Deep Learning

Redes Neuronales con TensorFlow

Agosto 30

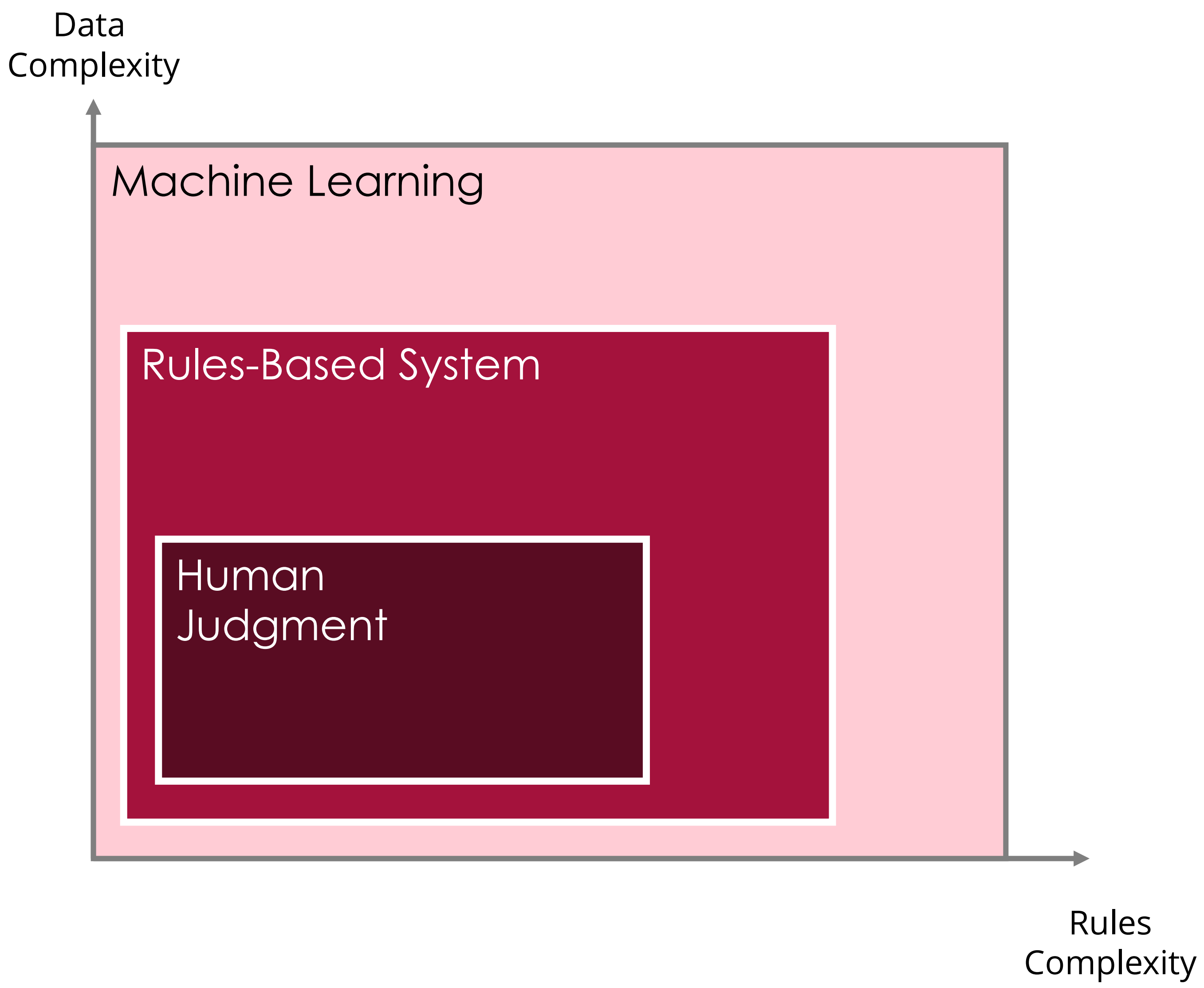
Deep Learning

Redes Neuronales con TensorFlow

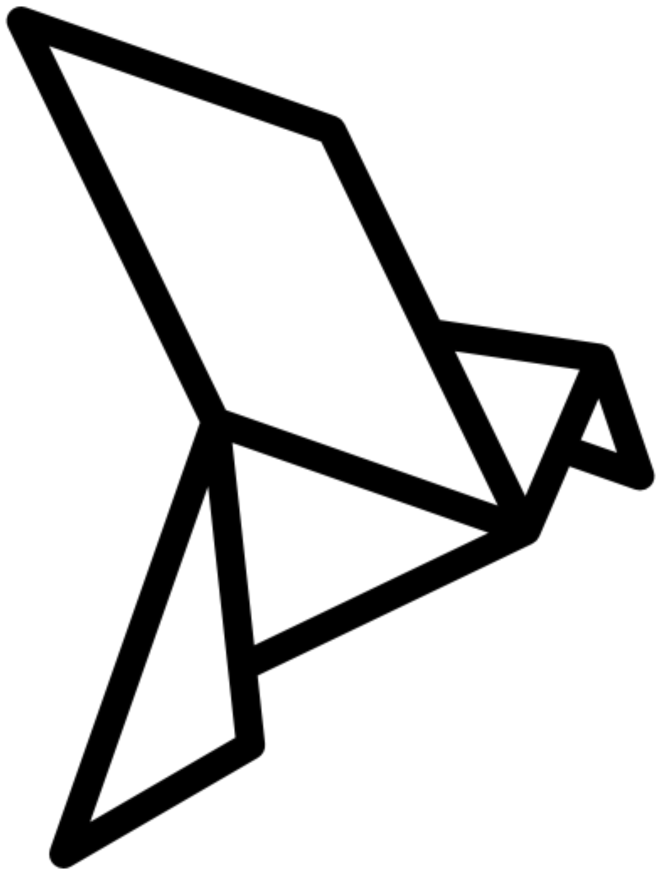
Agosto 30

- Deep Learning vs Machine Learning
- Teoría del Deep Learning
- Cuatro grandes teoría
- Arquitectura de una red neuronal
- ¿Qué hace una red neuronal por dentro?
- Backpropagation

Machine Learning como campo de la AI



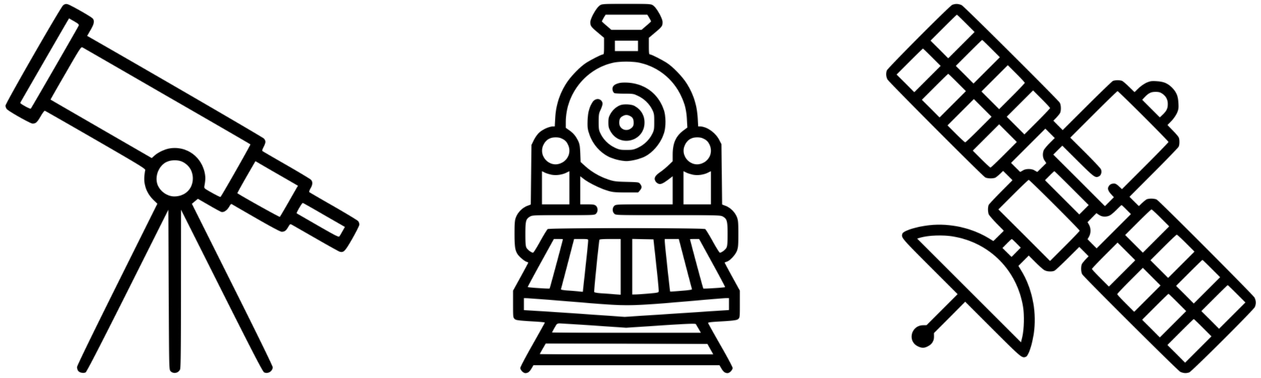
*Tomado de Meor Amer (2022)



Deep Learning inspirado por el cerebro, pero no en todas sus funciones, tal y como las aves y los aviones

¿Qué pasa cuando la teoría viene **después** de la práctica?

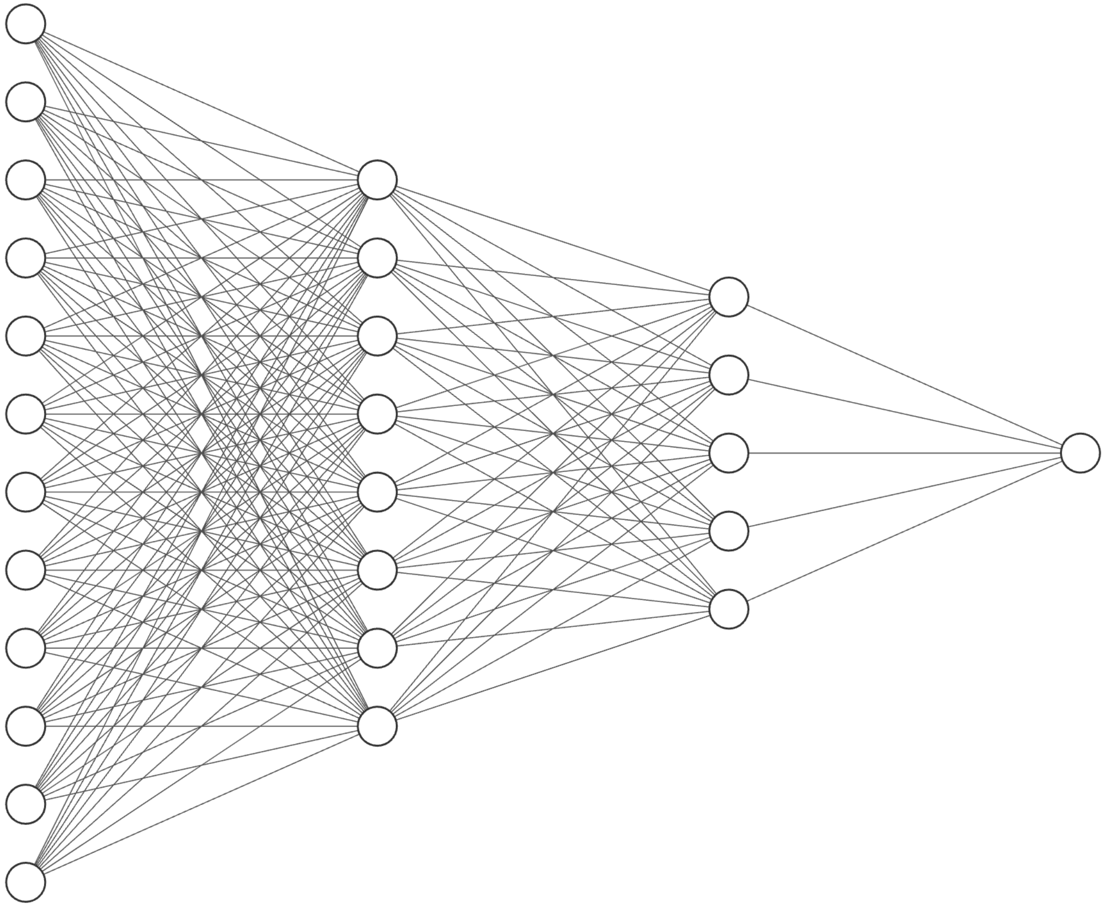
HD INSTITUTO HUMAN DELTA MR



"In the history of science and technology, the engineering artifact often comes first: the telescope, the steam engine, digital communication. The theory that explains its function and its limitations often appears later: the laws of refraction, thermodynamics, and information theory."

\- Yann LeCunn

Deep Learning vs Machine Learning

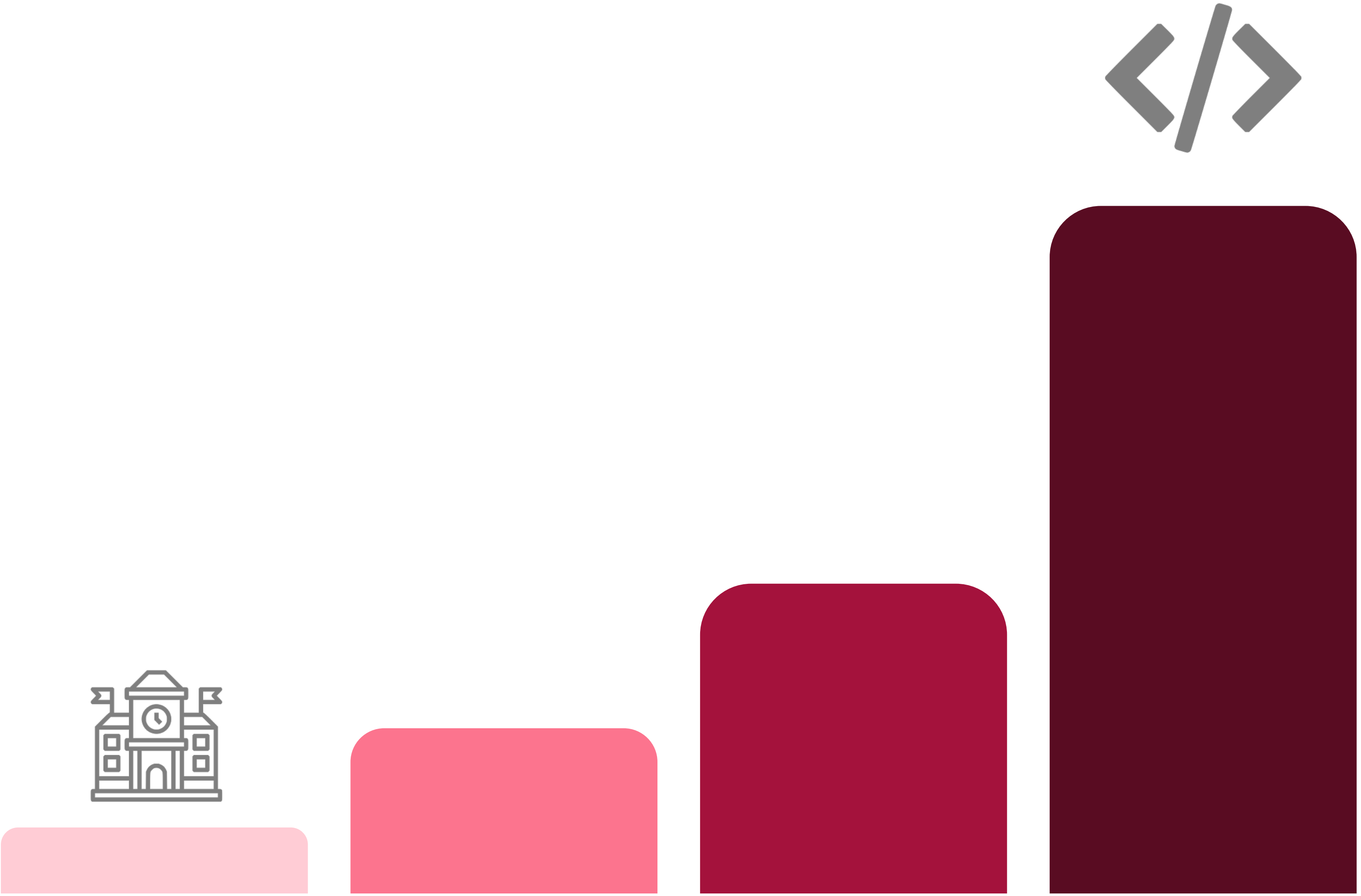


El verdadero poder de las redes neuronales viene de lo secuencial

Steam navigation brings nearer together the most distant nations. ...their theory is very little understood, and the attempts to improve them are still directed almost by chance. ...We propose now to submit these questions to a deliberate examination.

