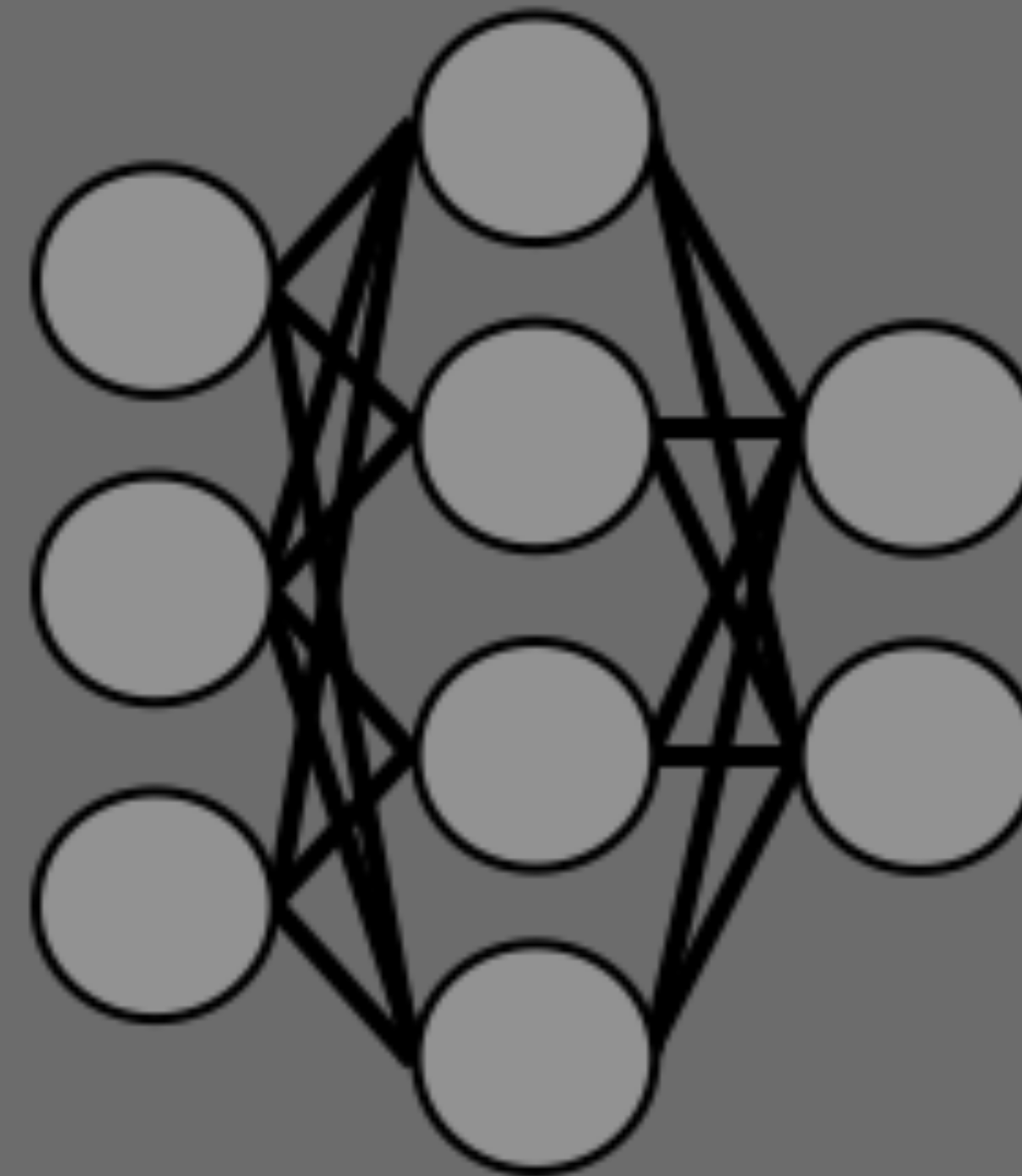


1. Introducción

Redes neuronales

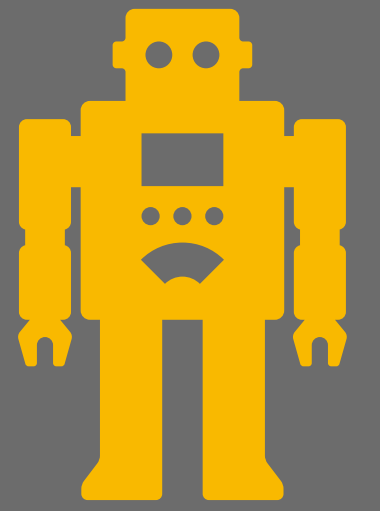


Sergio Fernández, Fernanda Sánchez.

22 de Septiembre de 2021

Paradigmas de aprendizaje automático

“El aprendizaje no se reduce a supervisado y no supervisado”



- Aprendizaje supervisado.
- Aprendizaje no supervisado.
- Aprendizaje semi-supervisado.
- Aprendizaje auto-supervisado (*self-supervised learning*).
- Aprendizaje por refuerzo (*reinforcement learning*).
- Aprendizaje “basado en la física” (*physics-based -informed- learning*).*



Supervisado

Consideremos

$$D = \{(x_1, y_1), \dots, (x_n, y_n)\} \subseteq \mathcal{X} \times \mathcal{Y},$$

$$(x_i, y_i) \sim \mathcal{P} \quad \forall i \in \{1, \dots, n\} \text{ (i.i.d)}$$

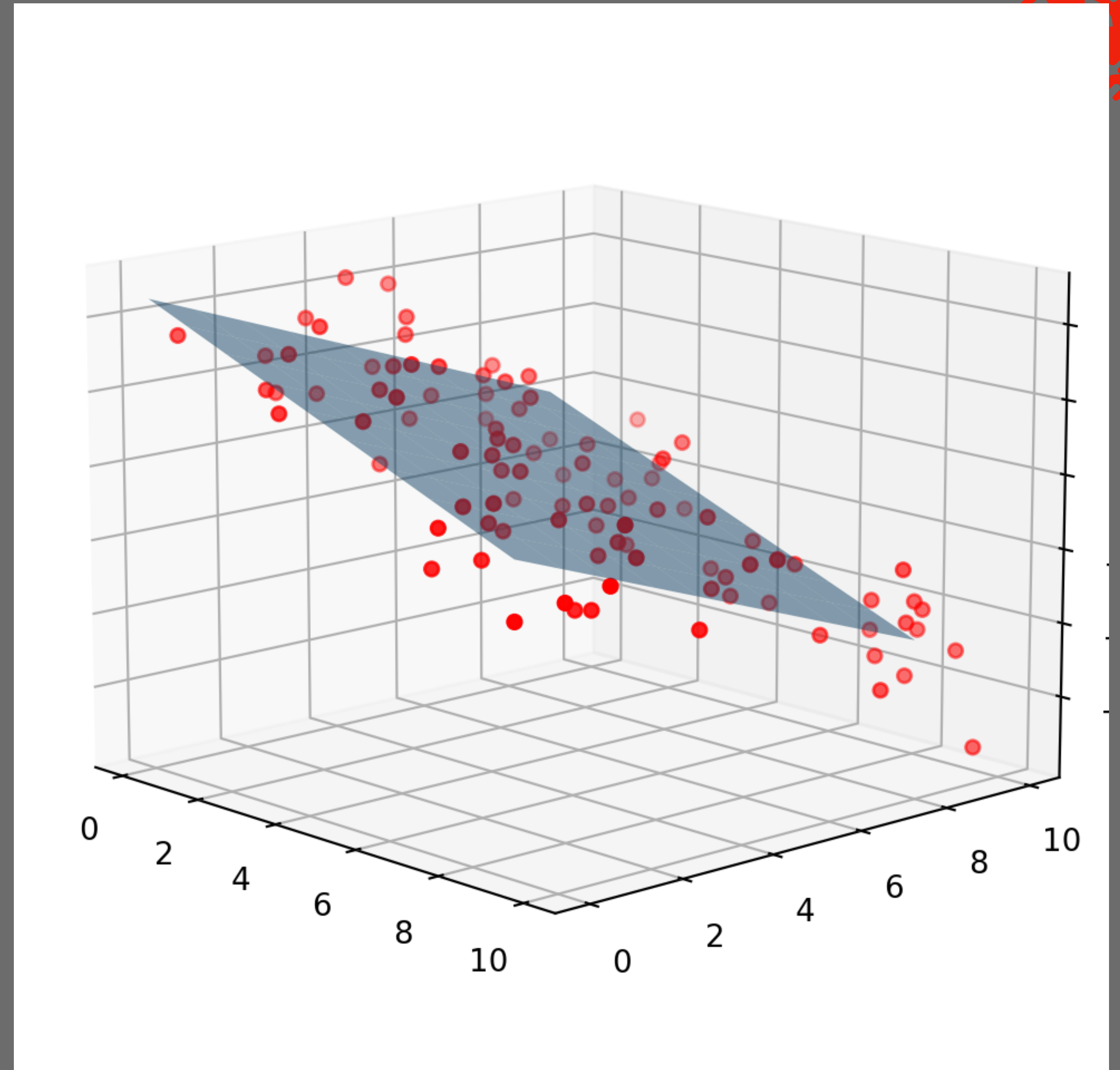
- x_i representa a los rasgos del muestra i .
- y_i es la etiqueta o valor asociado al muestra i .

Objetivo: Encontrar una hipótesis

$h \in \mathcal{H}$, $h : \mathcal{X} \rightarrow \mathcal{Y}$ tal que

$$h(x_i) \approx y_i.$$

- \mathcal{H} representa el espacio de hipótesis posibles.



