

Programación asistida por inteligencia artificial

Paradigmas de la Programación
2023

la programación asistida ha existido siempre

IDEs

autocomplete

los traductores tenían las memorias de traducción

la programación asistida ha existido siempre

IDEs

autocomplete

los traductores tenían las memorias de traducción

PERO

son deterministas

no capturan las causas latentes

Inteligencia Artificial

Inteligencia Artificial

En 2019 la Comisión Mundial de Ética del Conocimiento Científico y la Tecnología (COMEST) de la UNESCO definió la inteligencia artificial como un campo que implica máquinas capaces de imitar determinadas funcionalidades de la inteligencia humana, incluidas características como la percepción, el aprendizaje, el razonamiento, la resolución de problemas, la interacción lingüística e incluso la producción de trabajos creativos.

Aprendizaje Automático

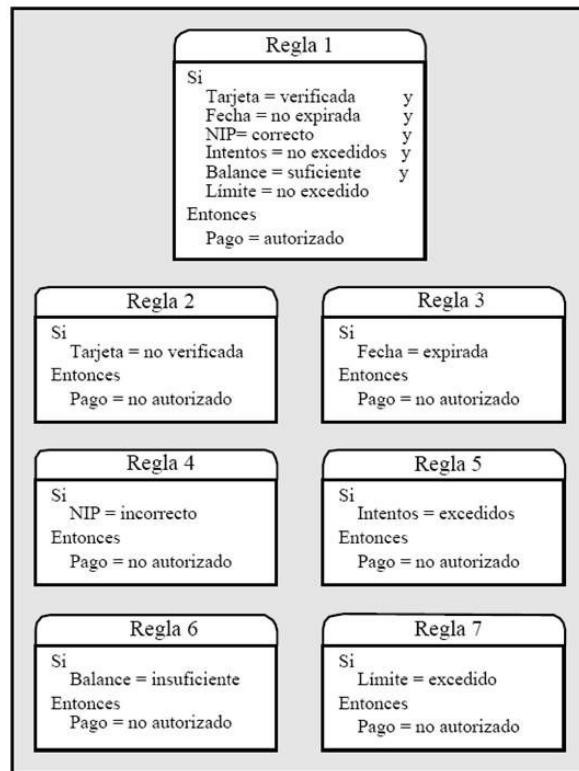
Aprendizaje Automático

En el aprendizaje de máquinas un computador observa datos, construye un modelo basado en esos datos y utiliza ese modelo a la vez como una hipótesis acerca del mundo y una pieza de software que puede resolver problemas

Russell & Norvig (2021)

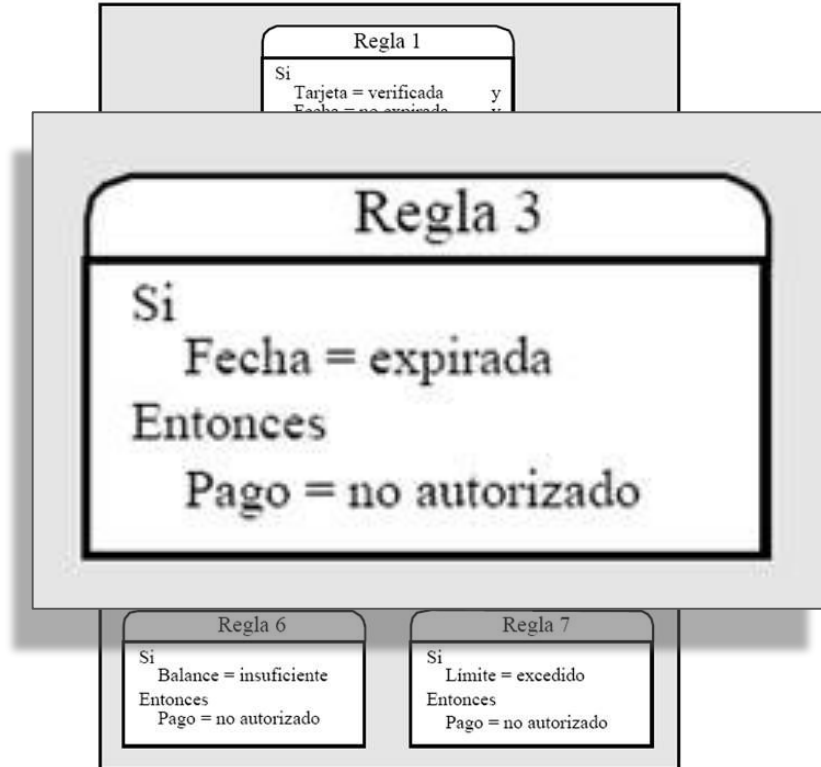
Cómo funciona un sistema de IA?