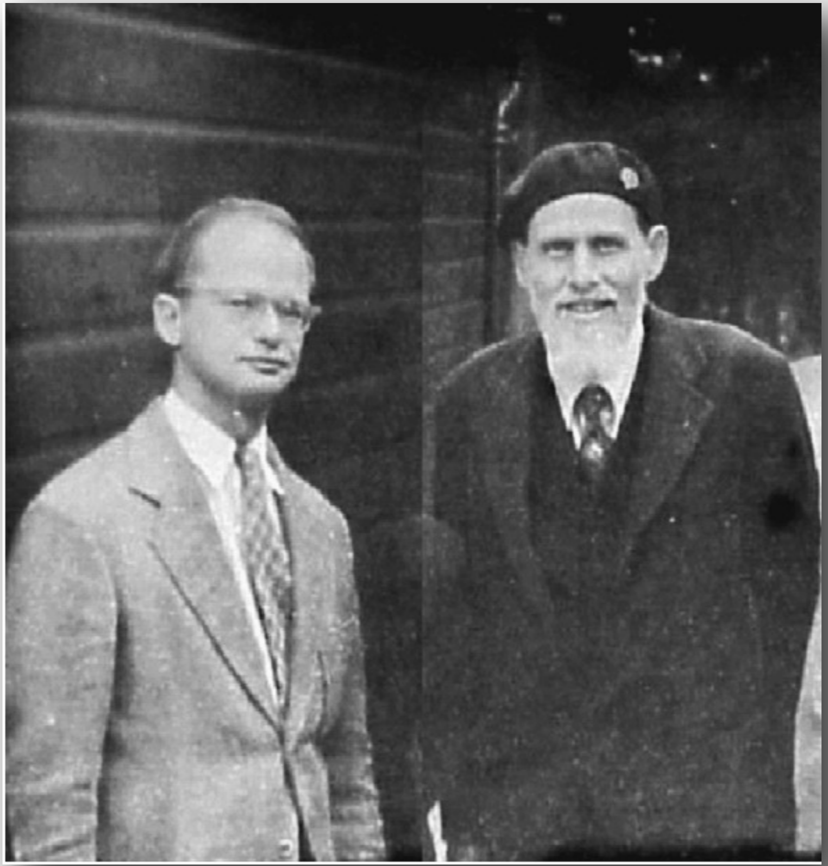
- Sadi Carnot

El verdadero problema, es que no existe una teoría del Deep Learning

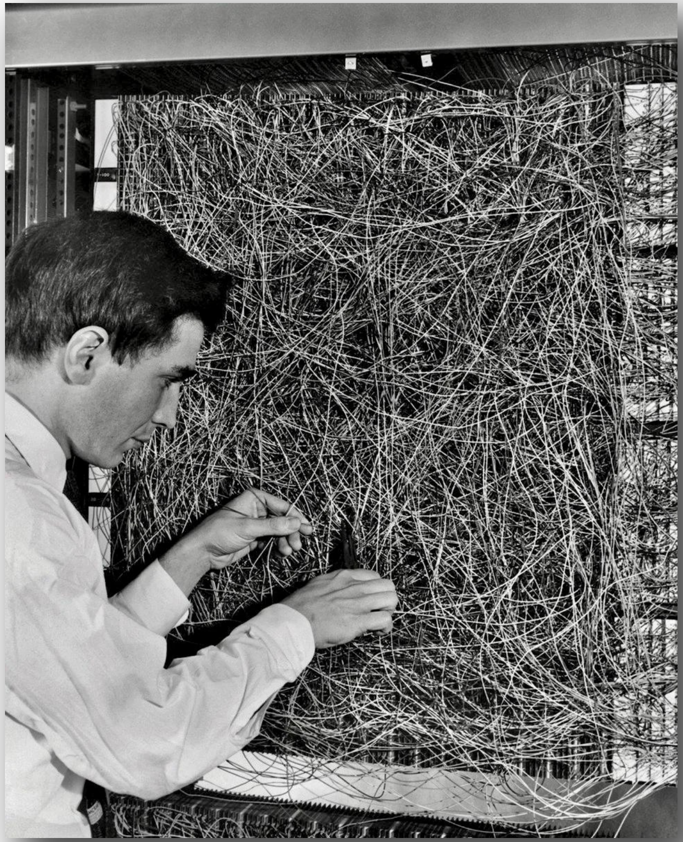


"Indeed, there is a serious disconnect between theory and practice: while practitioners have reached amazing milestones, they have far outpaced the theorists, whose analyses often involve assumptions so unrealistic that they lead to conclusions that are irrelevant to understanding deep neural networks as they are typically used."