No supervisado

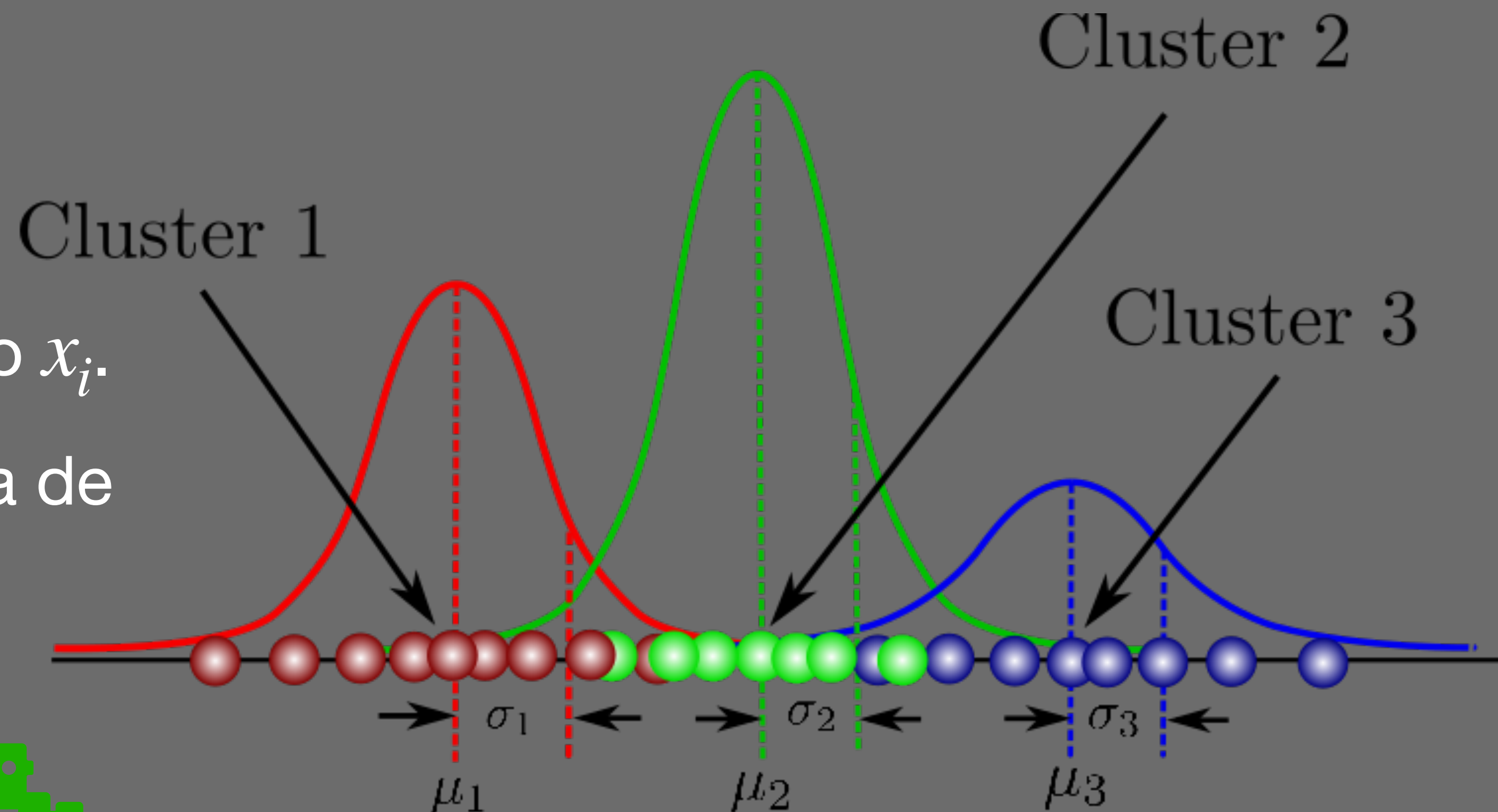
Consideremos

$$X = \{x_1, \dots, x_n\} \subseteq \mathcal{X},$$

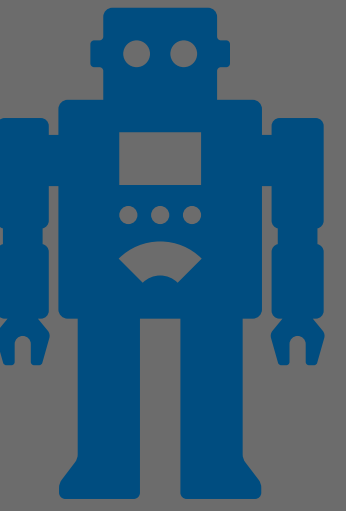
$$x_i \sim \mathcal{P} \quad \forall i \in \{1, \dots, n\} \text{ (i.i.d)}$$

- x_i representa a los rasgos del muestra i .
- No hay etiqueta o valor asociado x_i .

Objetivo: Encontrar una estructura de interesante de X .



<https://towardsdatascience.com/gaussian-mixture-models-explained-6986aaf5a95>



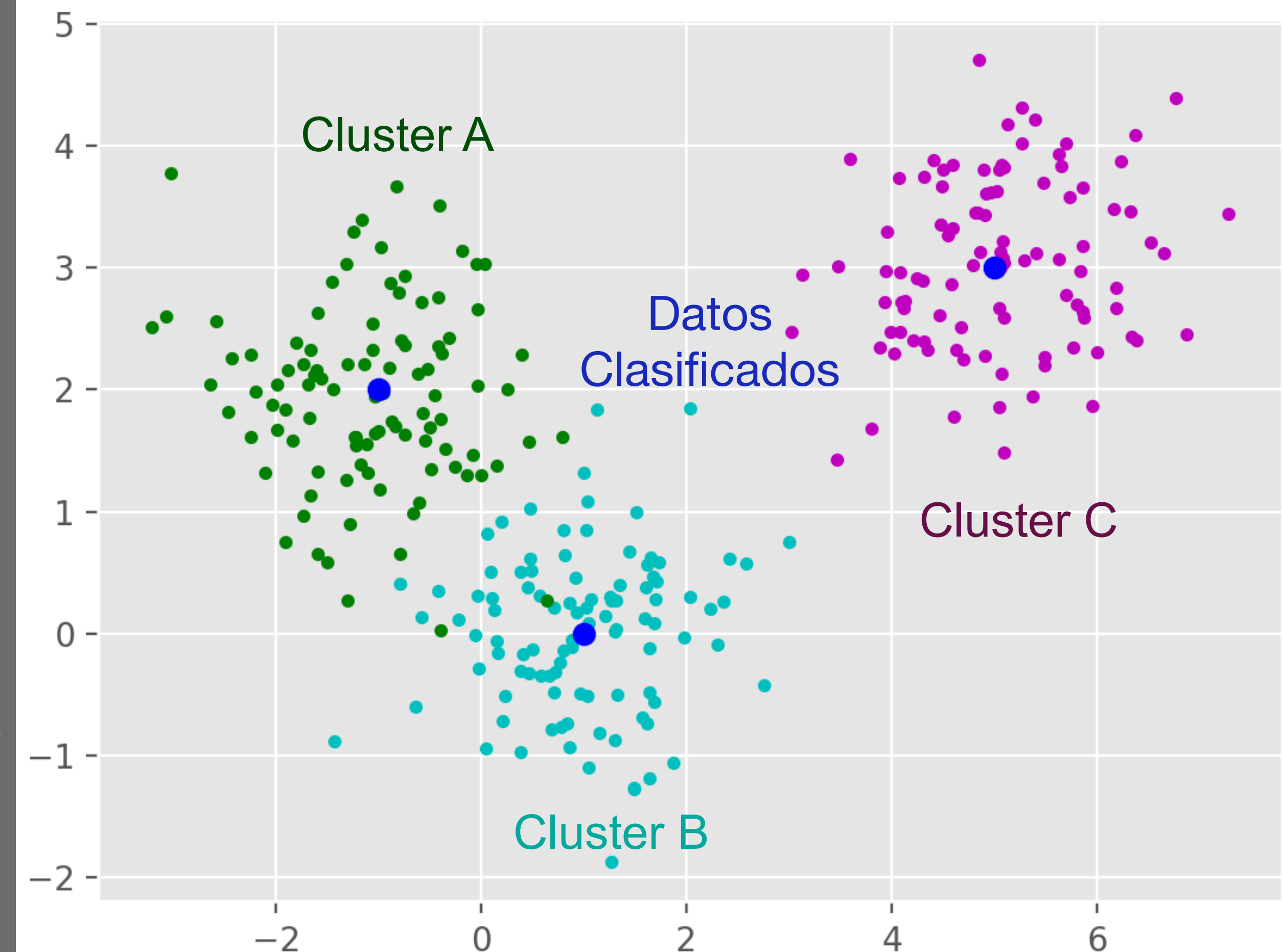
Semi-supervisado

Dividamos X en X_ℓ y X_u donde:

$$X_\ell = \{x_1, \dots, x_\ell\} \quad X_u = \{x_{\ell+1}, \dots, x_{n=u+\ell}\}$$