Sistemas expertos



Cómo funciona un sistema de IA?

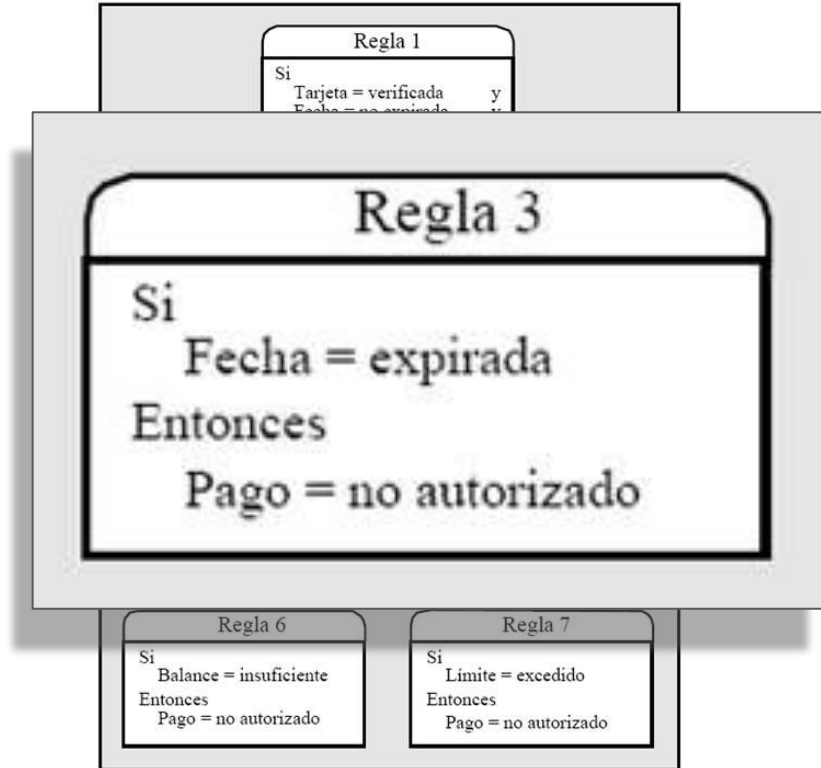
Sistemas expertos



Cómo funciona un sistema de IA?

Sistemas expertos

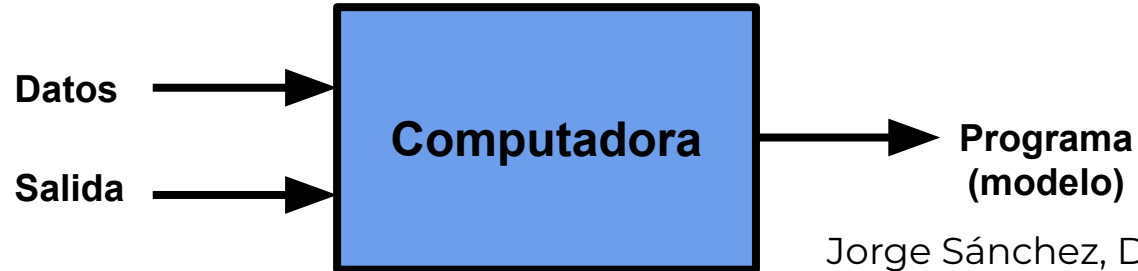
Los problemas humanos
son demasiado complejos