-Daniel A. Roberts



McCulloch & Pitts

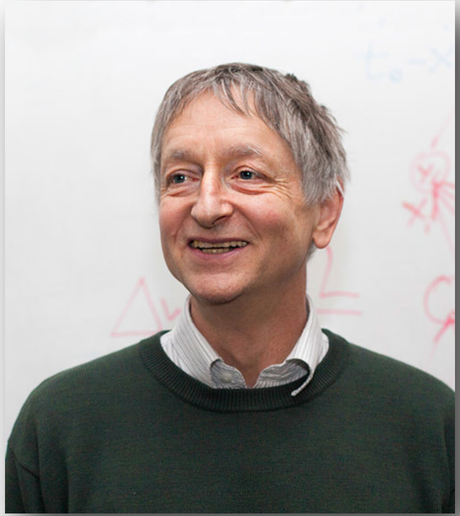


Frank Rosenblatt

Siglo XXI

Andrew NG

Yann LeCun

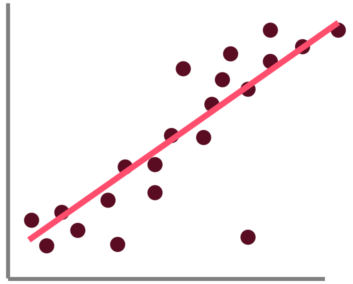


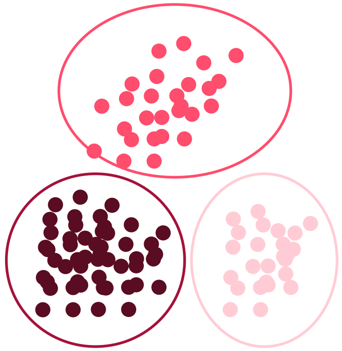
Geoffrey Hinton

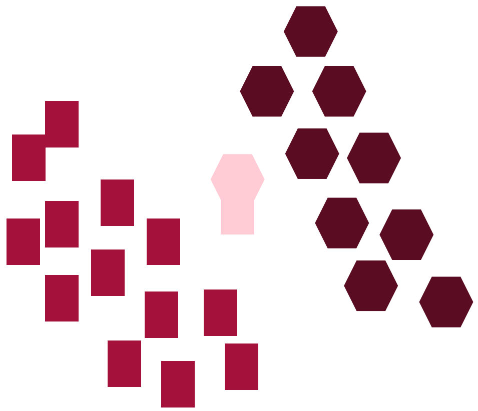
Yoshua Bengio

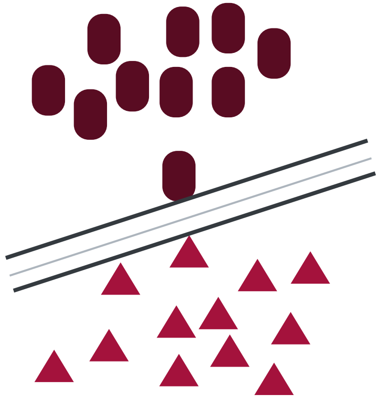
Ian Goodfellow

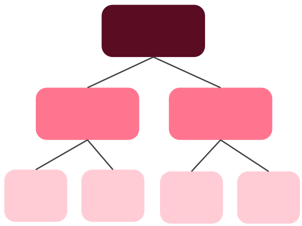
ML y DL

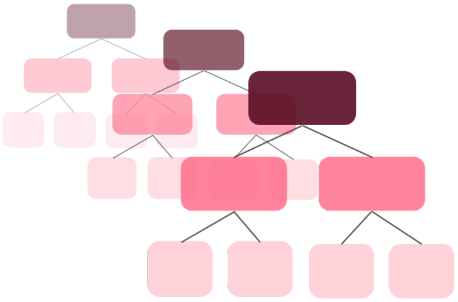

Regresión Lineal

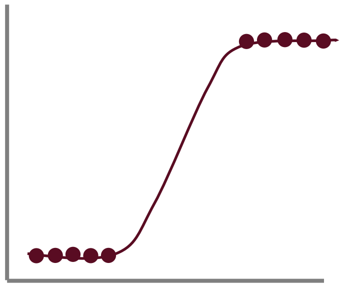

K-Means

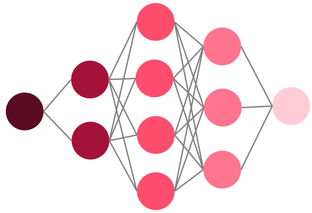

K-NN

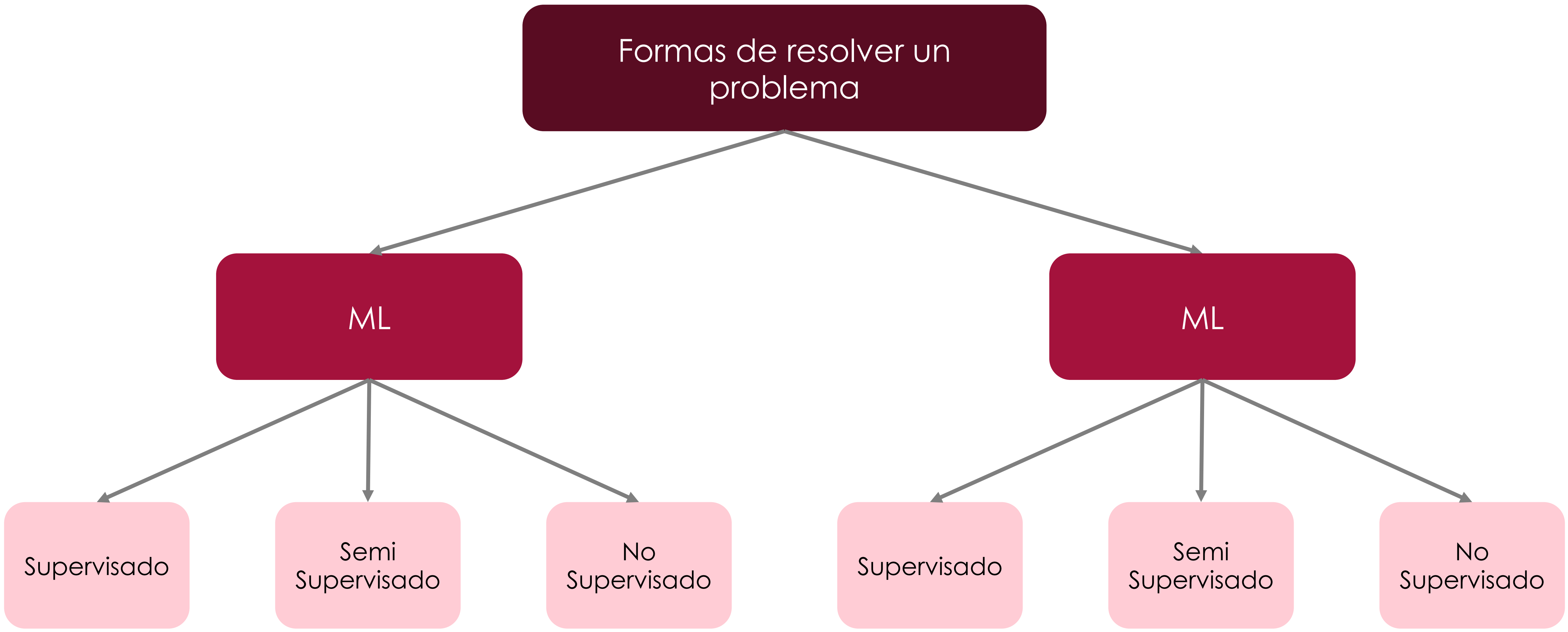

SVM


Decision Trees


Random Forests

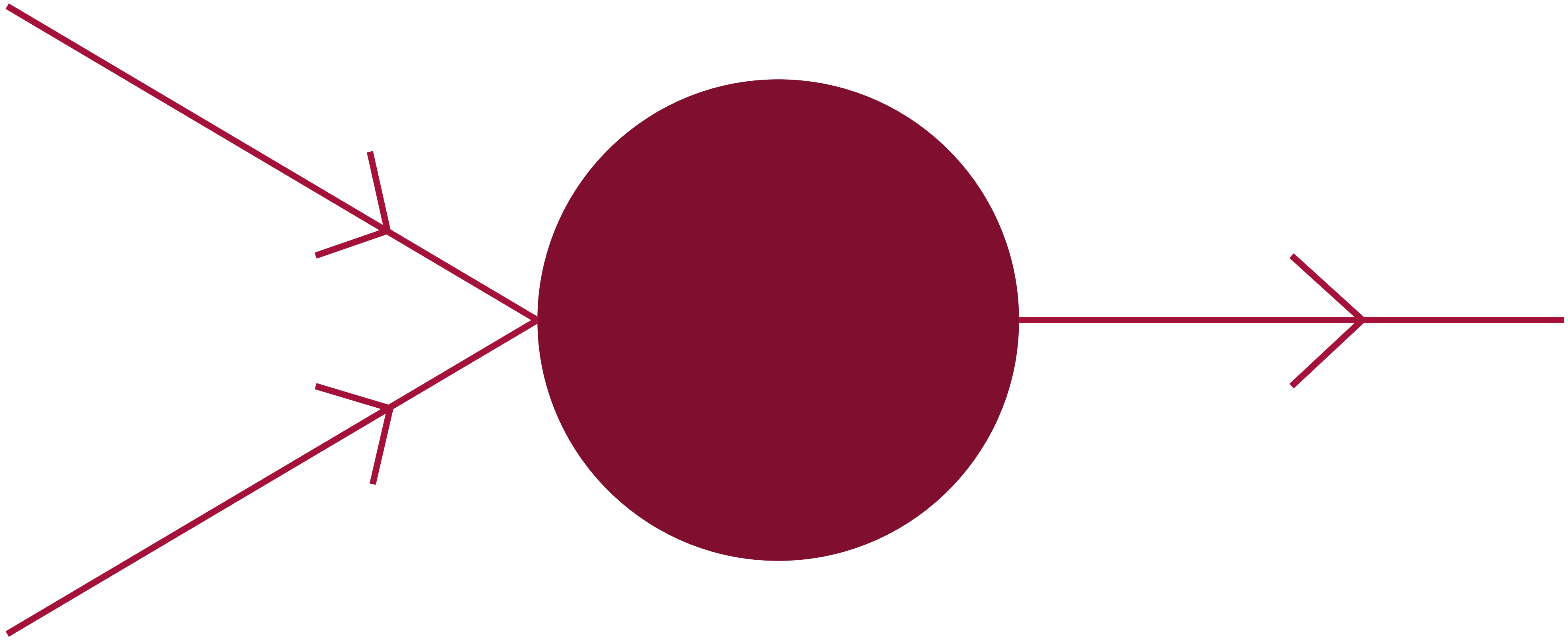

Logistic Regression


Neural Networks



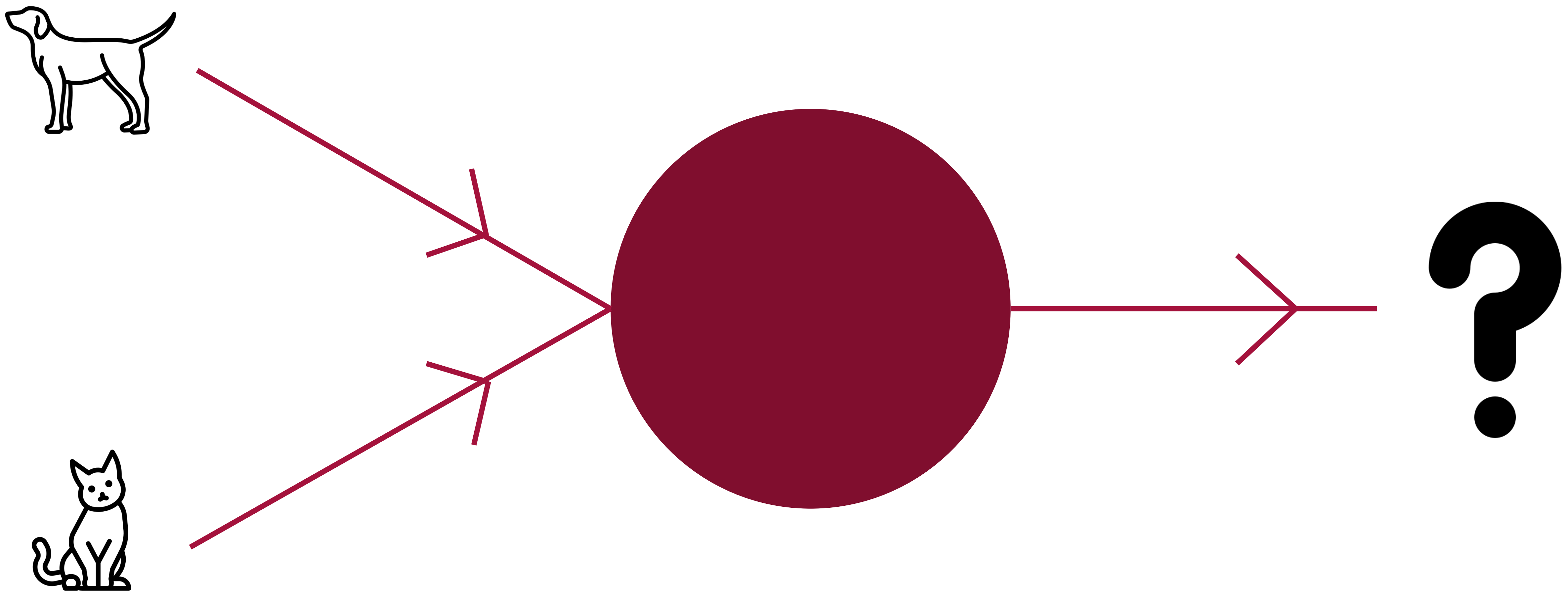
Formas de resolver un problema
ML
ML
Supervisado
Semi Supervisado
No Supervisado
Supervisado
Semi Supervisado
No Supervisado

Arquitectura



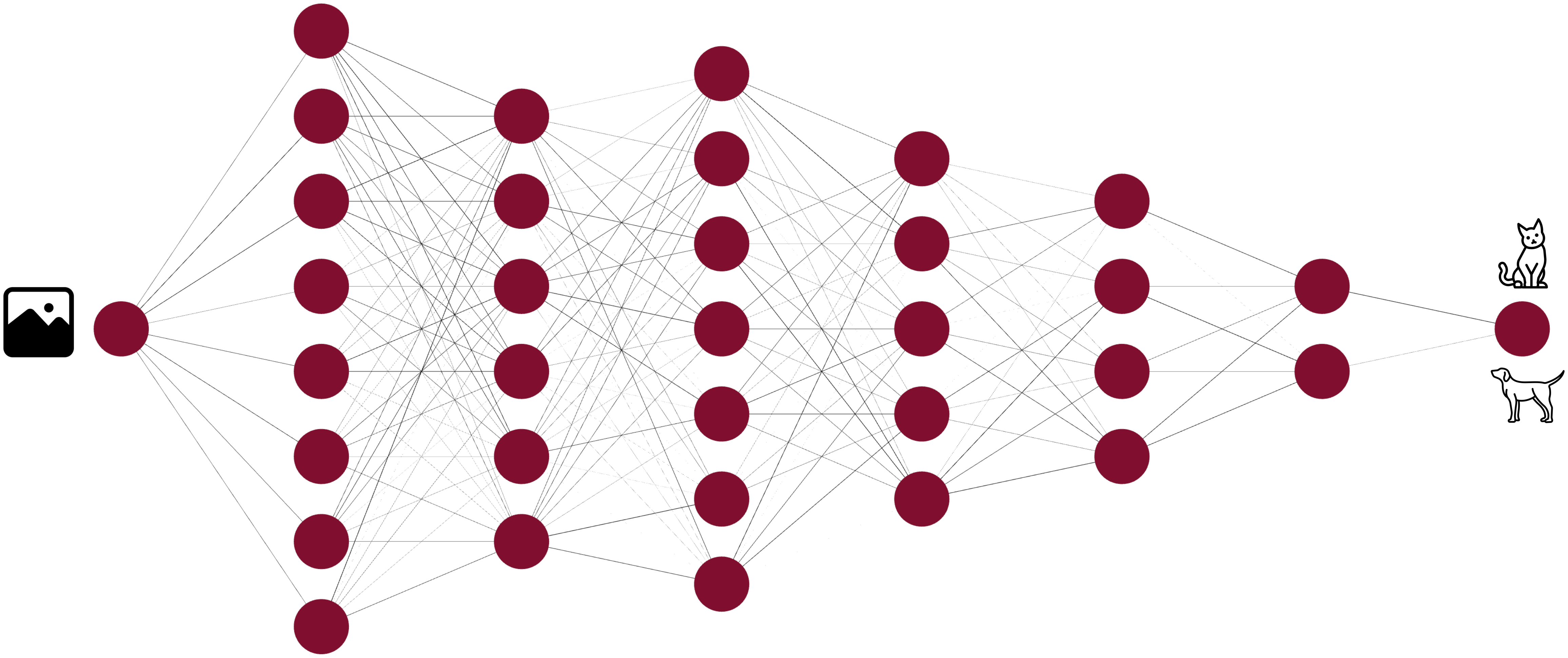
Una neurona recibe información y devuelve un resultado

*Tomado de Meor Amer (2022)