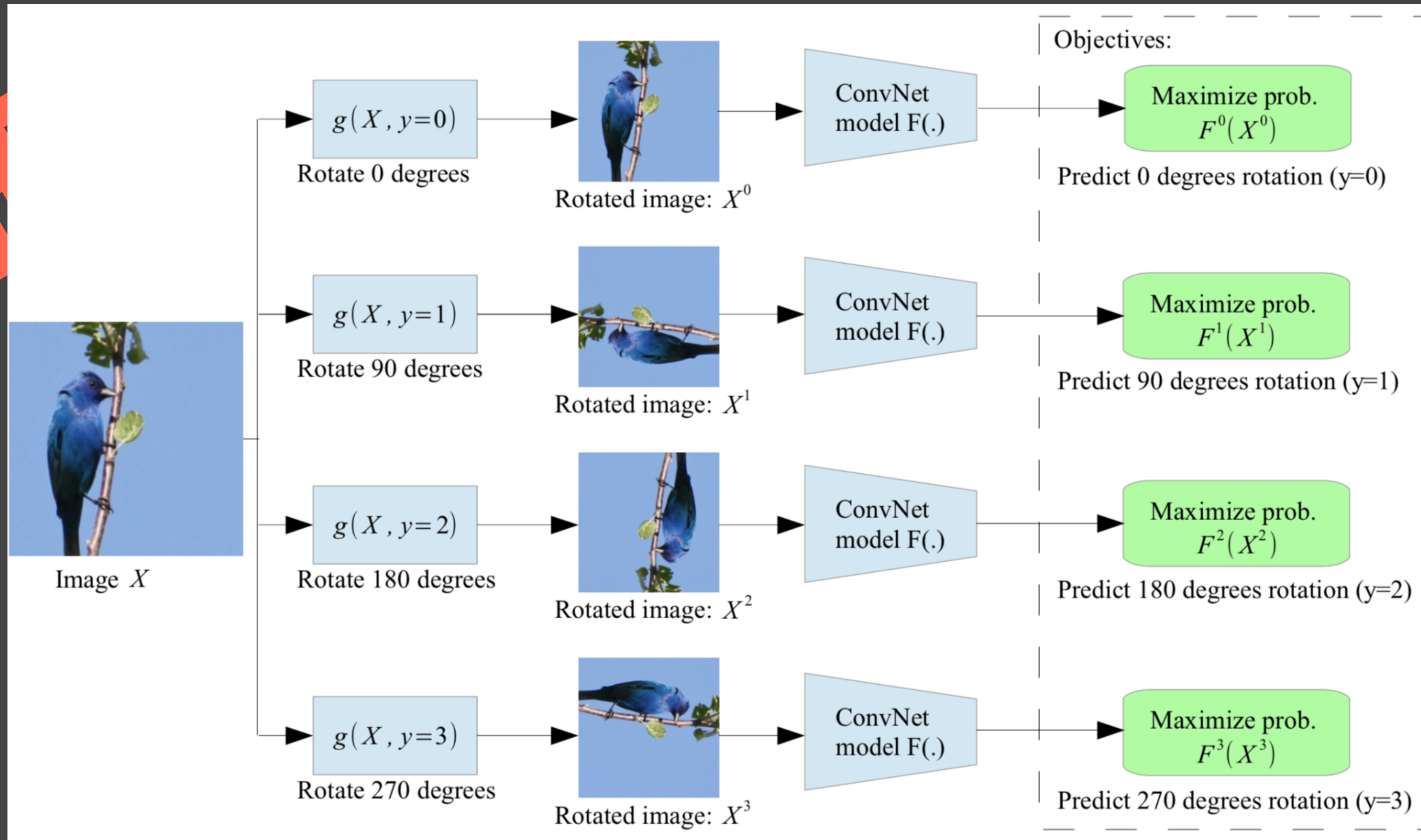
- A cada $x_i \in X_\ell$ se le asigna y_i .
- La idea es combinar métodos supervisados y no supervisados.

Objetivo: Encontrar una estructura de interesante de X_u guiados por los datos etiquetados X_ℓ .



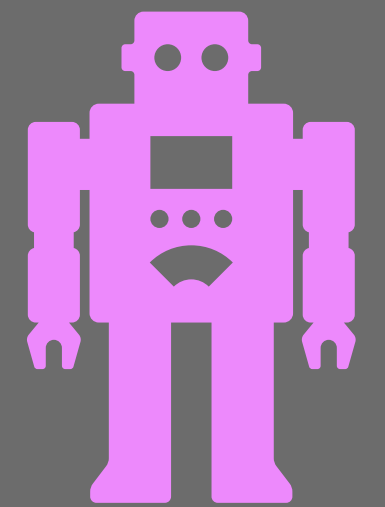
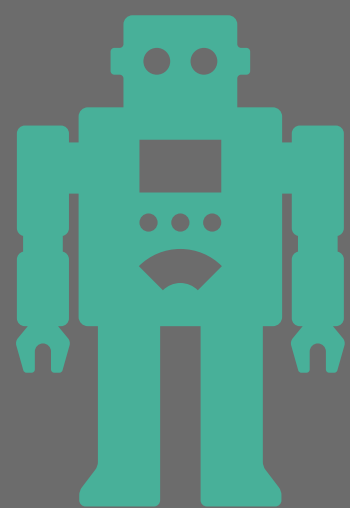
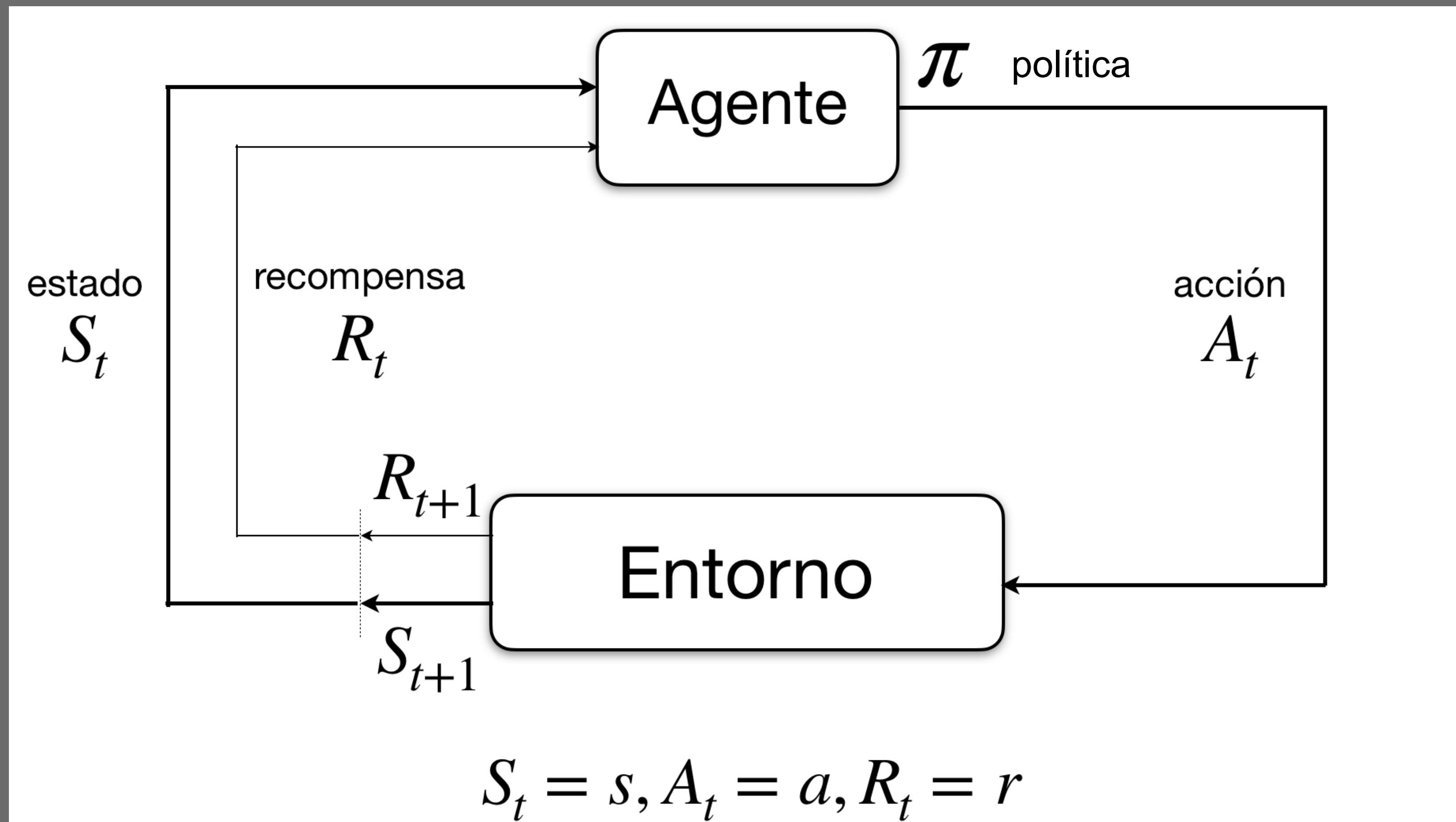
Auto-supervisado