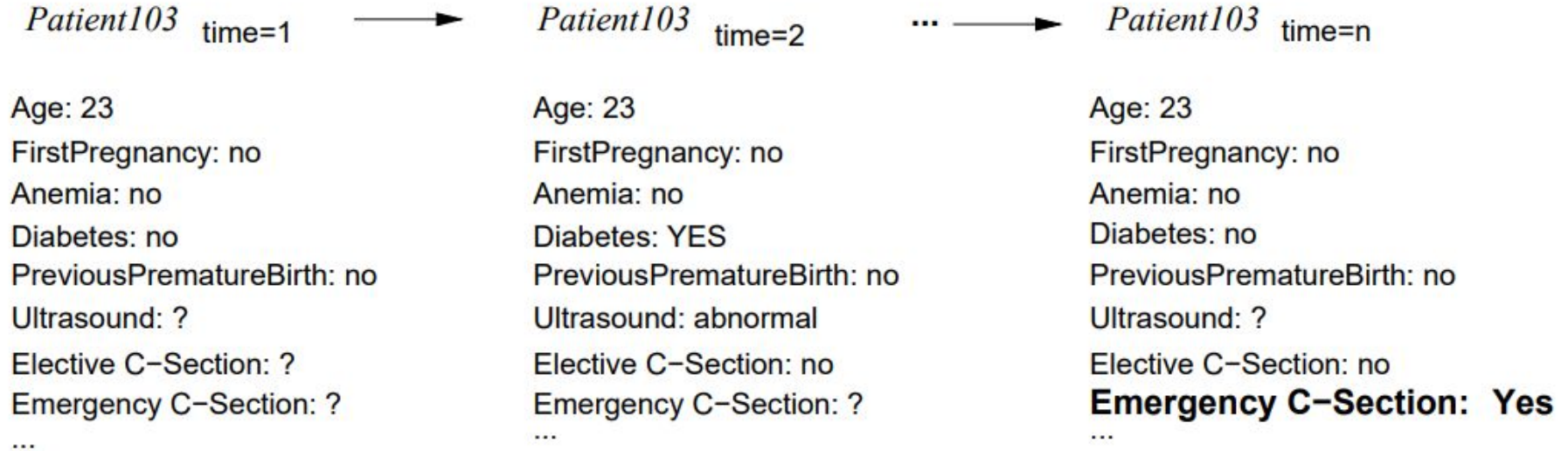
Aprendizaje Automático vs. IA basada en reglas



Aprendizaje automático



Cómo son los datos para aprendizaje automático



Cómo son los datos para aprendizaje automático

Patient103 time=1

Age: 23

FirstPregnancy: no

Anemia: no

Diabetes: no

PreviousPrematureBirth: no

Ultrasound: ?

Elective C-Section: ?

Emergency C-Section: ?

...

Given:

- 9714 patient records, each describing a pregnancy and birth
- Each patient record contains 215 features

Learn to predict:

- Classes of future patients at high risk for Emergency Cesarean Section

Cómo son los datos para aprendizaje automático

Patient103 time=1

Given:

A
F
A
D
P
U
E
E
...
If No previous vaginal delivery, and
Abnormal 2nd Trimester Ultrasound, and
Malpresentation at admission
Then Probability of Emergency C-Section is 0.6

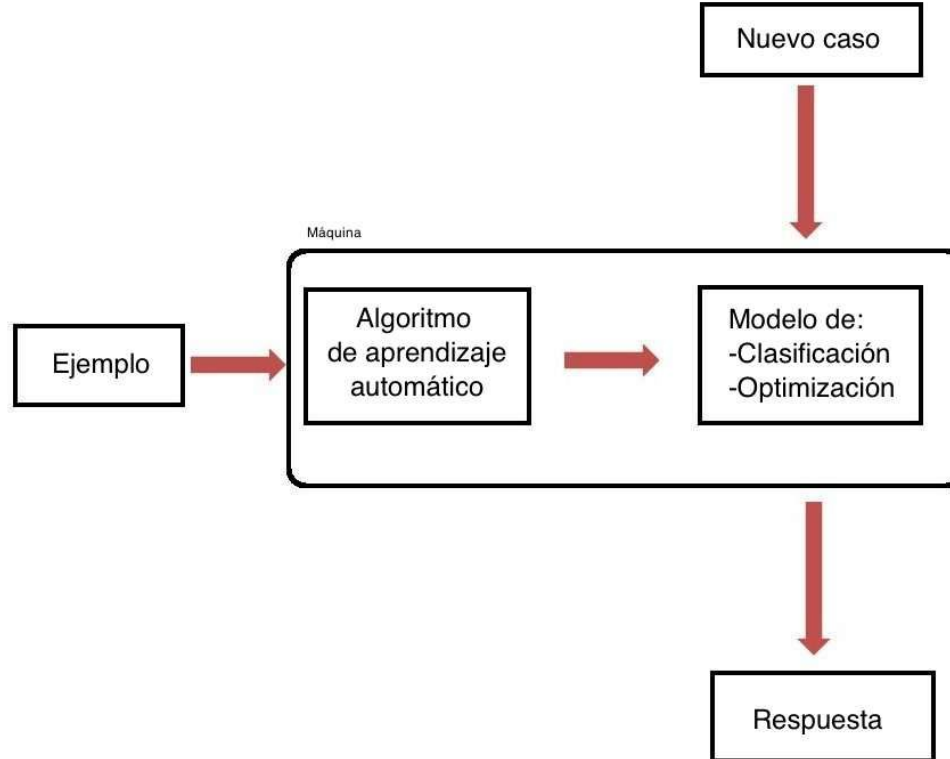
Over training data: $26/41 = .63$,

Over test data: $12/20 = .60$

Cómo se aprende?

