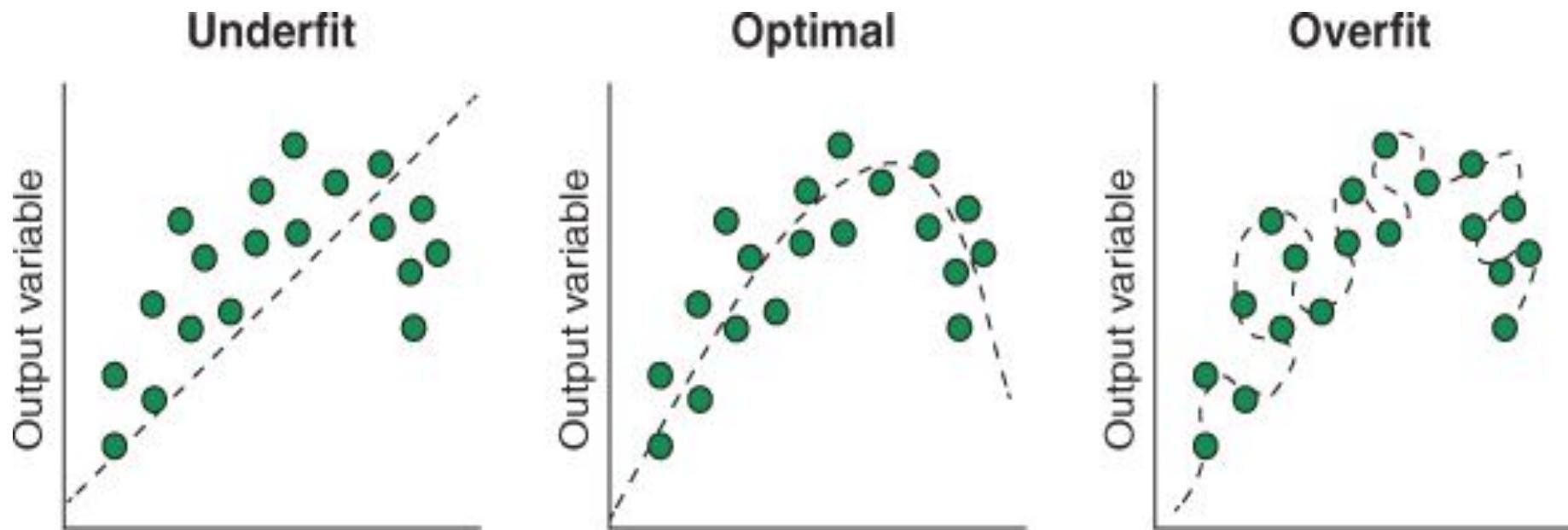


Resolviendo un problema de clasificación binaria

Entrenamiento del modelo de clasificación binaria

Regularización - Dropout

Overfitting

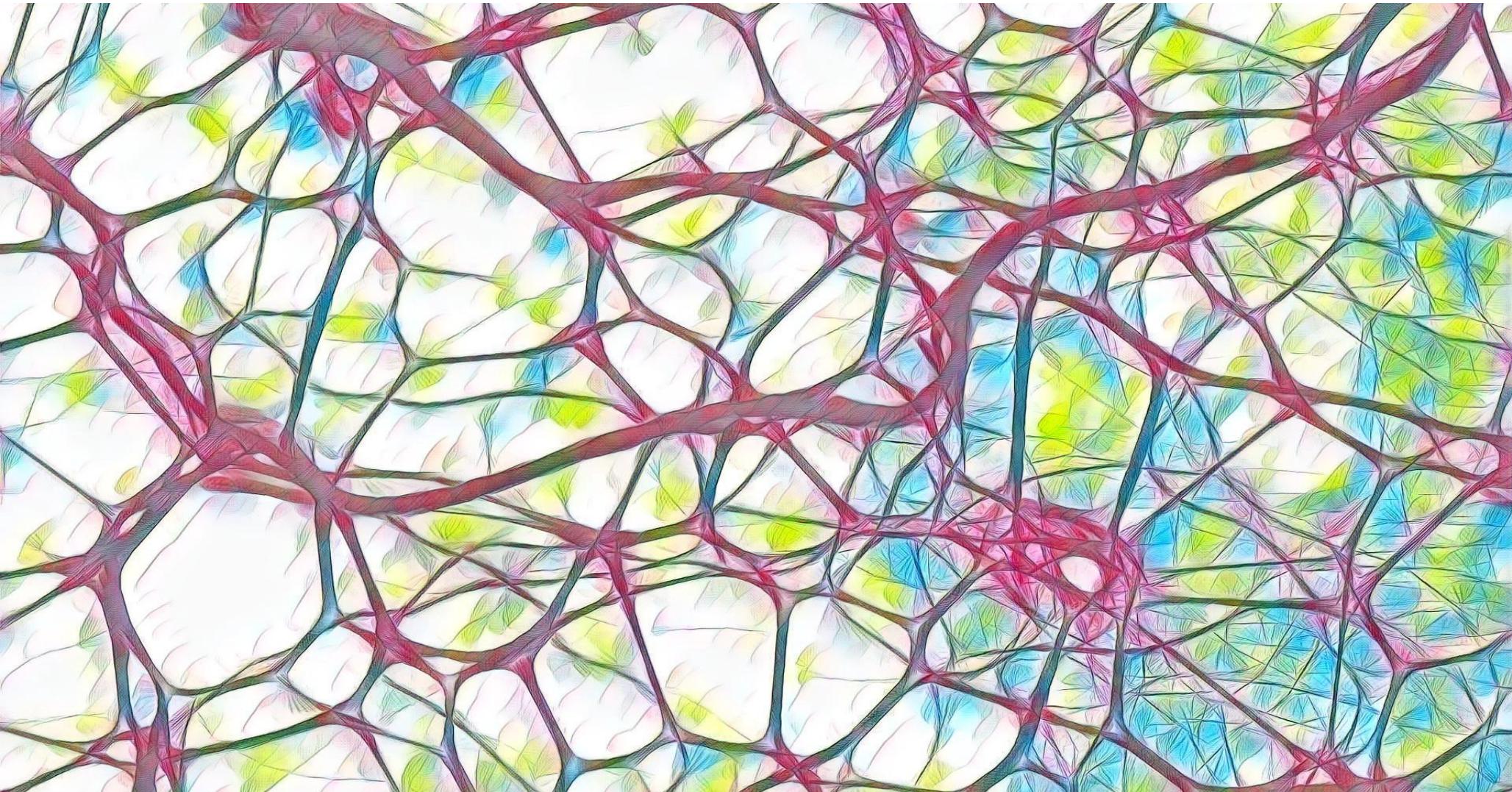


Overfitting



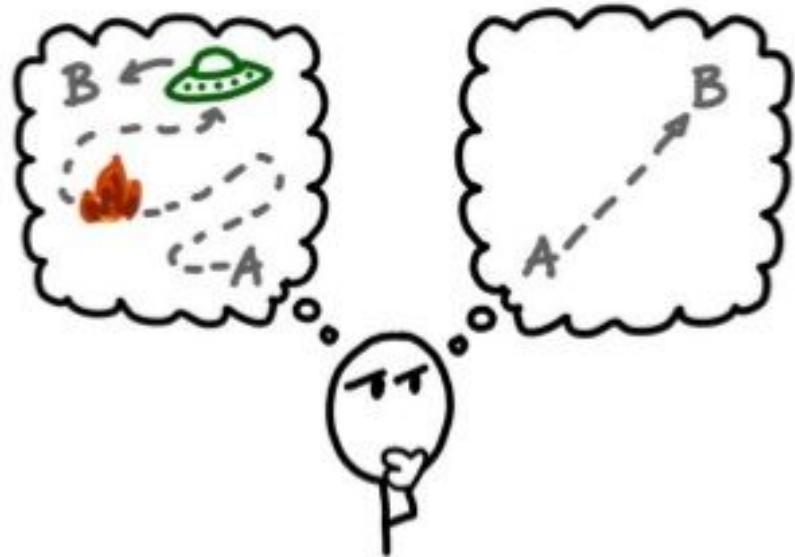
**THE BEST WAY TO
EXPLAIN OVERFITTING**

Overfitting



Regularización