Genera sus propias etiquetas

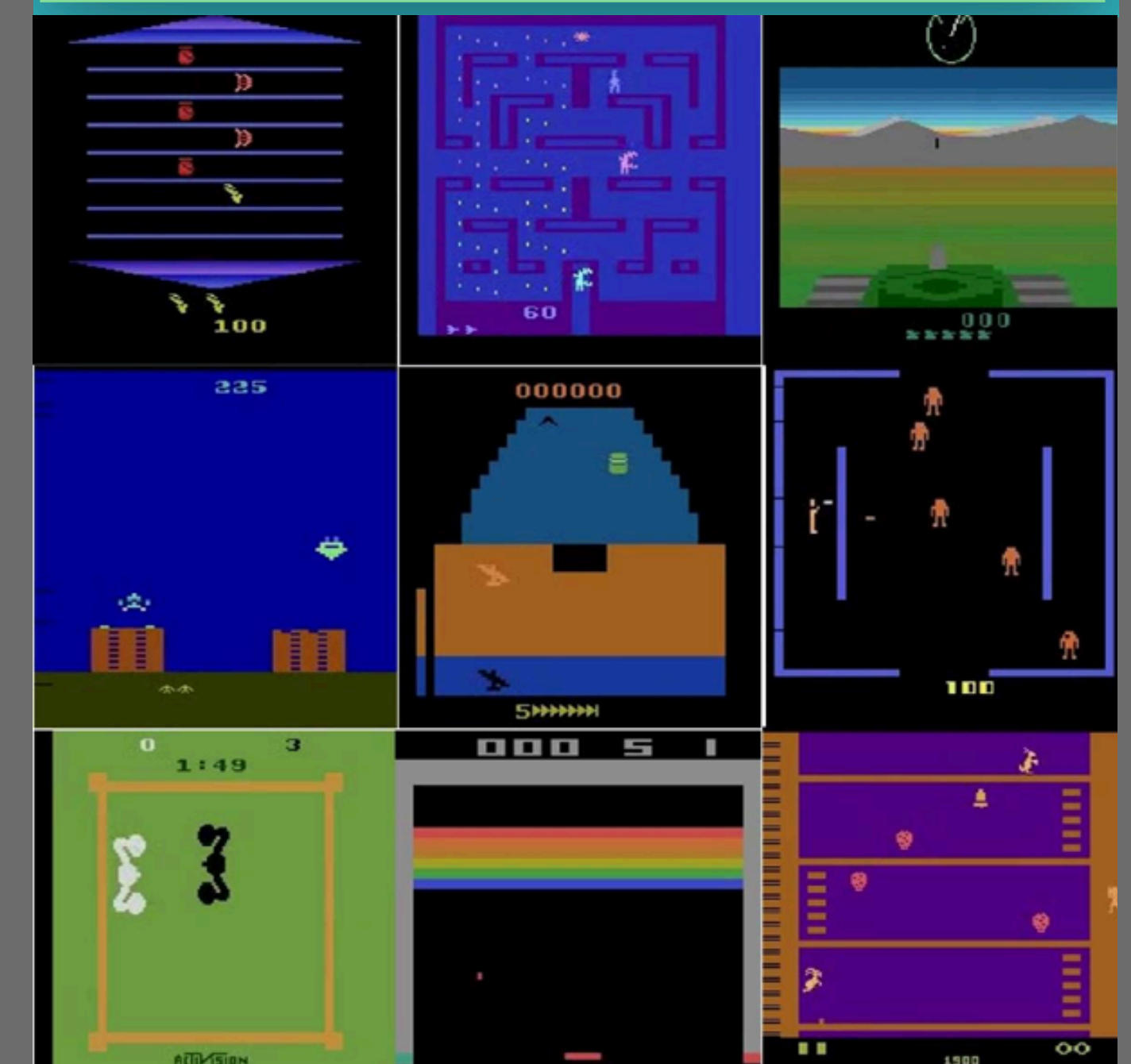
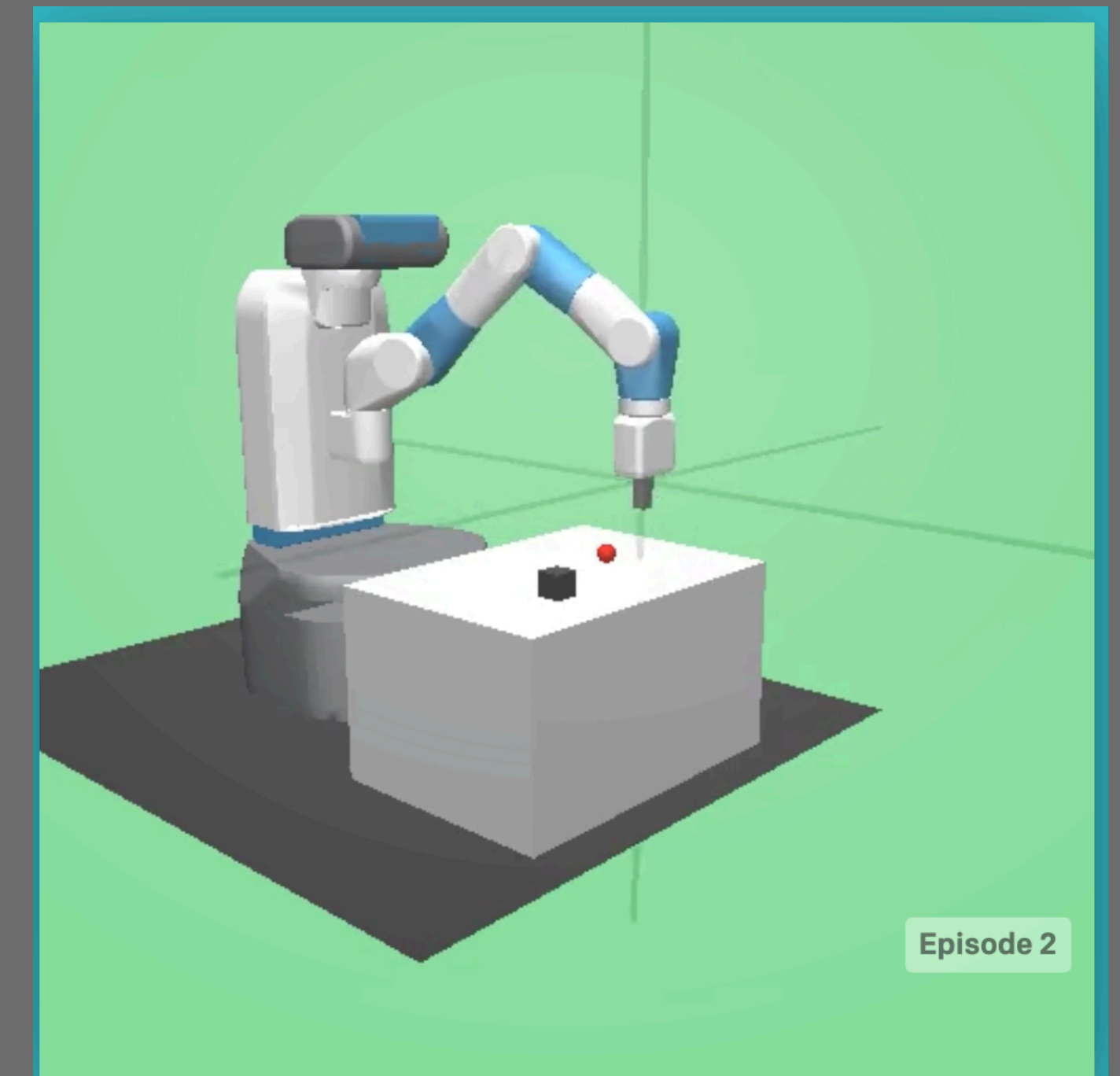
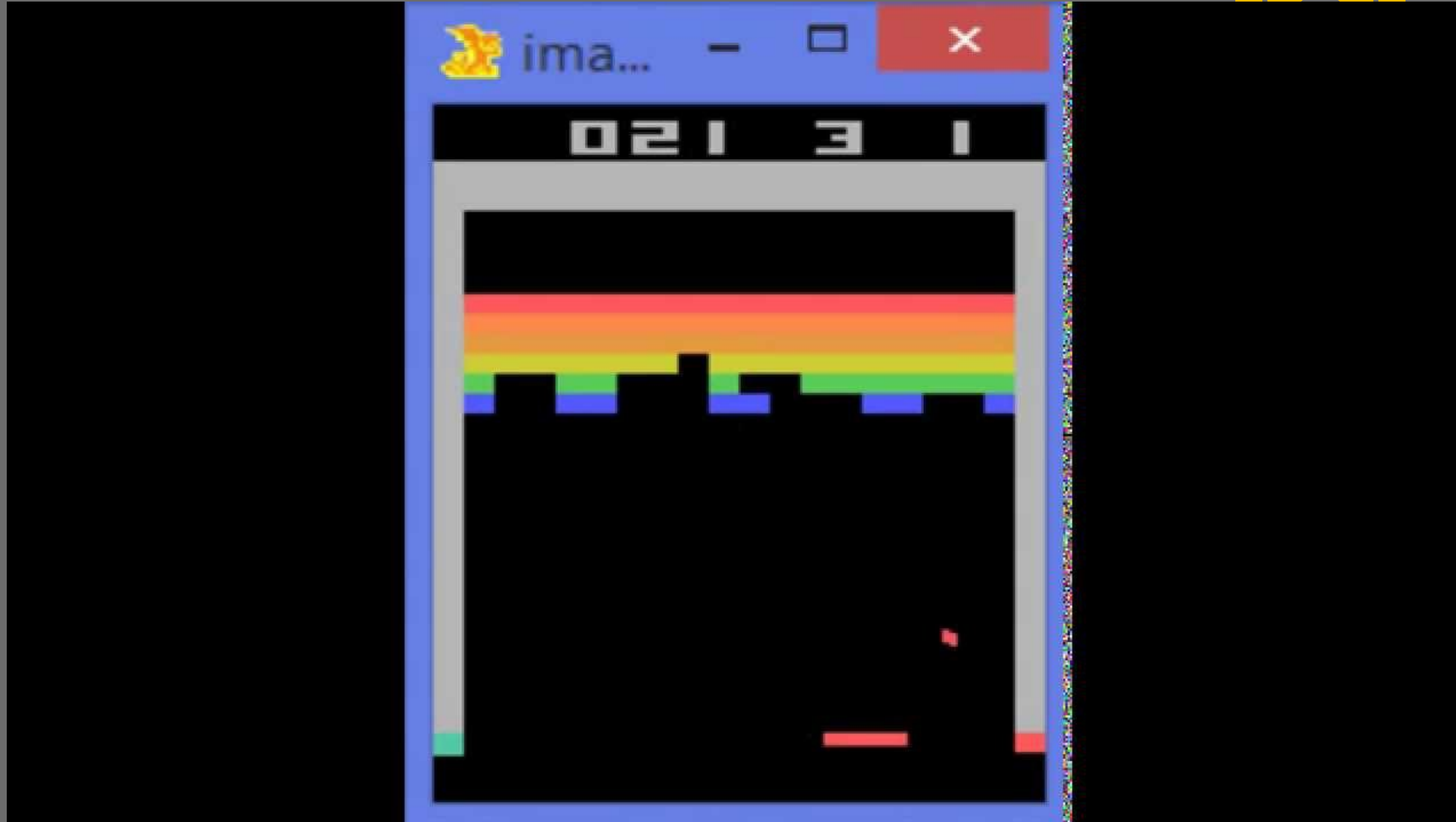


Aprendizaje por refuerzo

Objetivo: Mapear situaciones en acciones que permitan maximizar la recompensa a lo largo del tiempo. El diseño de este mapeo esta basado en la explotación y exploración.