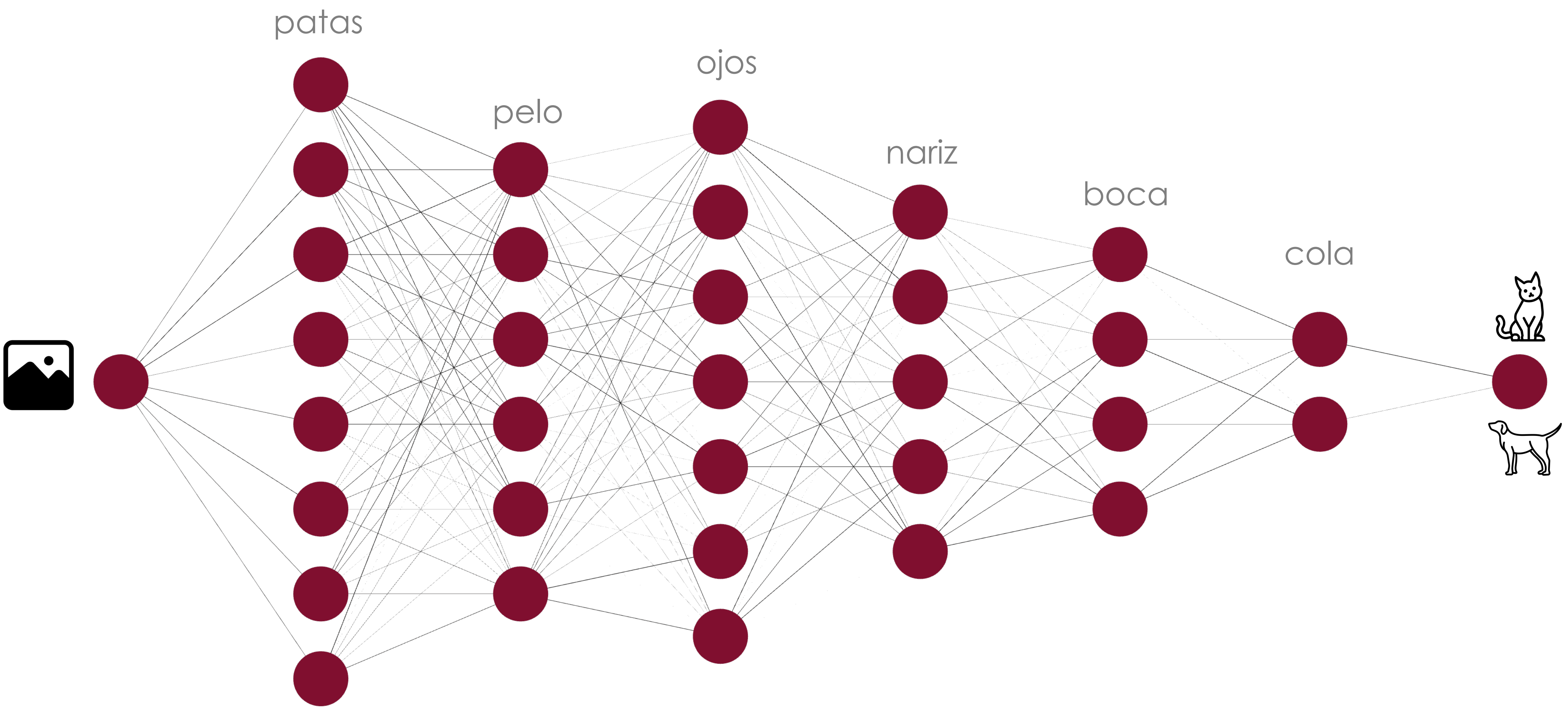


Una neurona recibe información y devuelve un resultado

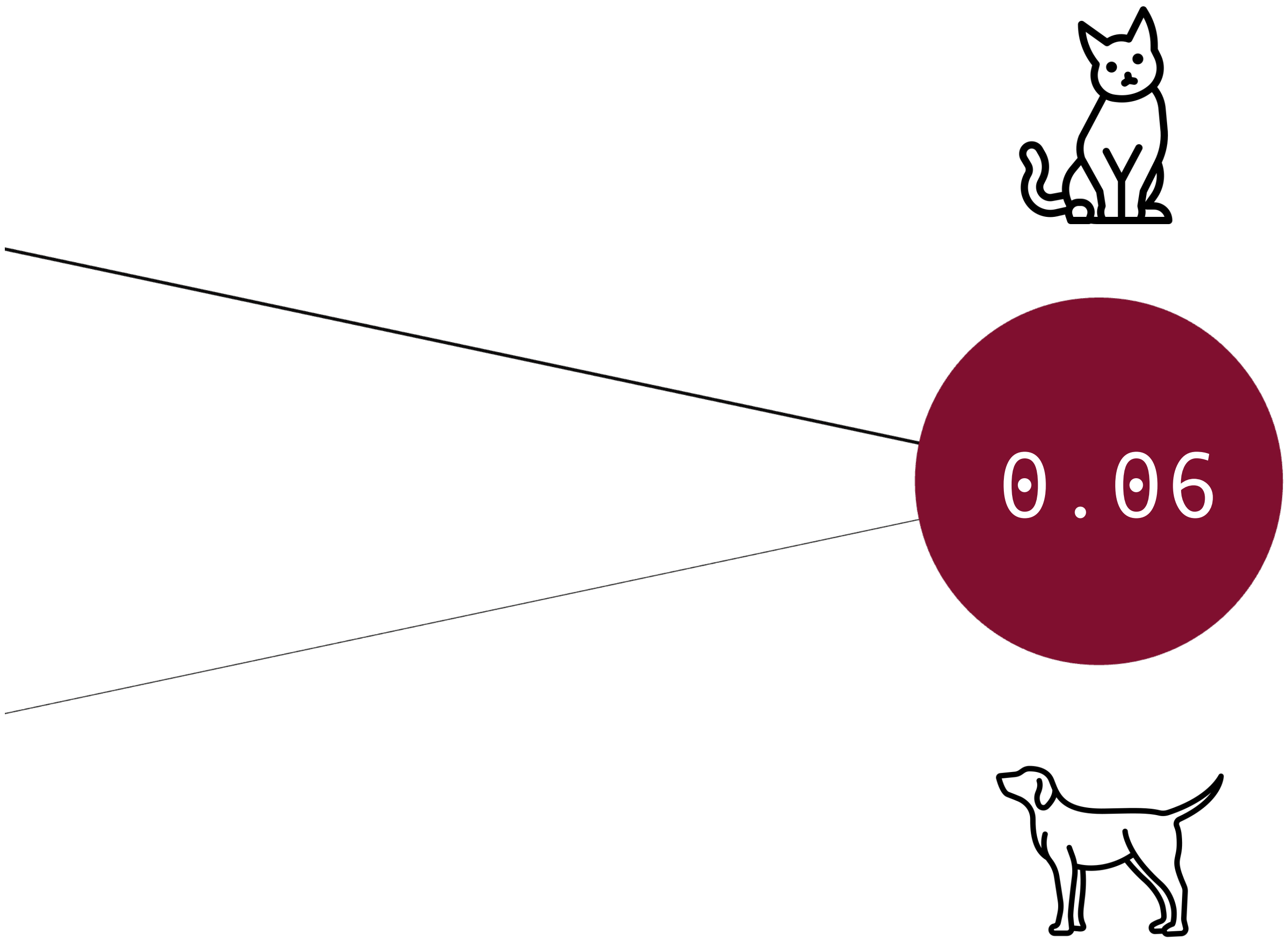
*Tomado de Meor Amer (2022)



Cada capa es una característica



Cada capa es una característica



No hay certeza absoluta, solo probabilidades

The New York Times Magazine

ON MEDICINE

This Cat Sensed Death. What if Computers Could, Too?

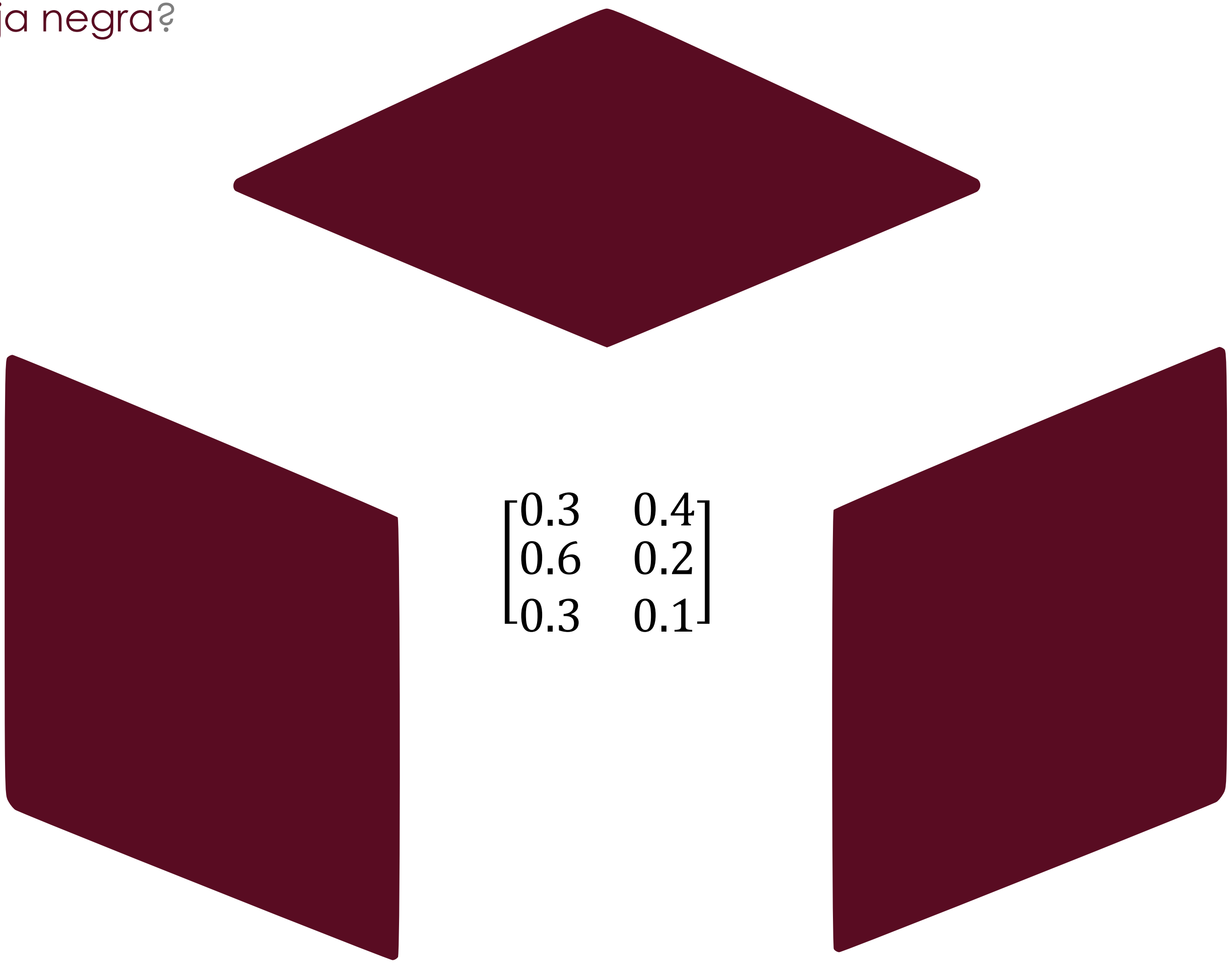
"[A deep learning system] learns, but it cannot tell us why it has learned; it assigns probabilities, but it cannot easily express the reasoning behind the assignment. Like a child who learns to ride a bicycle by trial and error and, asked to articulate the rules that enable bicycle riding, simply shrugs her shoulders and sails away, the algorithm looks vacantly at us when we ask, "Why?" It is, like death, another black box."

\- Siddhartha Mukherjee

¿Es la red neuronal una caja negra?

En deep learning, todo es un tradeoff

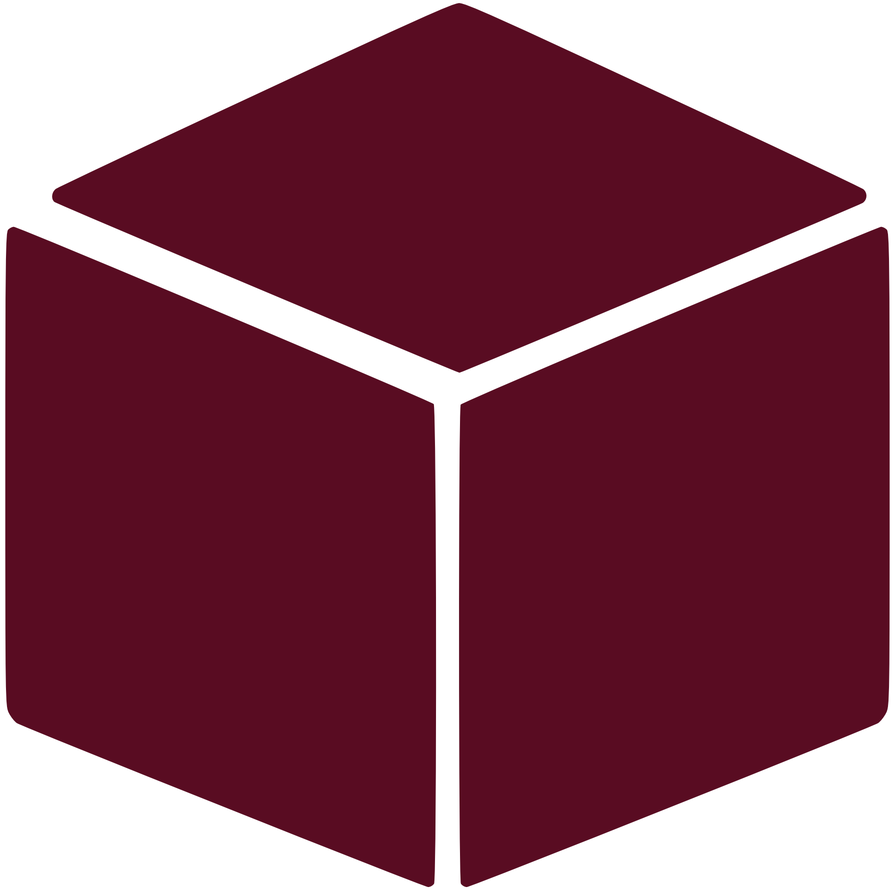
¿Es la red neuronal una caja negra?



$$\begin{bmatrix} 0.3 & 0.4 \\ 0.6 & 0.2 \\ 0.3 & 0.1 \end{bmatrix}$$

El corazon son los pesos y sesgos (weights & biases)

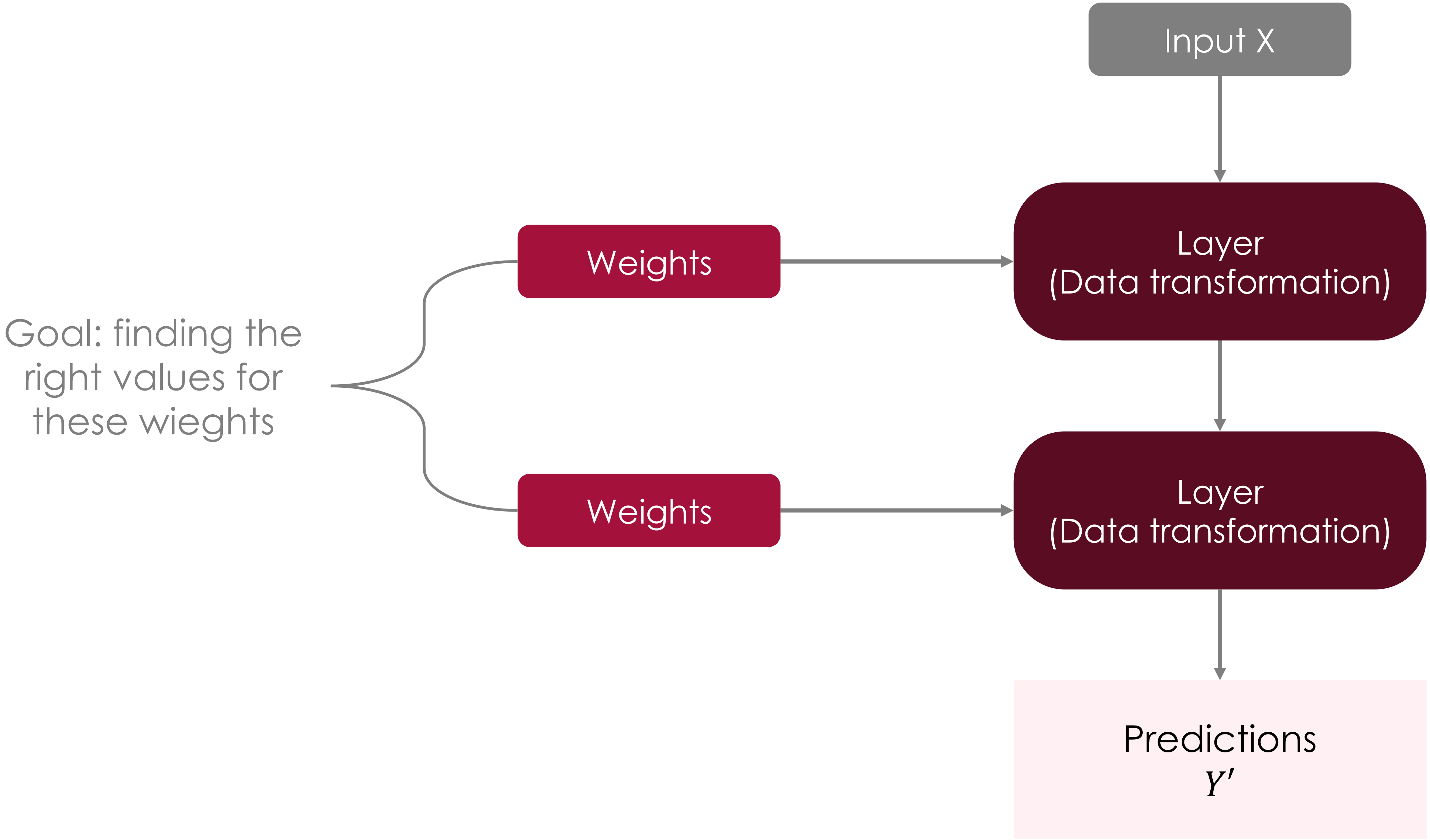
¿Es la red neuronal una caja negra?