Occam's Razor

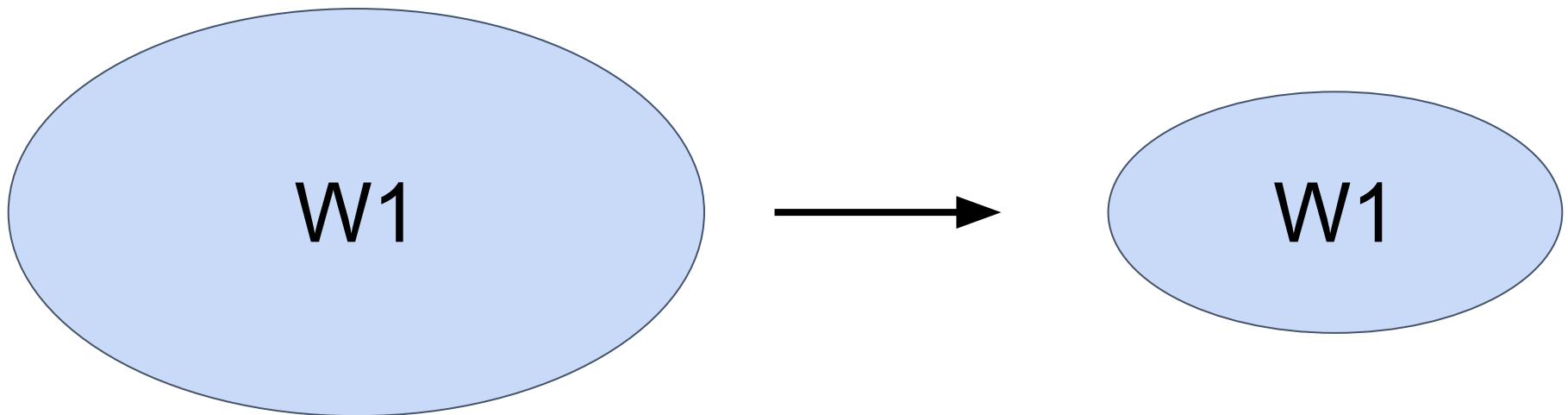


“When faced with two equally good hypotheses, always choose the simpler.”

Regularización

Reducir la complejidad
en el modelo

Regularización



Regularización

L1 Regularization

$$\text{Cost} = \sum_{i=0}^N (y_i - \sum_{j=0}^M x_{ij} W_j)^2 + \lambda \sum_{j=0}^M |W_j|$$

L2 Regularization

$$\text{Cost} = \sum_{i=0}^N (y_i - \sum_{j=0}^M x_{ij} W_j)^2 + \lambda \sum_{j=0}^M W_j^2$$

Loss function Regularization Term

Regularización

L1 Regularization

$$\text{Cost} = \sum_{i=0}^N (y_i - \sum_{j=0}^M x_{ij} W_j)^2 + \lambda \sum_{j=0}^M |W_j|$$

L2 Regularization

$$\text{Cost} = \sum_{i=0}^N (y_i - \sum_{j=0}^M x_{ij} W_j)^2 + \lambda \sum_{j=0}^M W_j^2$$