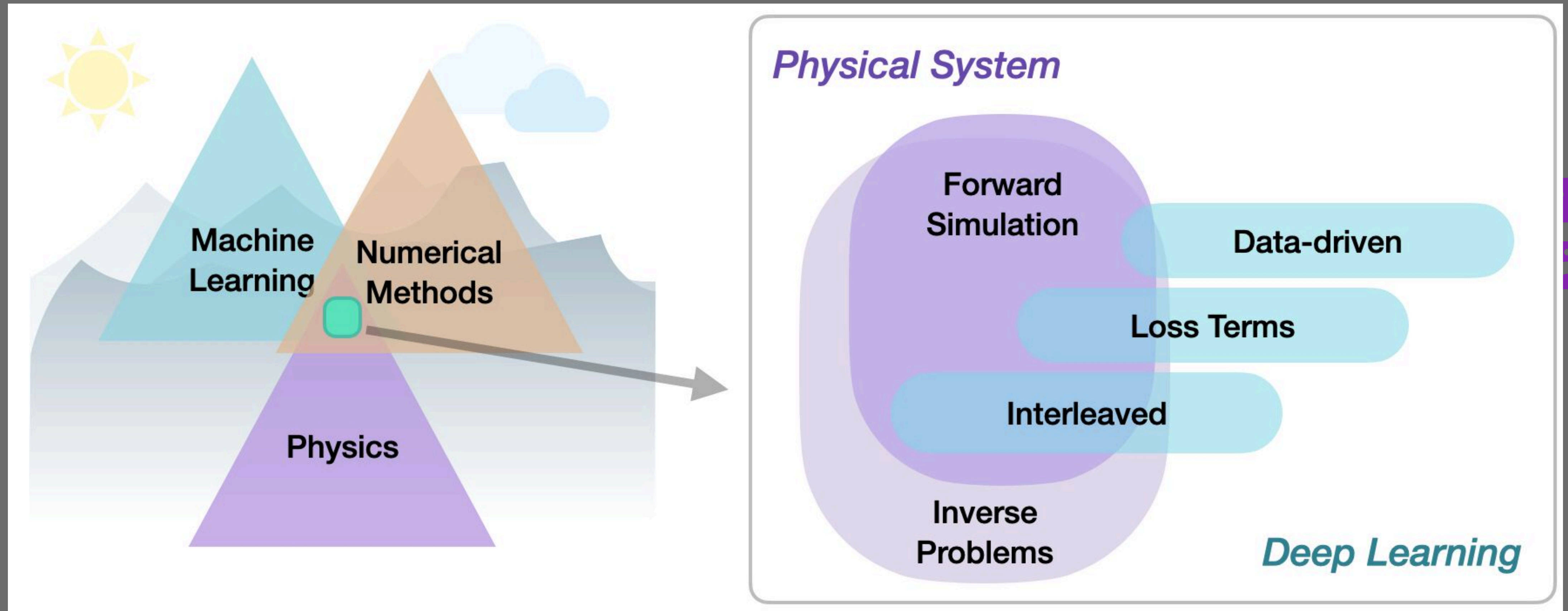


Aprendizaje por refuerzo



Physics-based learning

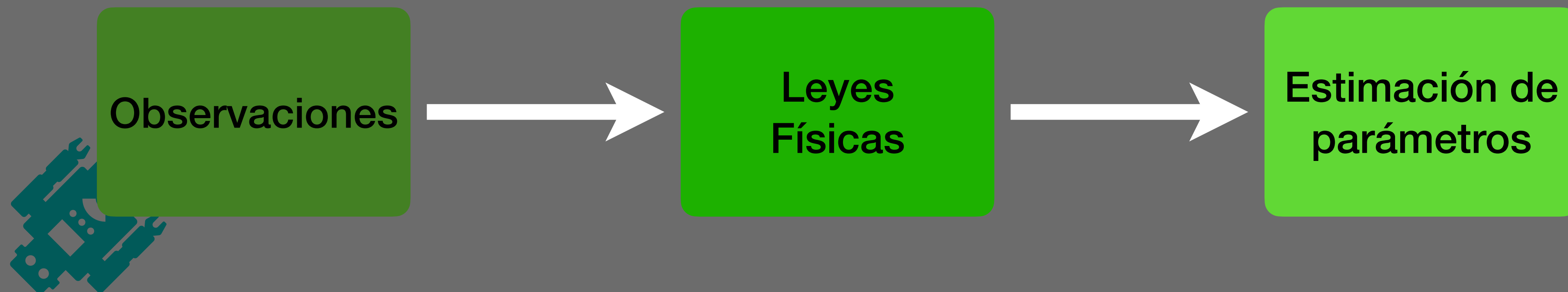
Es la integración entre métodos de ML, modelación matemática y simulación de fenómenos físicos. La idea es incorporar principios físicos en el proceso de entrenamiento.



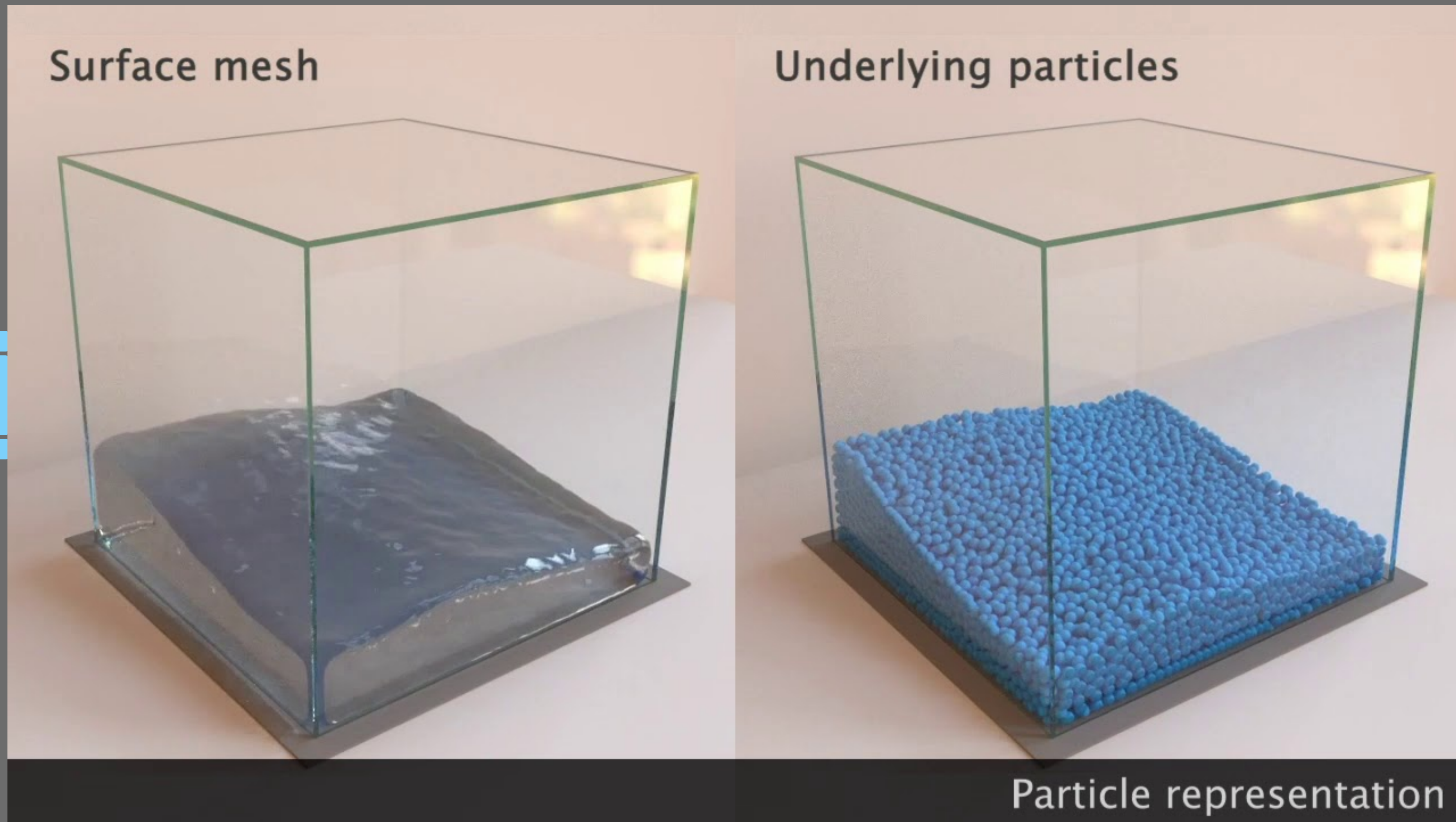
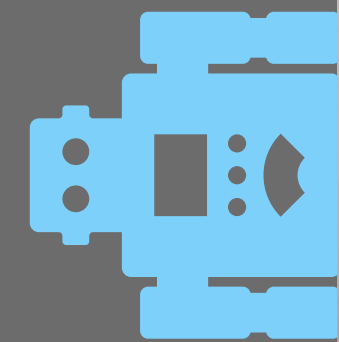
Problema directo (*forward problem*)



Problema inverso (*inverse problem*)