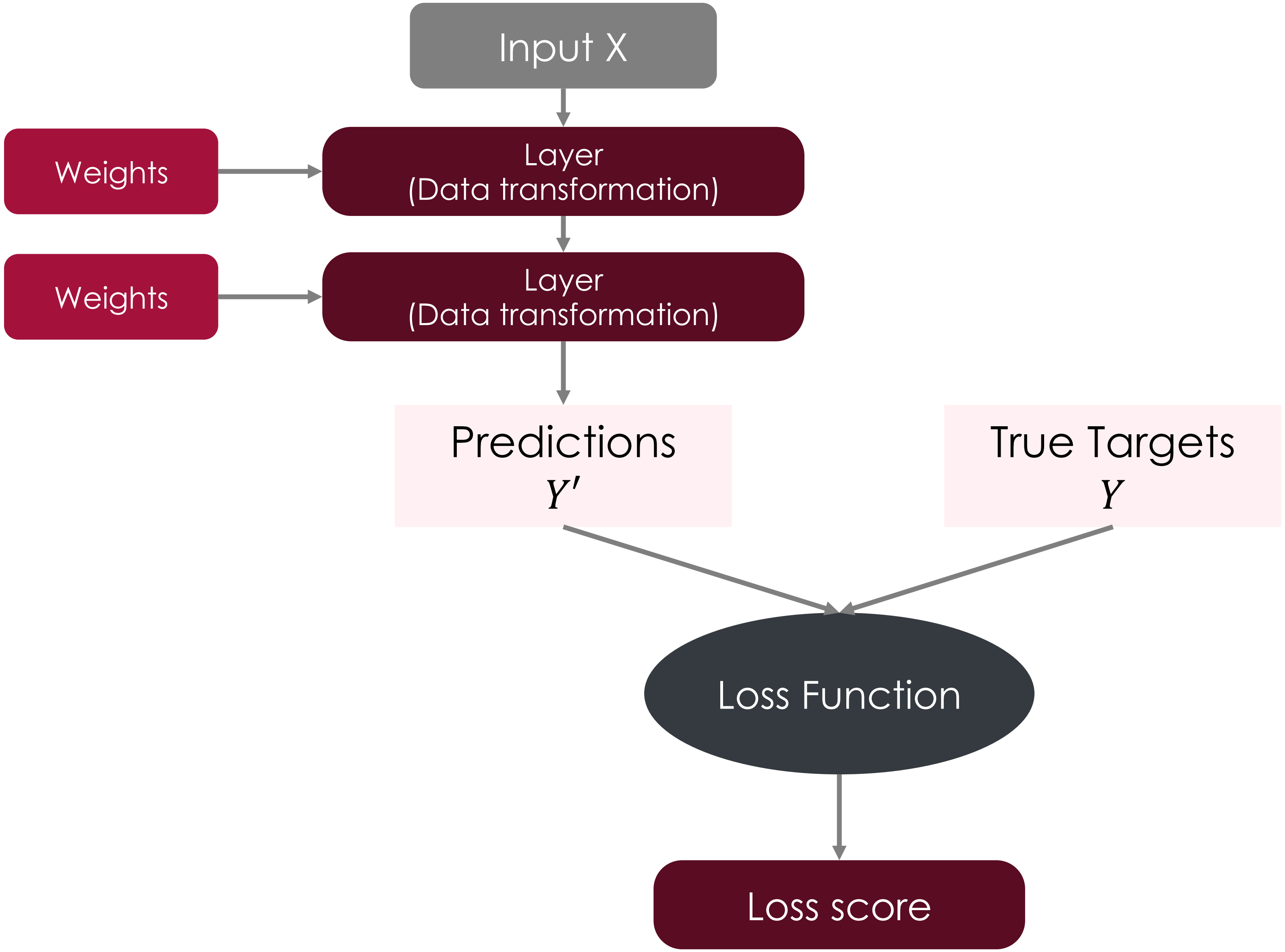
$$Y = \sum(Weights * input) + bias$$

Los pesos o weights controlan la fuerza de conexión entre dos neuronas.

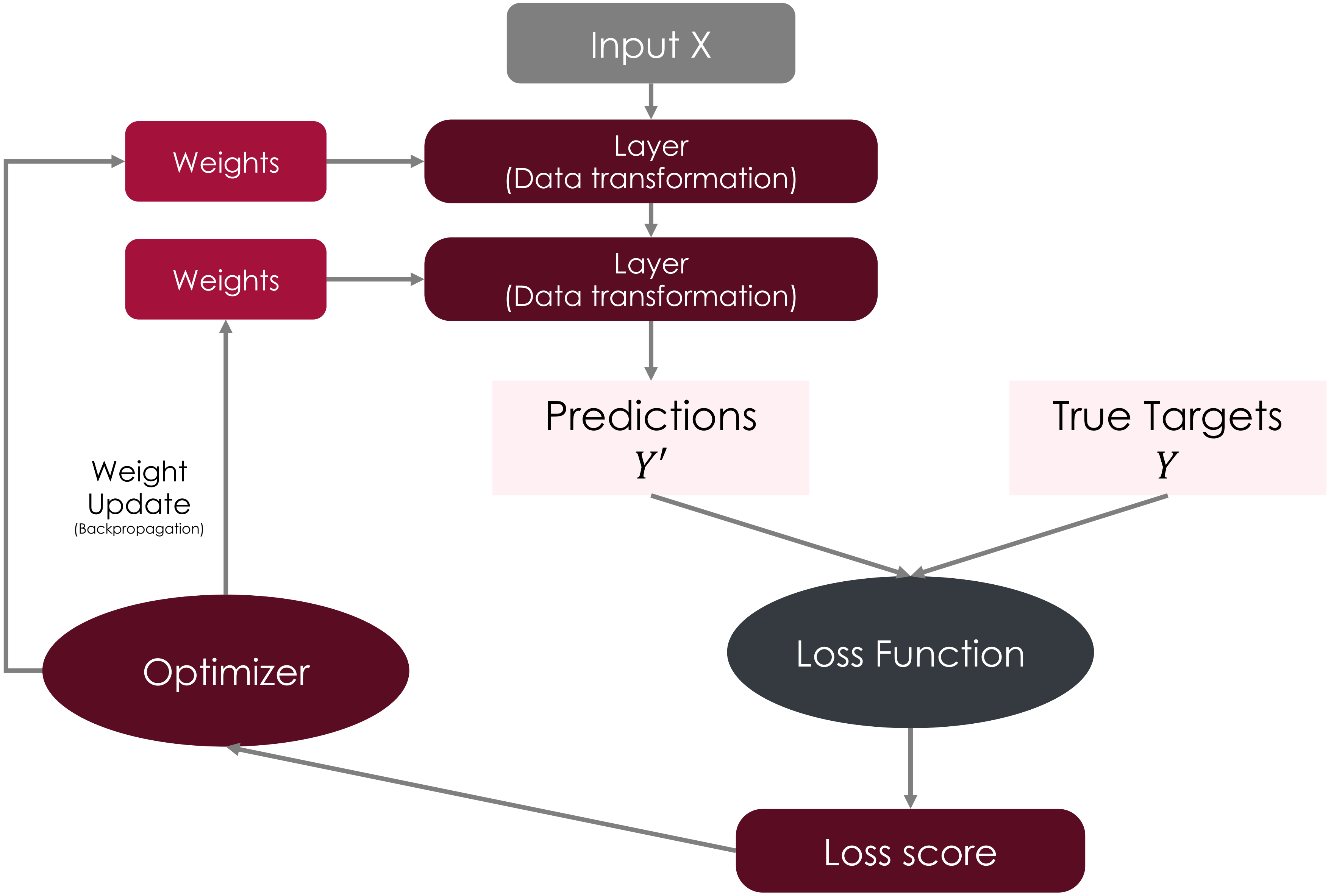
Los sesgos o bias son constantes que garantizan que a pesar de tener inputs iguales a 0, aun exista alguna activación en la neurona.



*Tomado de Francois Chollet (2021)



*Tomado de Francois Chollet (2021)



*Tomado de Francois Chollet (2021)

La hoja arrugada es el input data, y cada hoja de papel es un feature.

La red neuronal buscará la transformación de la bola de papel que pudiera desarrugarla y hacer cada hoja separable

Cuatro grandes fundamentos

Universal Approximation Theorem

Occam's Razor

Curse of Dimensionality

No Free lunch Theorem

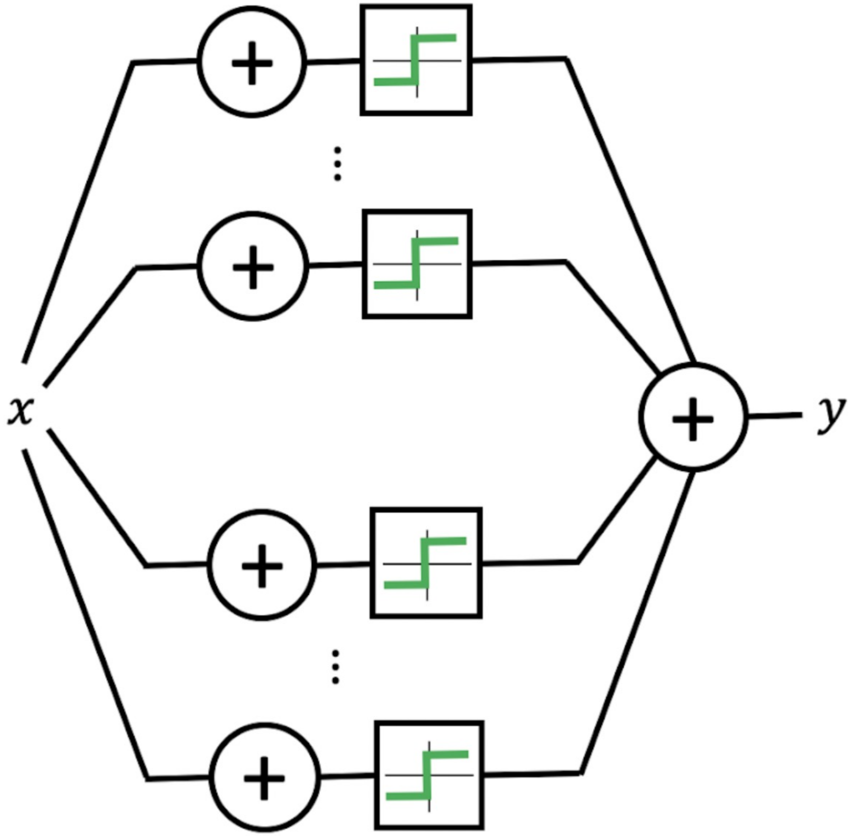
Aproximamos esa red neuronal
a través de *n* número de capas
y neuronas

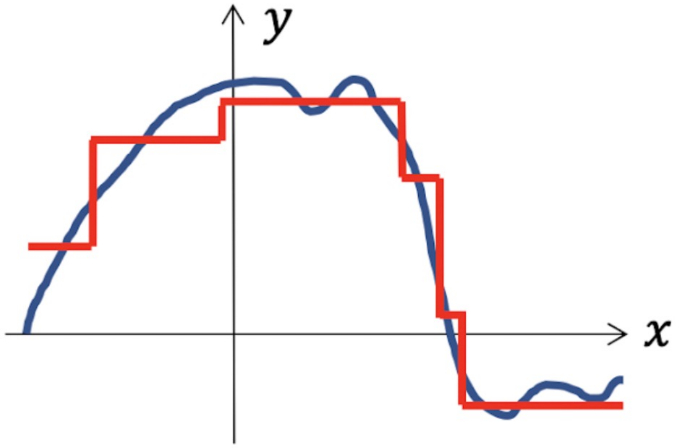
$$\underbrace{\text{Universal}}_{\text{Sin importar la función, una red neuronal puede hacer el trabajo}}\ \overbrace{\text{Approximation}}\ \text{Theorem}$$

Sin importar la
función, una red
neuronal puede
hacer el trabajo

Supongamos que una función $f(x)$ es dada y que deseamos aproximar esa función a través de un cómputo con una precisión $\varepsilon > 0$