Mejorar en una tarea con respecto a una métrica de desempeño a partir de experiencia

aprendizaje automático



aprendizaje automático

Funciones \mathcal{F}

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$

Datos de entrenamiento

$$\{(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}\}$$

Aprendizaje

$$\begin{array}{l} \text{find } \hat{f} \in \mathcal{F} \\ \text{s.t. } y_i \approx \hat{f}(x_i) \end{array}$$



Predicción

$$y = \hat{f}(x)$$

$$x$$

Nuevo dato

¿Cómo se aprende?

Mejorar en una tarea con respecto a una métrica de desempeño a partir de experiencia

sesgos en la definición de

- tarea

- métrica

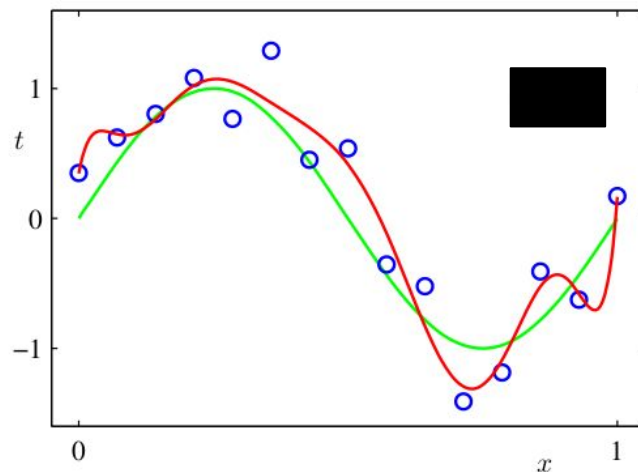
- experiencia

Tipos de aprendizaje automático

- Aprendizaje **supervisado** (inductivo)
Datos de entrenamiento + salida esperada
- Aprendizaje **no supervisado**
Datos de entrenamiento (sin salida esperada)
- Aprendizaje **semi-supervisado**
Datos de entrenamiento + pocas salida esperadas
- Aprendizaje **auto-supervisado**
Datos de entrenamiento auto generados (tareas pretexto)
- Aprendizaje **por refuerzo**
"Recompensas" a determinada secuencias de acciones