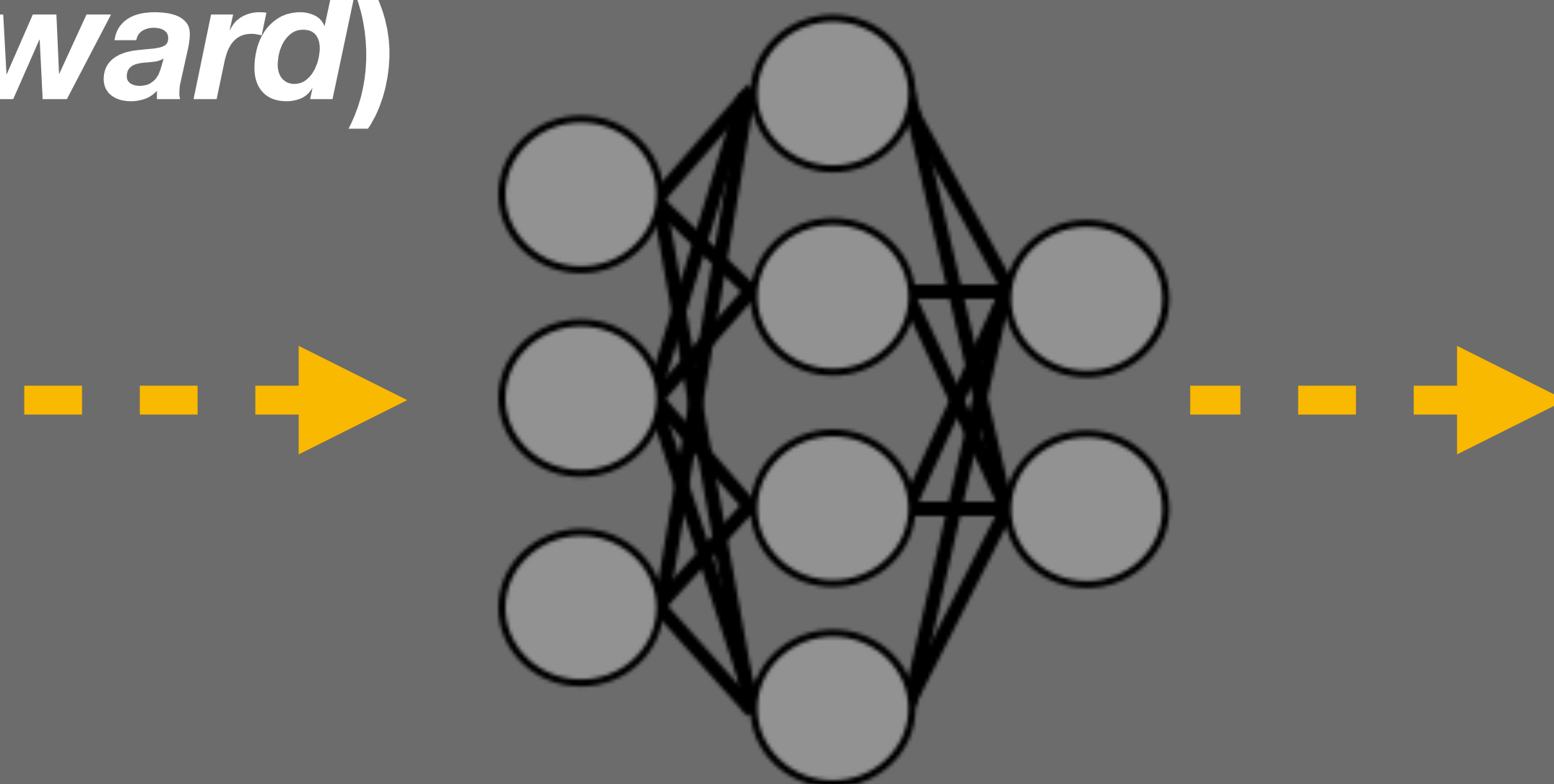


Learning to Simulate Complex Physics with Graph Networks



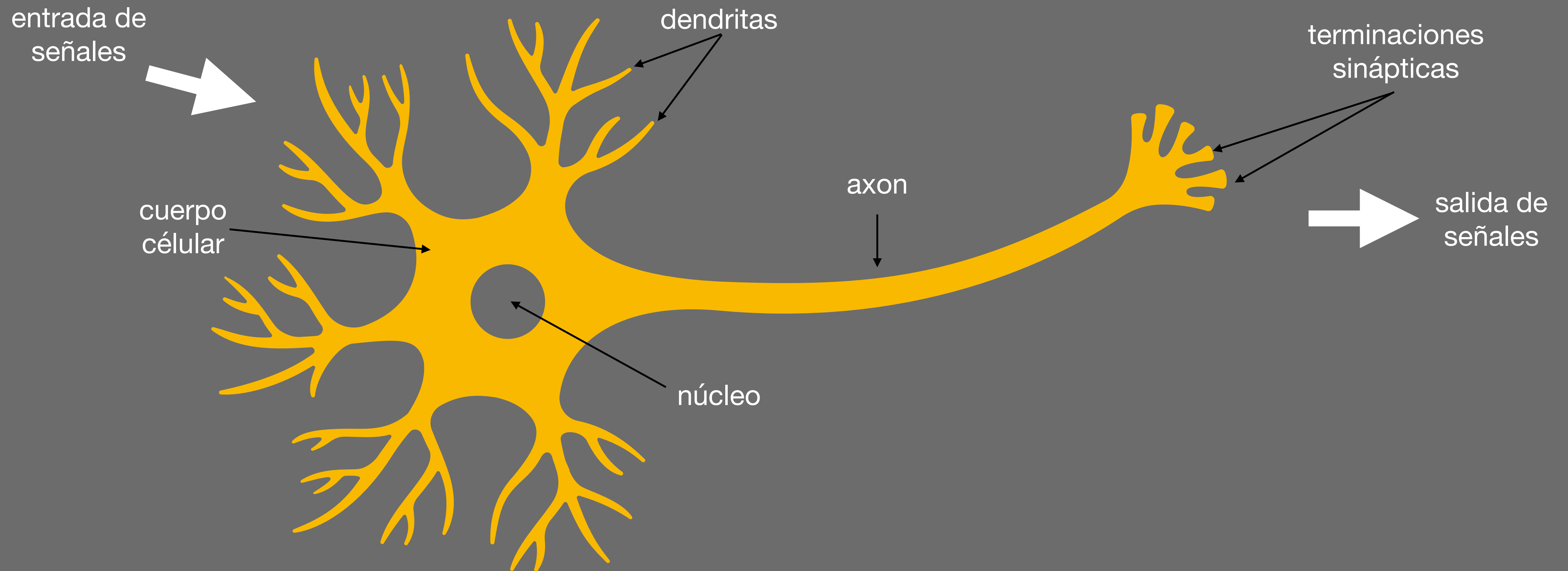


Redes neuronales artificiales unidireccionales (*feedforward*)



Antecedentes

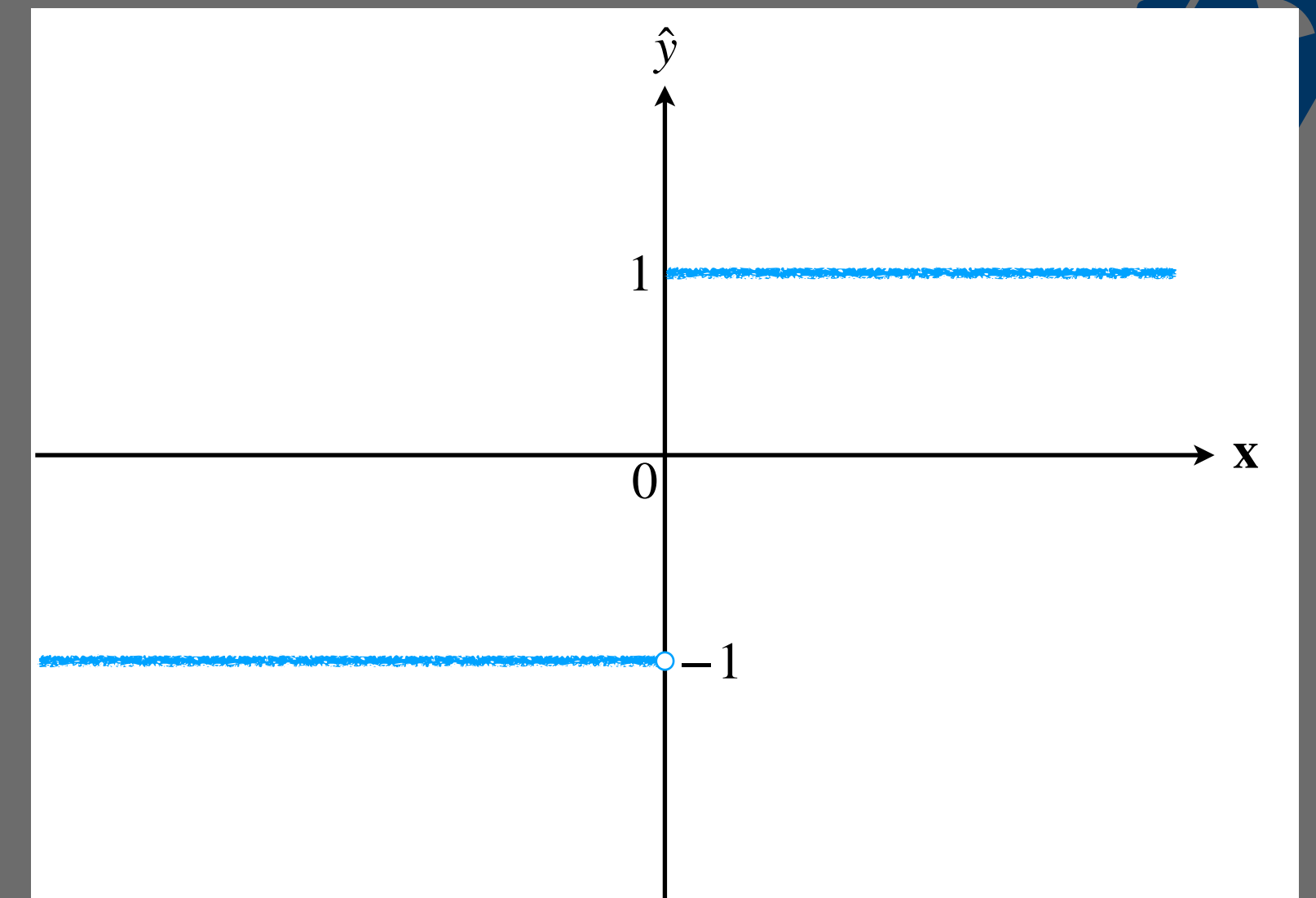
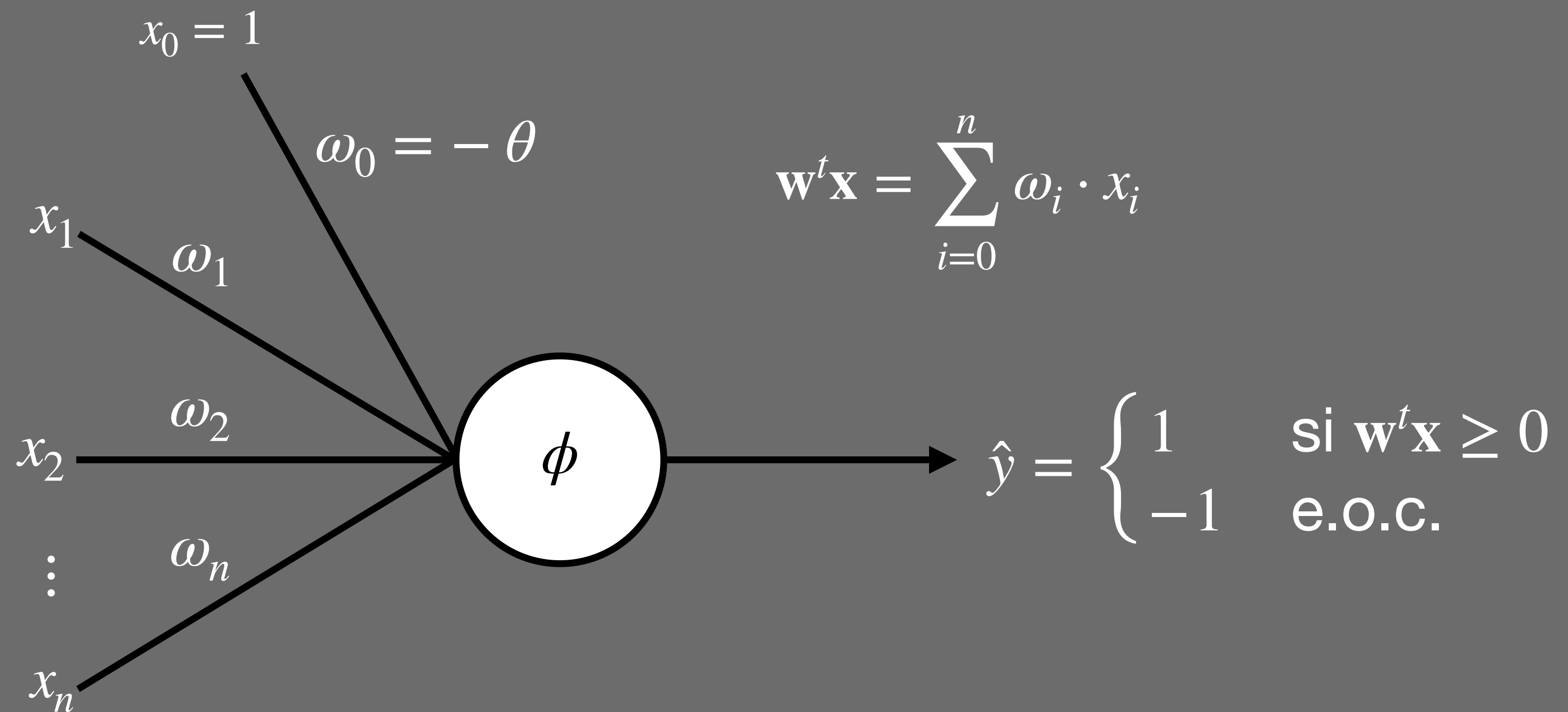
Inspiración en redes neuronales biológicas