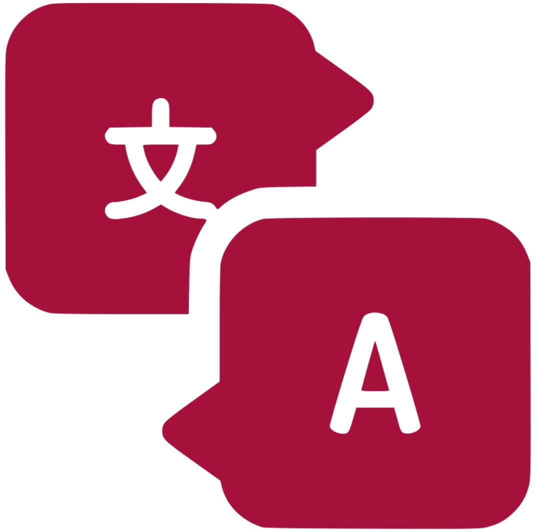
Podemos garantizar que usando suficientes neuronas ocultas siempre podemos encontrar una red neuronal para la cual su salida $g(x)$ logre satisfacer $|g(x) - f(x)| < \varepsilon$ para todo x



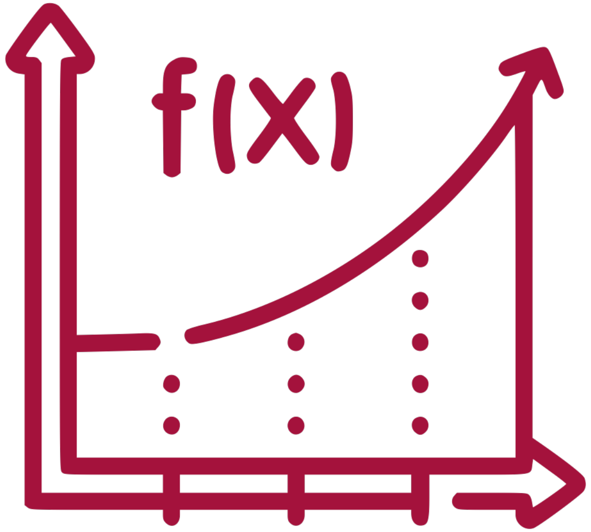


*Tomado de Michael M. Bronstein (2022)

Consideraciones



No porque sepamos que existe una red neuronal que pueda traducir del español al chino, significa que tengamos buenas técnicas para poder constuir esa red neuronal



Todas las funciones deben de ser continuas para poder aproximarlas

(en la práctica es dificil encontrar funciones discontinuas)

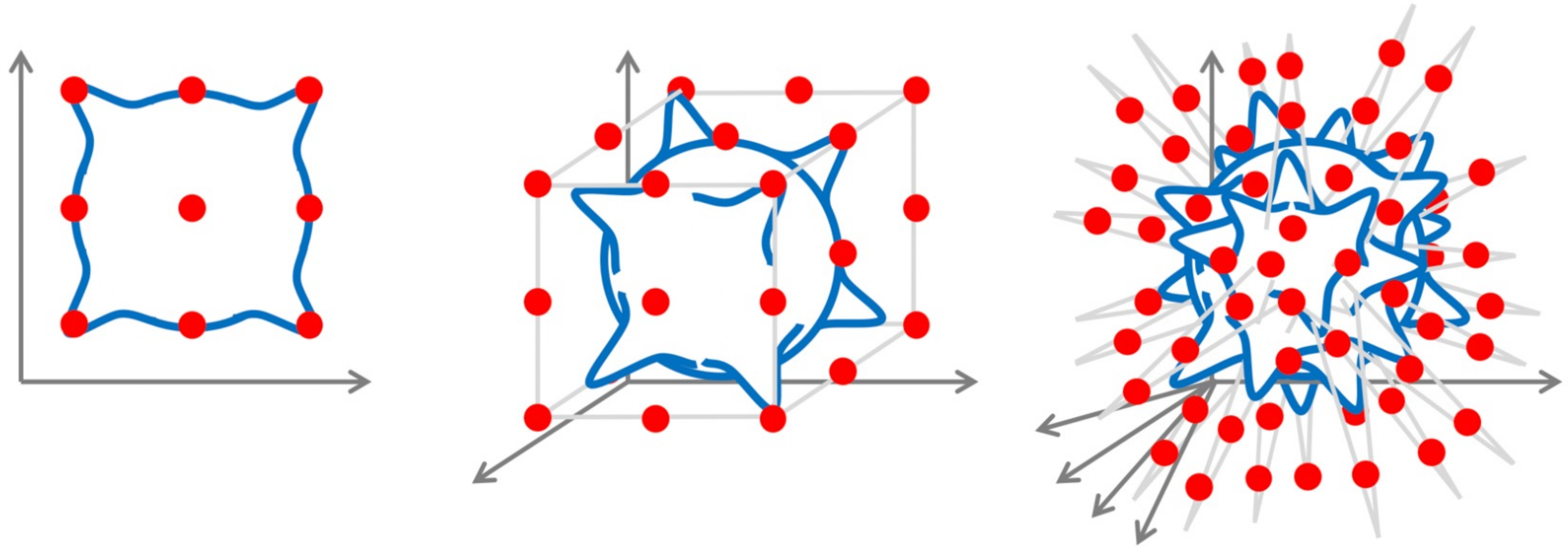
Las entidades no deben de ser multiplicadas más allá de lo necesario

Entia non sunt multiplicanda praeter necessitatem

Es inútil hacer con más cosas lo que se puede hacer con menos

Frustra fit per plura quod potest fieri per pauciora

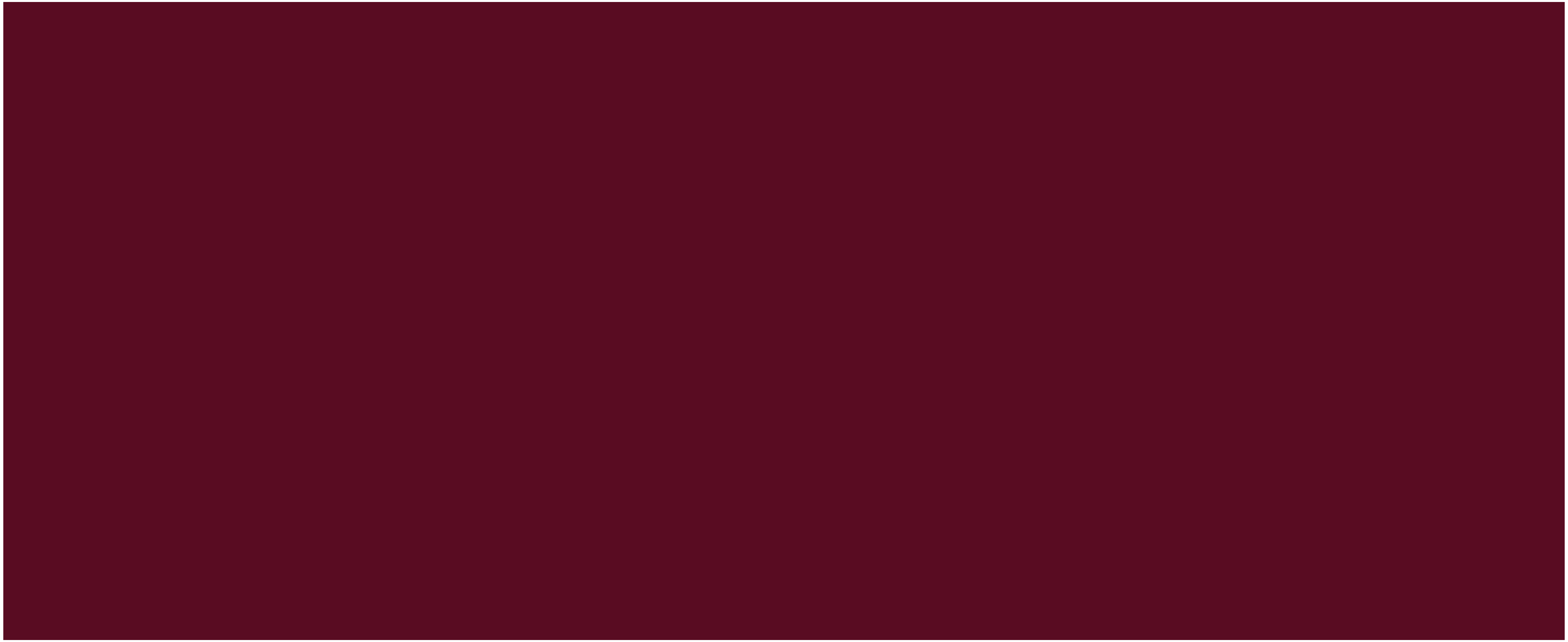
The curse of dimensionality



*Tomado de Michael M. Bronstein (2022)