Loss function Regularization Term

Dropout



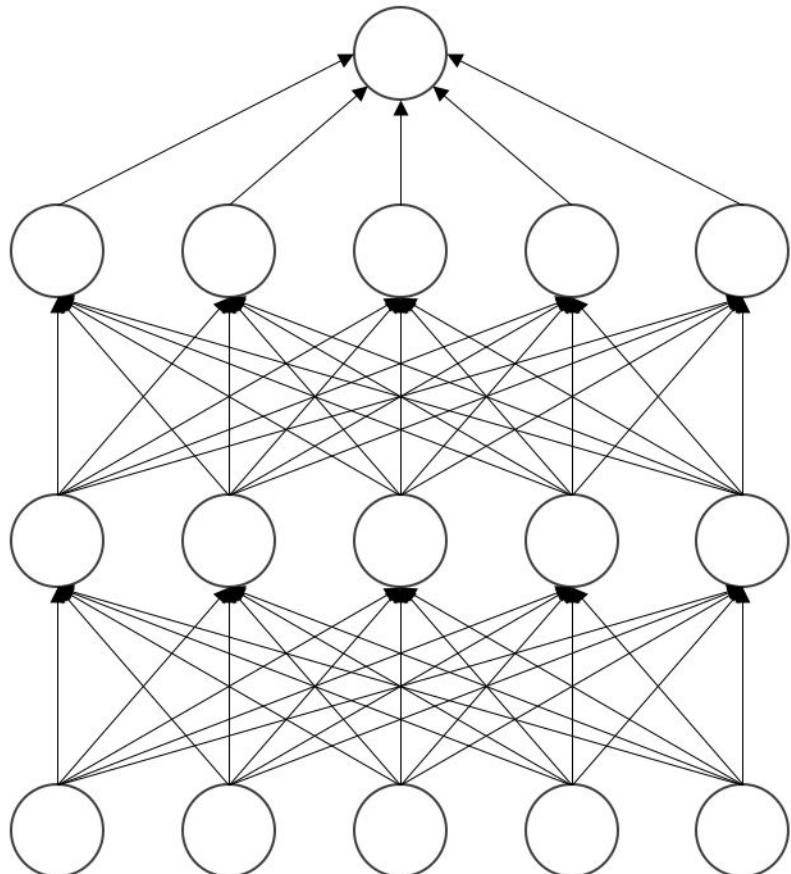
Dropout

0.5	2.1	0.6	1.5
0.6	0.2	1.1	2.3
3.0	0.8	1.5	1.2
2.2	0.7	2.2	0.9

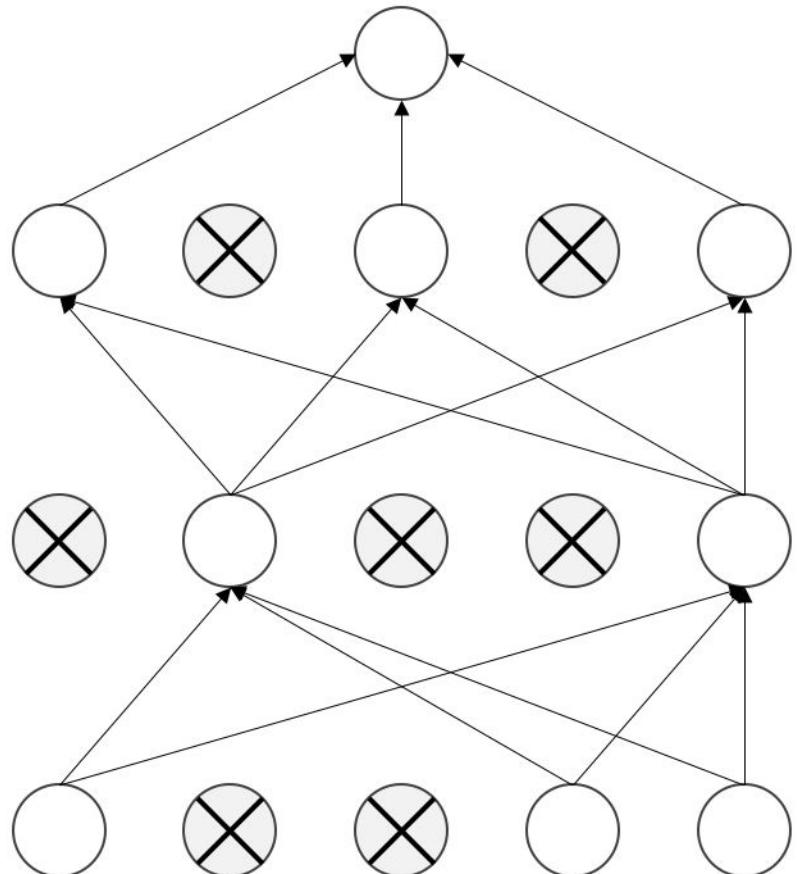
Dropout 50%
→

0	2.1	0.6	0
0	0.2	0	2.3
3.0	0	1.5	0
2.2	0	0	0.9

Dropout



Standard Neural Net



After applying dropout

Reduciendo el overfitting

Resolviendo un problema de clasificación múltiple

Entrenamiento del modelo de clasificación múltiple

Resolviendo un problema de regresión

Entrenamiento del modelo de regresión

Análisis de resultados del modelo de regresión

¿Qué sigue por aprender
de redes neuronales?

■ ¿Qué funciones utilizar?

Problem type	Last-layer activation	Loss function
Binary classification	sigmoid	binary_crossentropy
Multiclass, single-label classification	softmax	categorical_crossentropy
Multiclass, multilabel classification	sigmoid	binary_crossentropy
Regression to arbitrary values	None	mse
Regression to values between 0 and 1	sigmoid	mse or binary_crossentropy

Más por aprender

- CNN (Convolutional neural network)
- NLP (Natural language process)
- Sequences
- Hyper parameters tuning
- Gans (Generative adversarial networks)
- LSTM (Long short-term memory)
- Callbacks
- Tensorboard
- Y un mundo más de posibilidades

Carlos Andrés Alarcón



@Alarcon7a