Antecedentes

Perceptrón - F. Rosenblatt (1958)

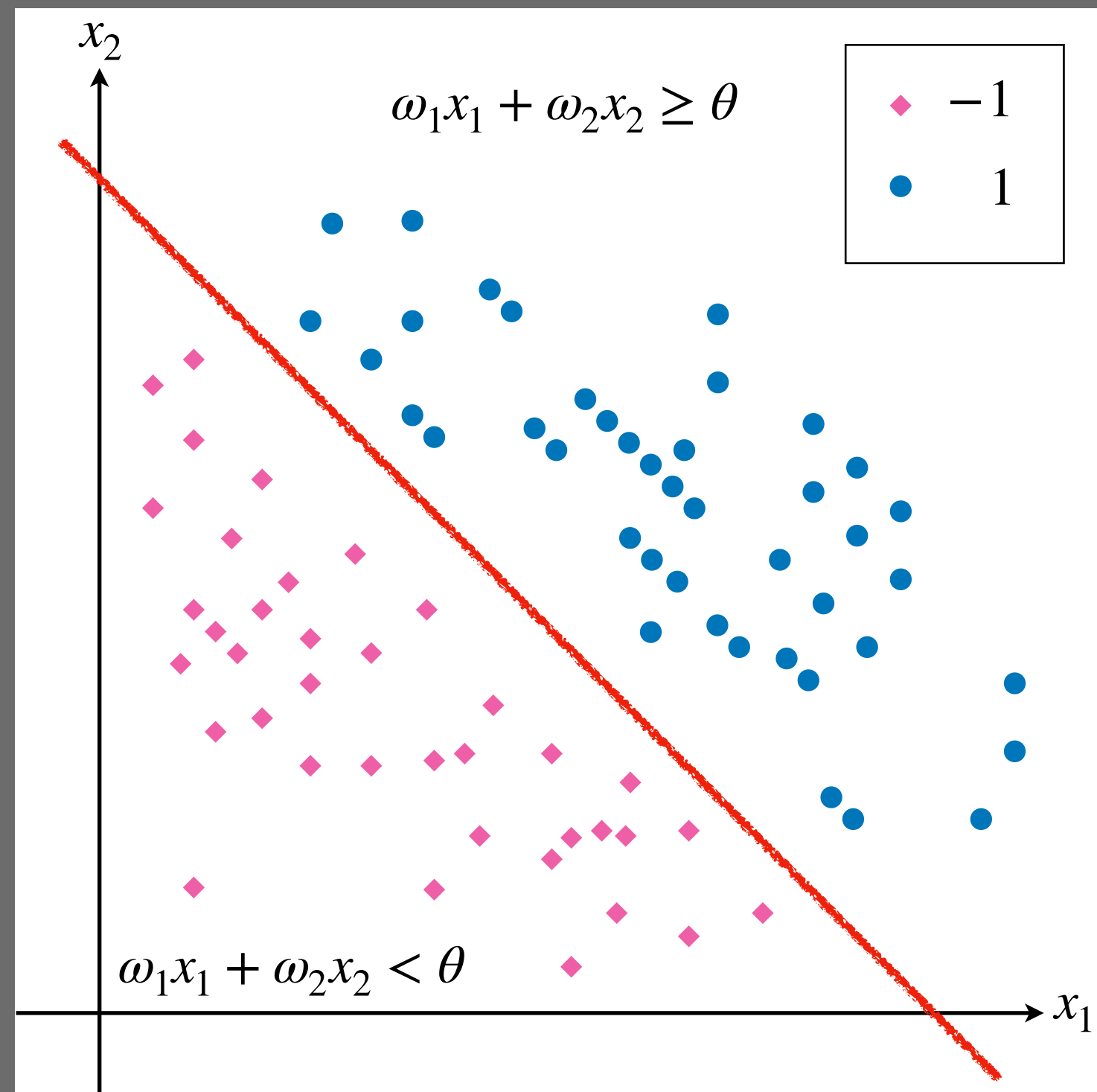
Sean $\mathbf{x} = [x_0, x_1, \dots, x_n]^t$, $\mathbf{w} = [\omega_0, \omega_1, \dots, \omega_n]^t \in \mathbf{R}^{n+1}$. A cada $\mathbf{x}^{(r)}$ le corresponde $y^{(r)} \in \{-1, 1\}$.



Perceptrón - F. Rosenblatt (1958)

Si $n = 2$, se observa una región de decisión representada por la recta

$$x_2 = -\frac{\omega_1}{\omega_2}x_1 + \frac{\theta}{\omega_2}$$

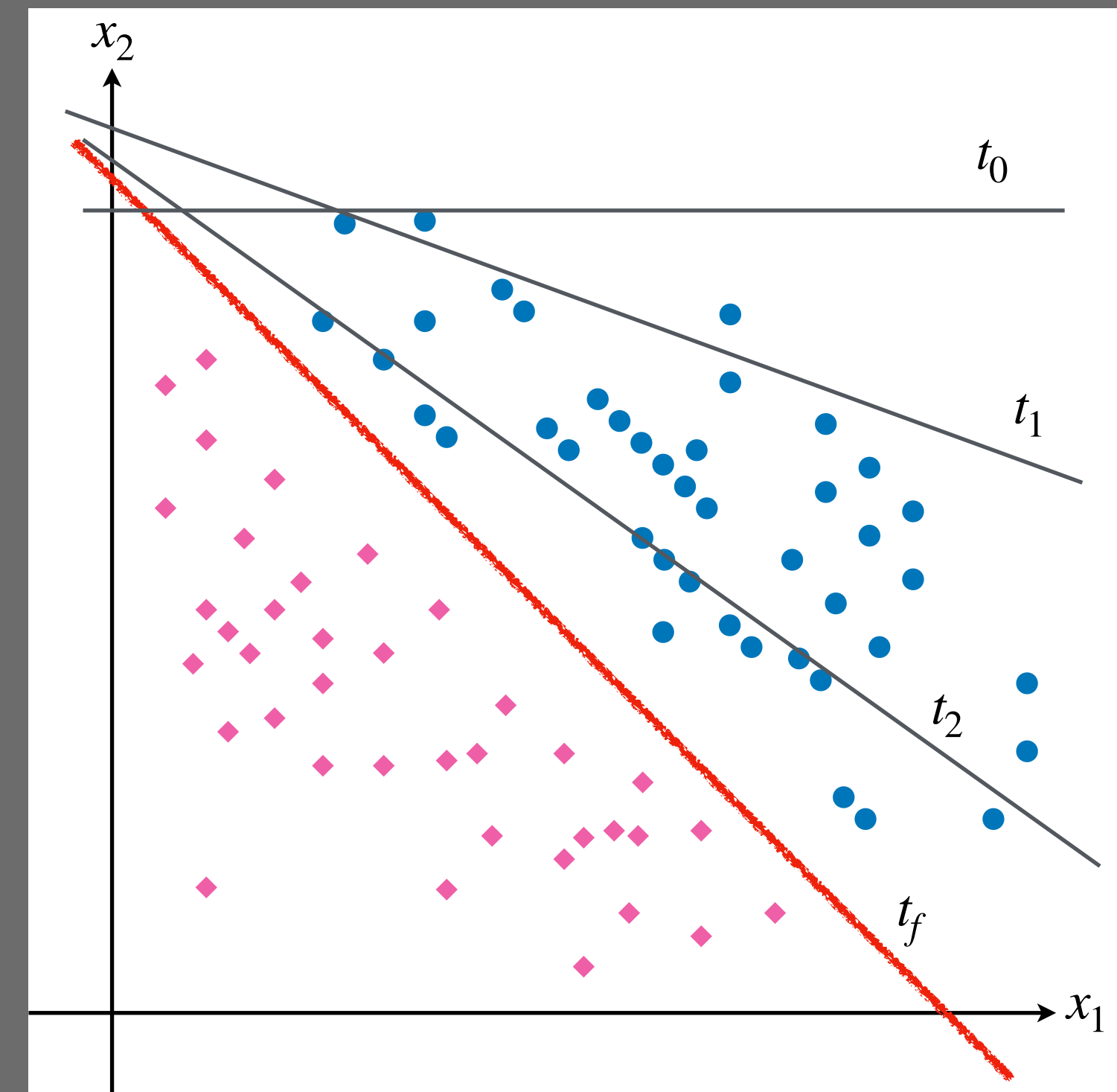


Discriminador lineal

Regla de Aprendizaje:

$$\omega_j^{(t+1)} = \omega_j^{(t)} + \Delta\omega_j$$

$$\Delta\omega_j = \eta (y^{(r)} - \hat{y}^{(r)}) x_j^{(r)} \quad \eta \in (0,1)$$



Convergencia garantizada si los datos son linealmente separables

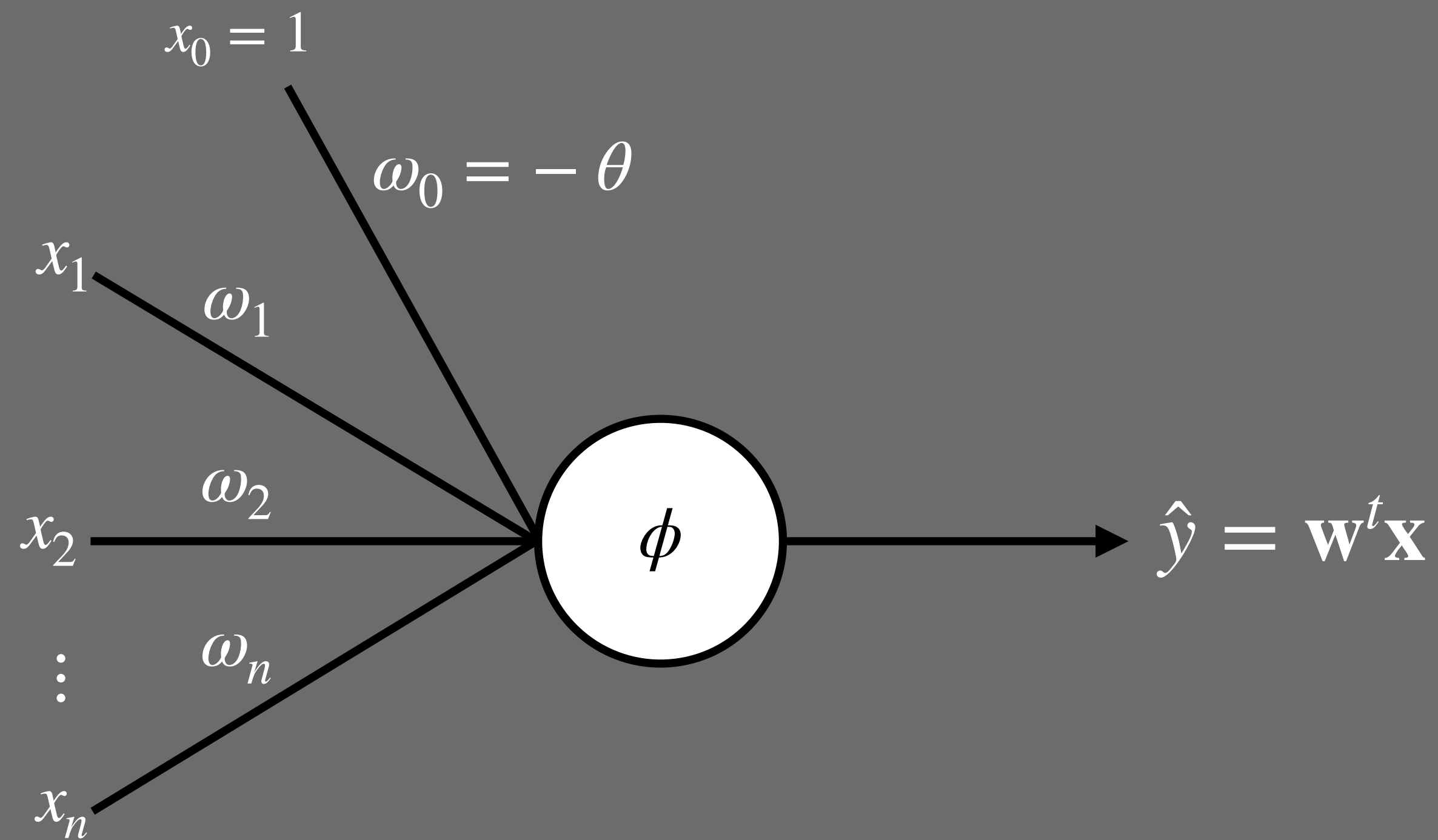
Antecedentes



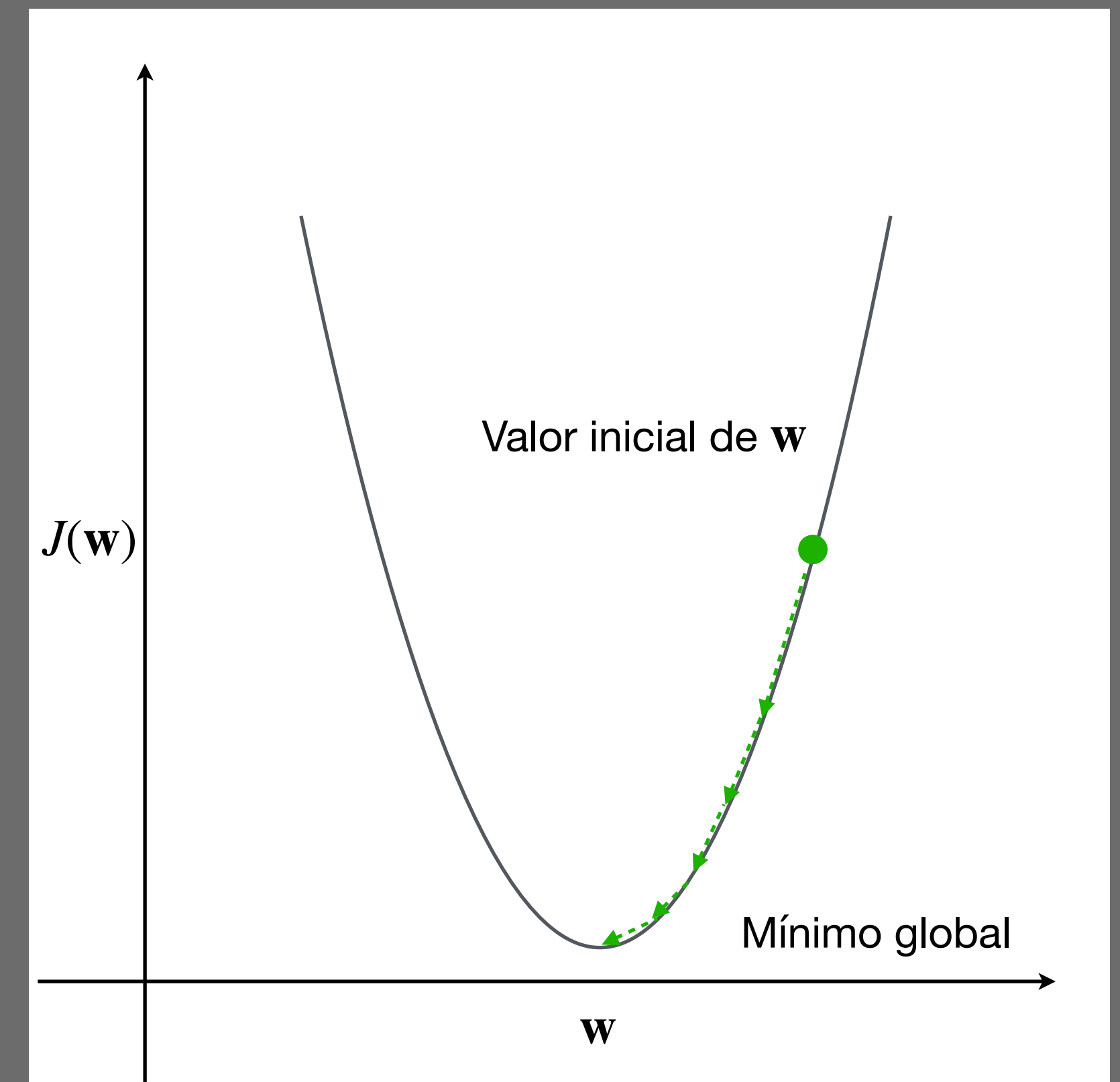
Adaline: ADaptive Linear NEuron - B. Widrow, T. Hoff (1960)

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \eta \nabla J(\mathbf{w}^{(t)})$$

$$\frac{\partial J}{\partial w_j} = \sum_{r=1}^N (y^{(r)} - \hat{y}^r) x_j^{(r)}$$



Función objetivo: $J(\mathbf{w}) = \frac{1}{2} \sum_{r=1}^N (y^{(r)} - \hat{y}^{(r)})^2$



Perceptrón multi-capa



Lectura

- A. http://www.d2l.ai/chapter_preliminaries/index.html (Capítulo 2)
- B. <https://pytorch.org/assets/deep-learning/Deep-Learning-with-PyTorch.pdf> (Capítulo 3)