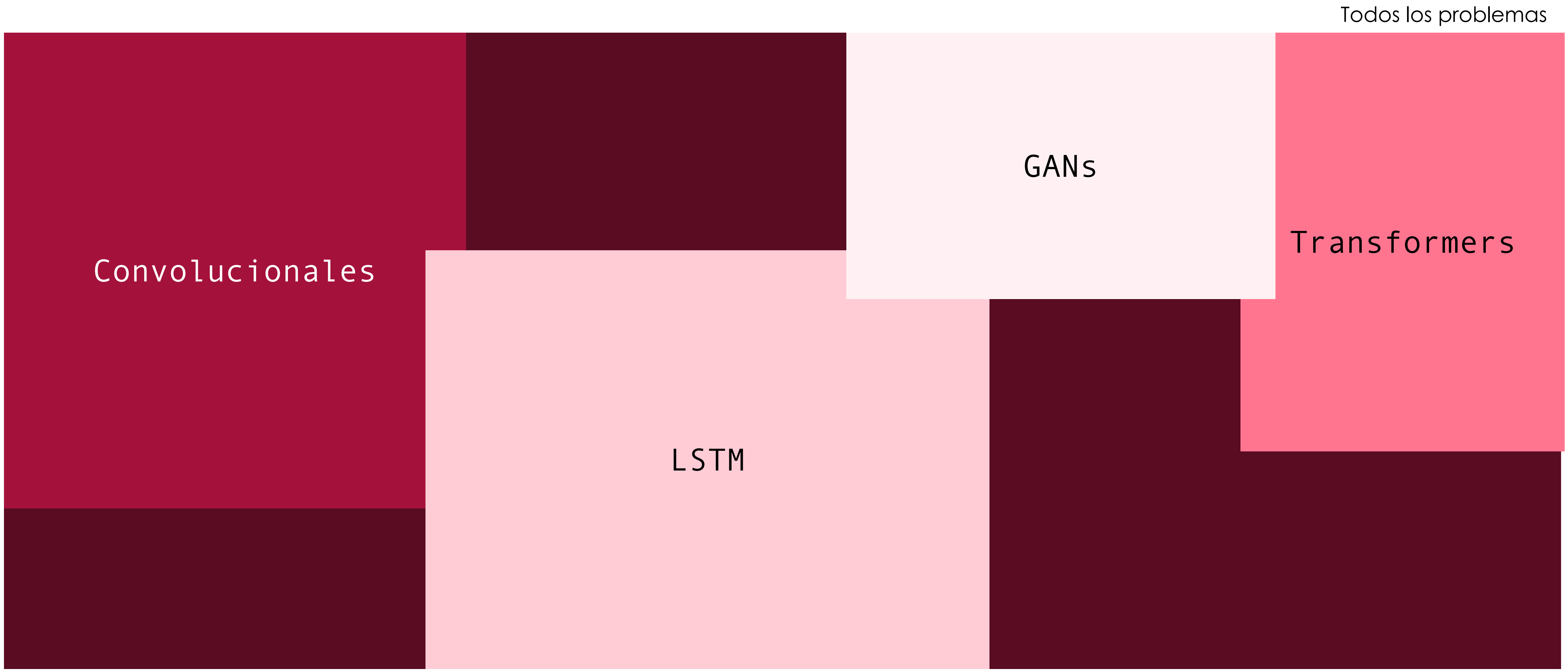
No free lunch theorem

Todos los problemas

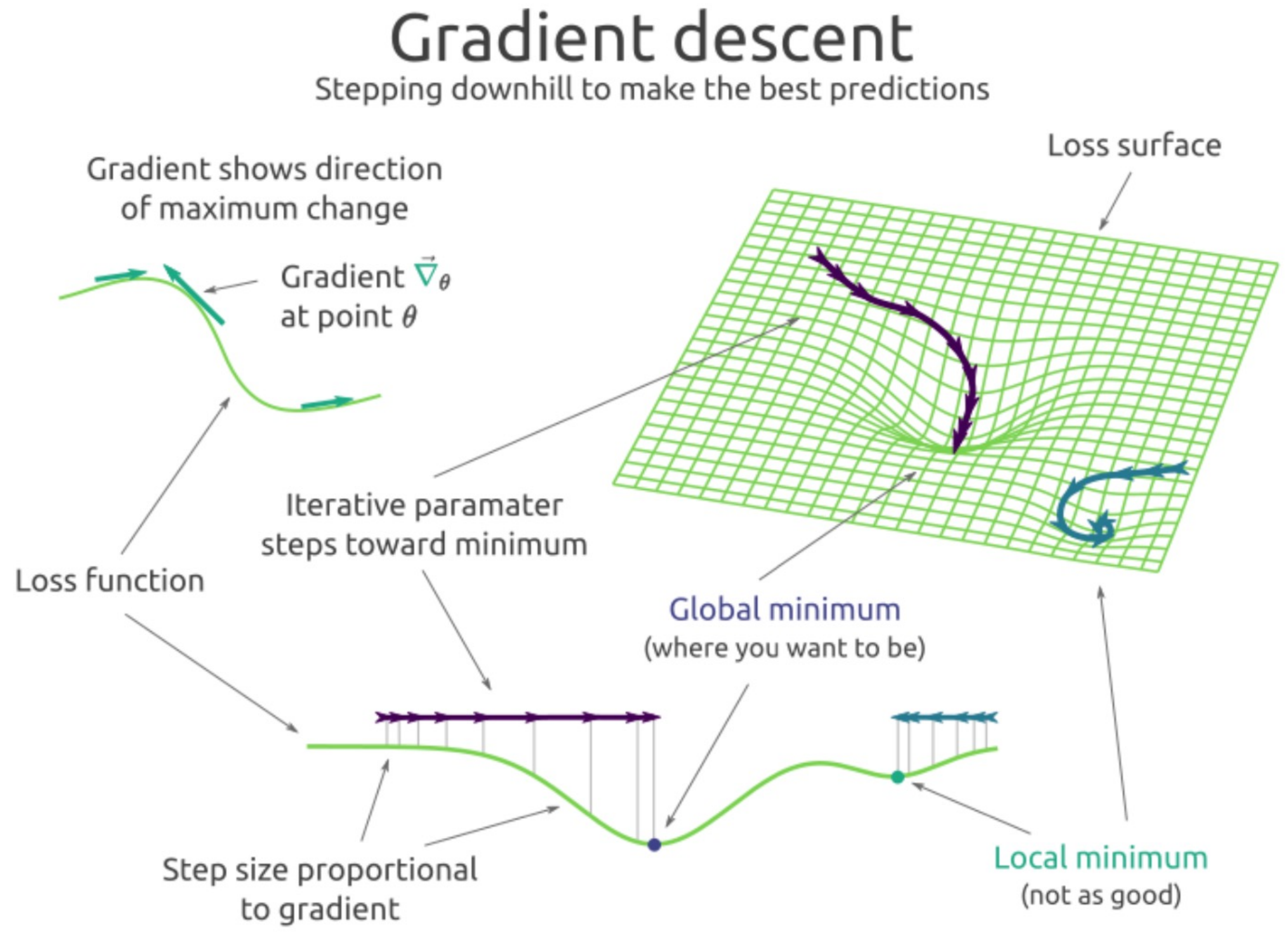


Ningún algoritmo es el mejor / no existe algoritmo universal…

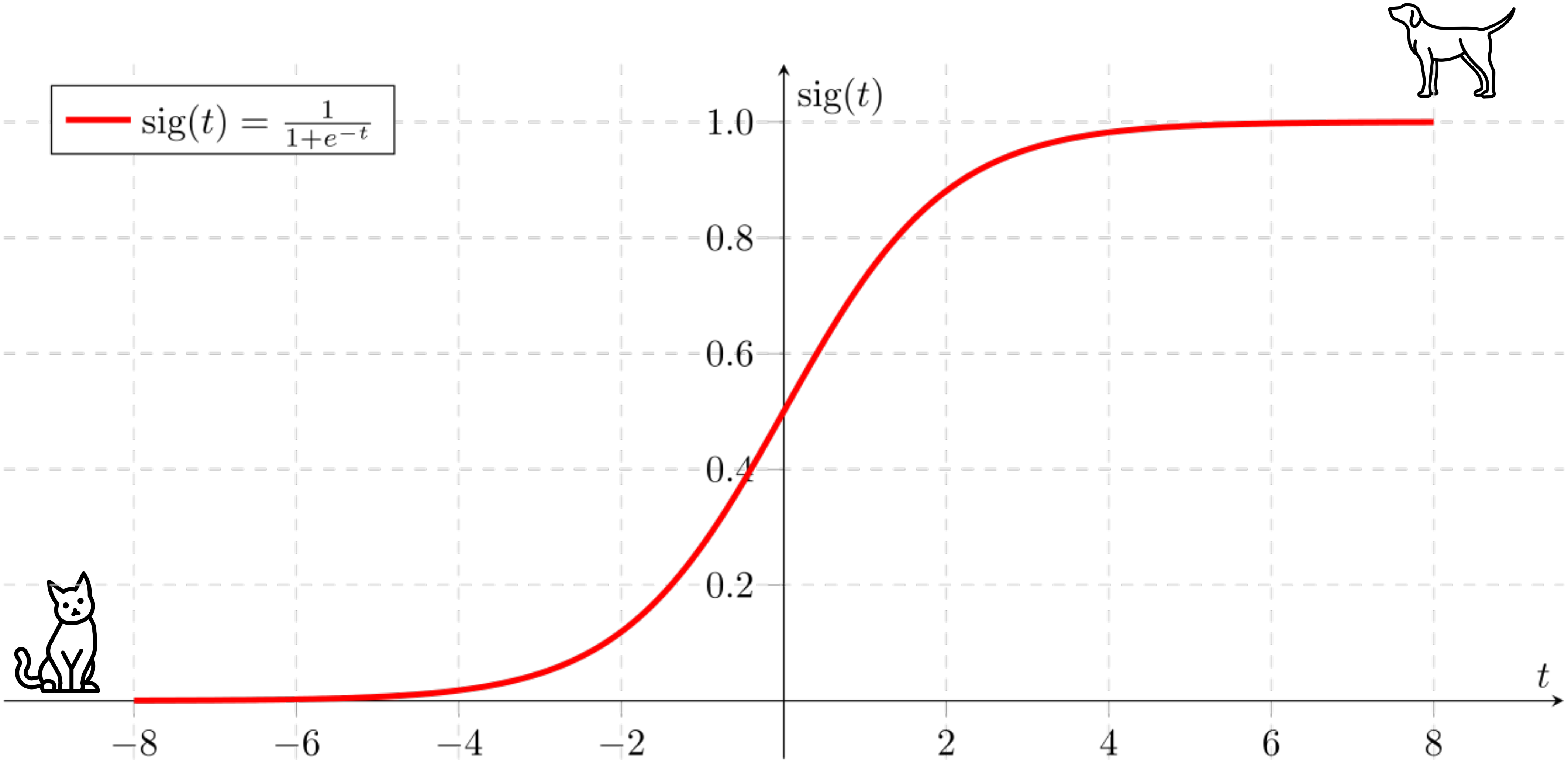
No free lunch theorem



...pero JAMAS se hace un solo modelo



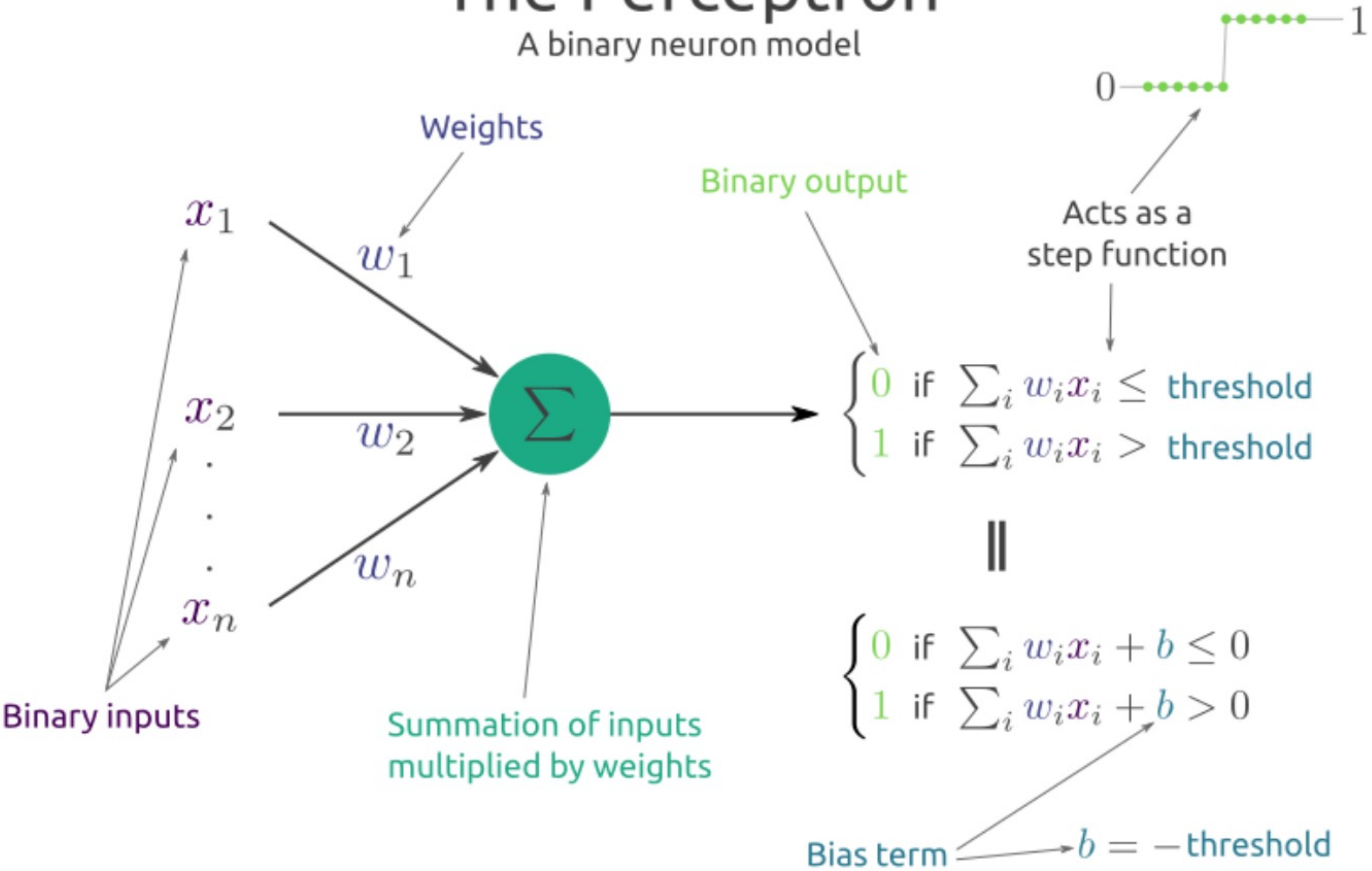
*Tomado de Roy Keyes (2022)

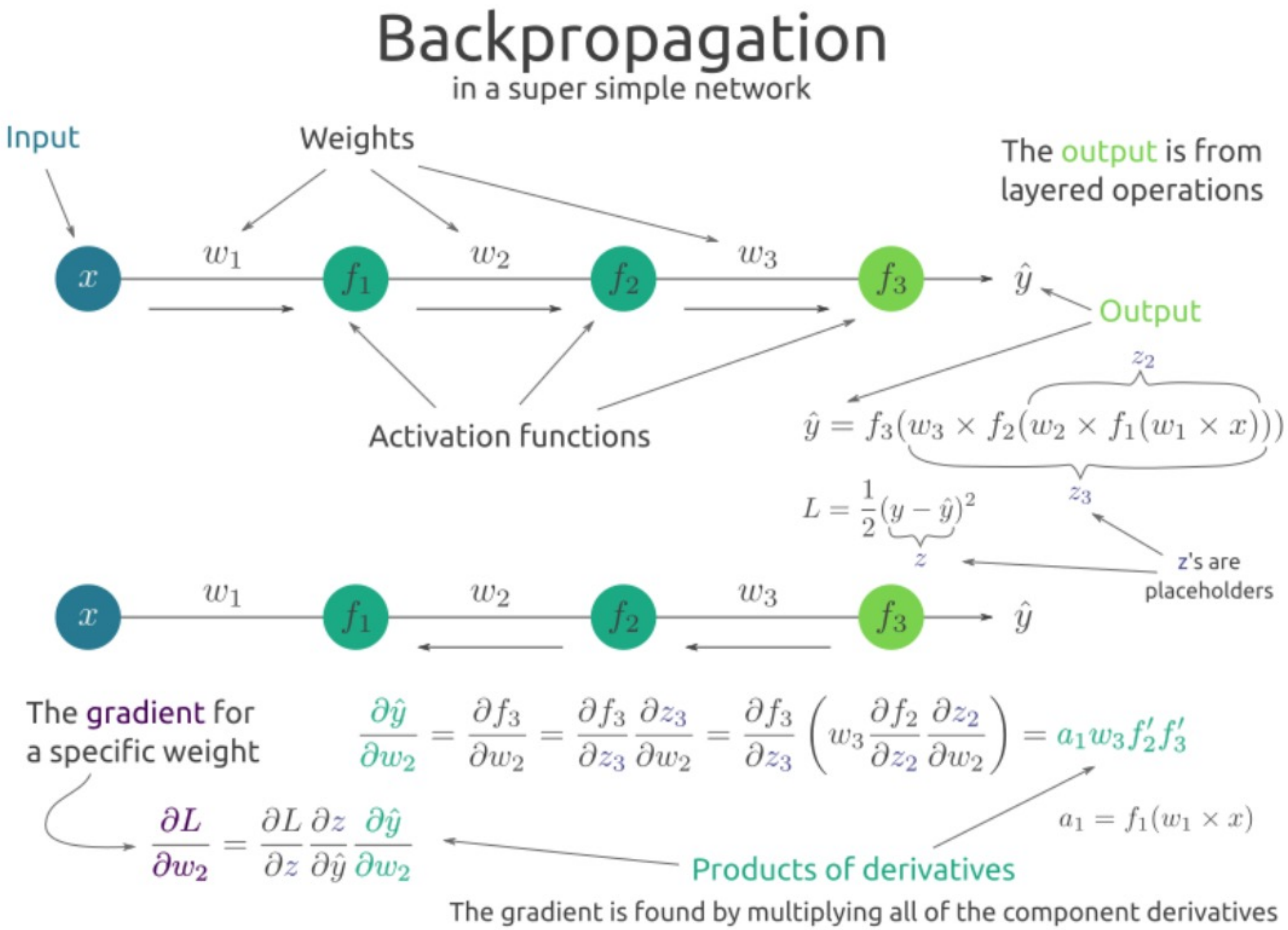


INSTITUTO HUMAN DELTA

The Perceptron

A binary neuron model





*Tomado de Roy Keyes (2022)

Backpropagation

The gradient in a single back pass

A change in weight here affects all downstream operations

$w_i + \Delta$

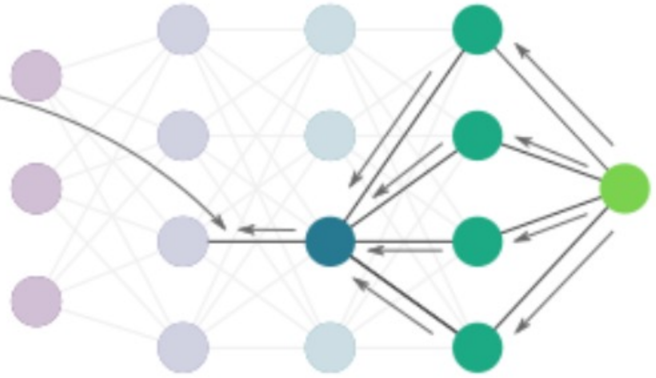
$$f_i(w_i+\Delta) = f_i(w_i) + \Delta \frac{\partial f_i}{\partial w_i}$$

Derivatives

$$\hat{y}(x, w_i+\Delta) = \hat{y}(x, w_i) + \Delta \frac{\partial \hat{y}}{\partial w_i}$$

$$L(x, w_i+\Delta) = L(x, w_i) + \Delta \frac{\partial L}{\partial w_i}$$

You can build the derivative anywhere in the network by knowing all of the downstream derivatives, thanks to the chain rule

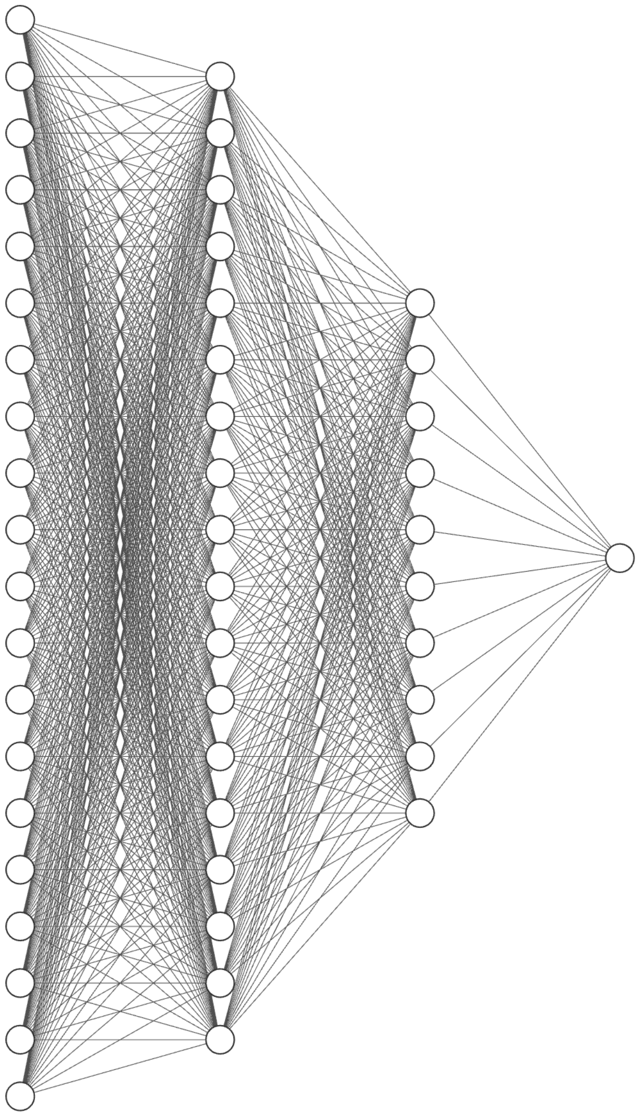


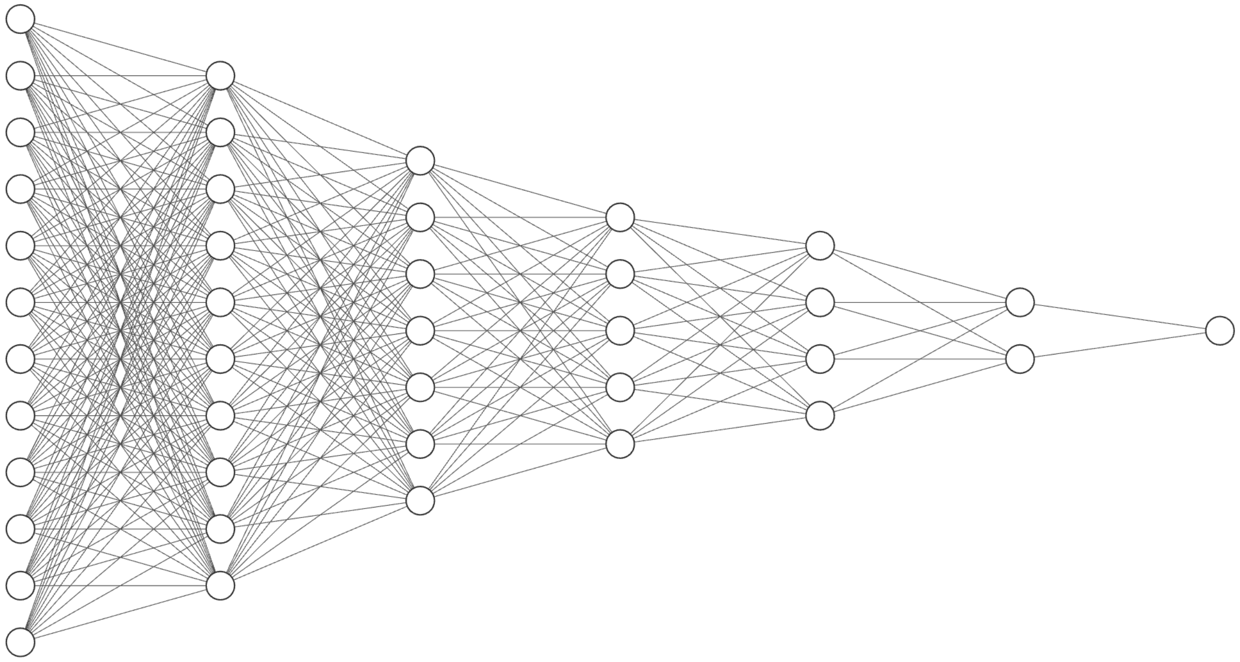
$$\frac{d}{dx} g(f(x)) = \frac{dg}{df} \frac{df}{dx}$$

$$\frac{\partial L}{\partial w_2^{(3)}} = (y - f_4) a_1^{(3)} \partial f_2^{(3)} \partial f_4 \sum_{j=1}^{4} w_3^{(j)} w_4^{(j)} \partial f_3^{(j)}$$

*Tomado de Roy Keyes (2022)

Width problem





¿Por qué las redes neuronales computan una función y no otra?

El eterno problema del comportamiento macroscópico de sistemas con múltiples componentes

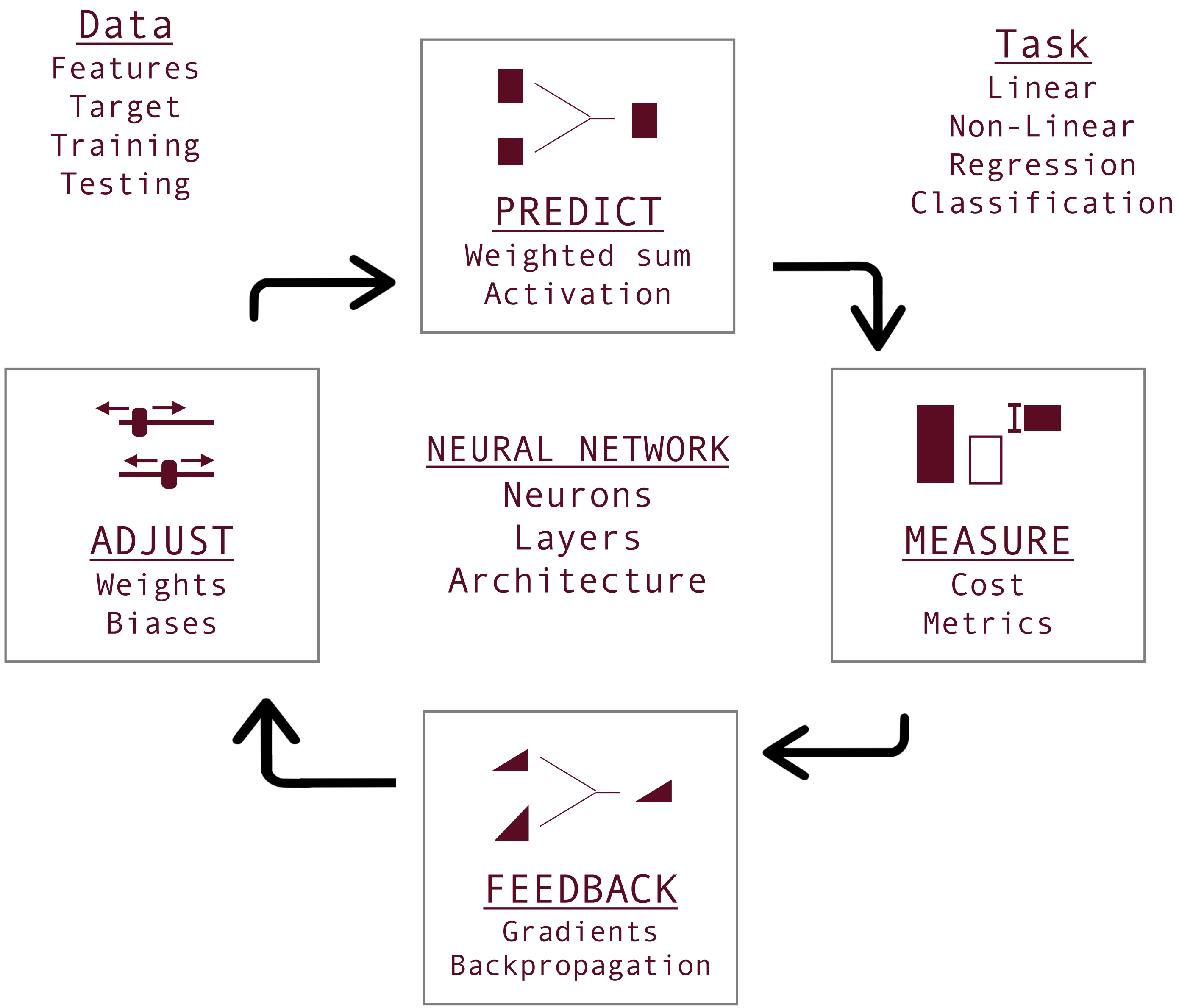
Teoría vs práctica

Es la razón de porqué se necesita una teoría y por la que las redes neuronales suelen fallar.

"The point of theoretical calculations is to connect measurable outcomes or observables directly to the fundamental underlying constants or parameters that define the theory. This perspective also implies a tradeoff between the predictive accuracy of a model and its mathematical tractability, and the former must take precedence over the latter for any theory to be successful: a short tether from theory to physical reality is essential."

-Daniel A. Roberts

Loop



*Tomado de Meor Amer (2022)

QUESTION:

Amanda wants to buy the perfect shirt.

She knows somebody at a warehouse, so she pulls a favor one Sunday afternoon, and to her amusement, they have thousands of different shirts.

If Amanda only cared about a size that fits her, and there are four different sizes, she could buy one right after trying four shirts.

But if Amanda also cared about the color, and there are ten different colors, Amanda would have to try $4 \times 10 = 40$ combinations to find the perfect shirt.

What would happen if she also wanted to take the material into account? Assuming three different types, Amanda could only find the perfect shirt after trying $4 \times 10 \times 3 = 120$ shirts.

All of a sudden, Amanda is overwhelmed.

Which of the following ideas explains what's happening to Amanda?

- Occam's Razor
- No Free Lunch Theorem
- Curse of Dimensionality
- Universal Approximation Theorem

- Occam’s Razor
- No Free Lunch Theorem
- **Curse of Dimensionality**
- Universal Approximation Theorem

Explanation:

Occam's Razor is the idea that, given two solutions with similar characteristics, the simplest and most direct one is the correct answer. This idea is unrelated to Amanda's problem, so it's not the right choice.

The No Free Lunch Theorem implies that no single algorithm is universally the best-performing algorithm for all problems in machine learning. This idea is also unrelated to what's happening here.

The Universal Approximation Theorem states that a feed-forward network with a single hidden layer containing a finite number of neurons can approximate any continuous function. Not a correct choice either.

Finally, the **Curse of Dimensionality** refers to various phenomena when working with data in high-dimensional spaces. It states that, as the dimensionality of the data increases, the amount of data needed to train a learning algorithm grows exponentially.

That's what's happening to Amanda: as she wants to consider more shirt attributes, it becomes more problematic to pick the perfect shirt.

Referencias

- Foto de portada: Mahdi Bafande
- Meor Amer, (2022), **“A visual Introduction to Deep Learning”**
- Roy Keyes, (2022), **“Deep Learning”**
- Bronstein, Michael M., et al. **“Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges”.** arXiv, el 2 de mayo de 2021. arXiv.org, http://arxiv.org/abs/2104.13478.

GRACIAS

emilio.sandpal@gmail.com

milioe