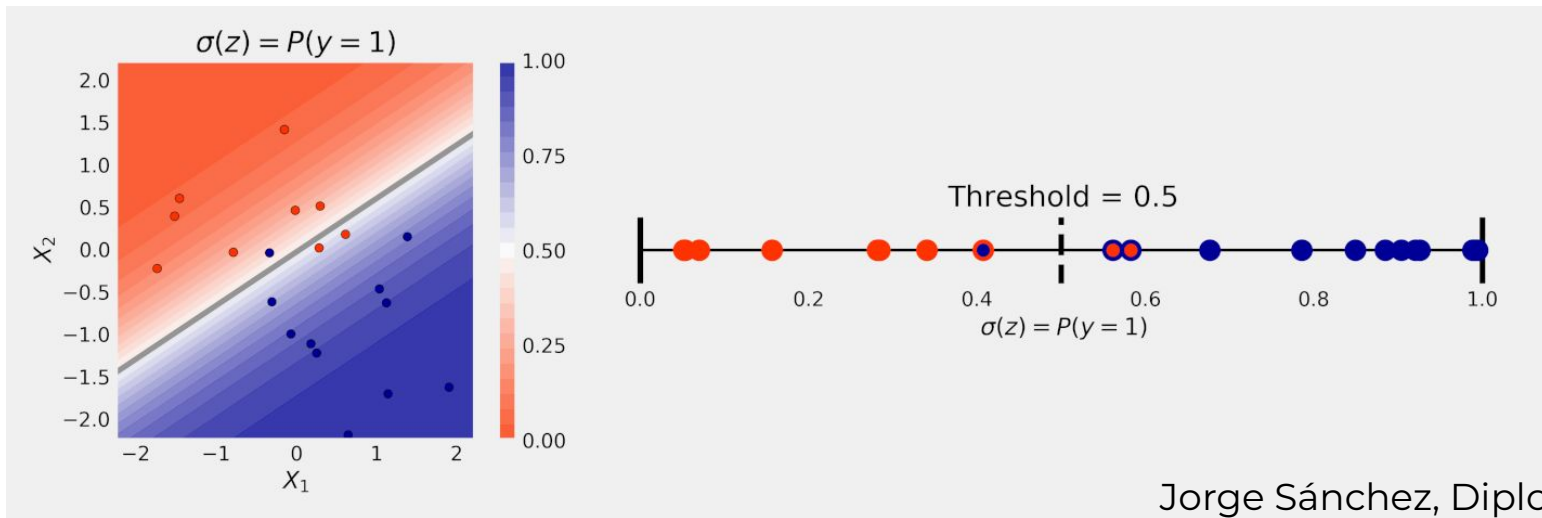
Aprendizaje supervisado: Regresión

- Datos $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$
- Aprender una $f(x)$ que permita predecir y a partir de x
 - Si y está en $\mathbb{R}^n \rightarrow$ **regresión**



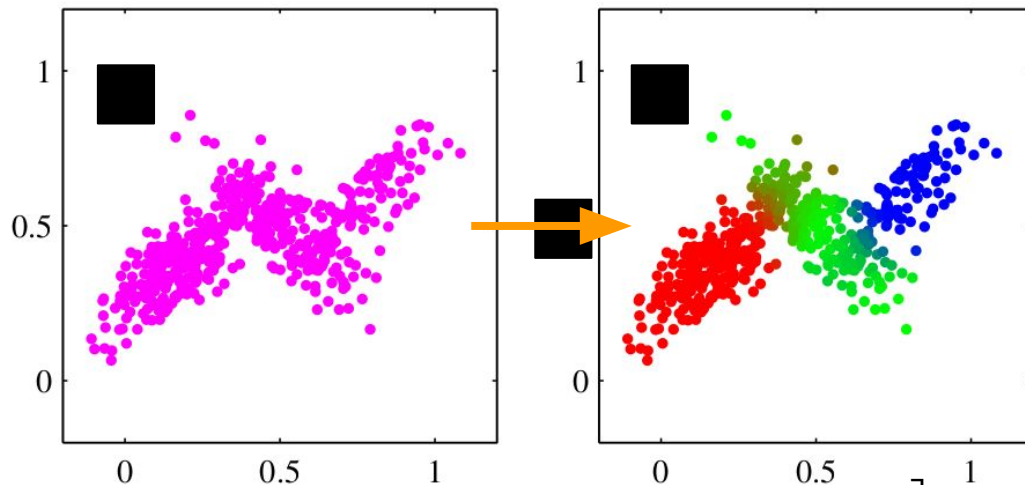
Aprendizaje supervisado: Clasificación

- Datos $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$
- Aprender una $f(x)$ que permita predecir y a partir de x
 - Si y es categórica \rightarrow **clasificación**



Aprendizaje no supervisado: Clustering

- Datos x_1, x_2, \dots, x_n
- Aprender la estructura interna de los datos
 - p.ej. *clustering*



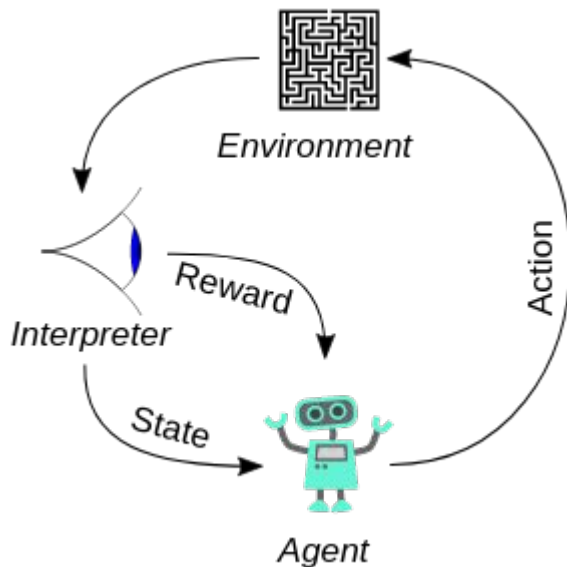
Aprendizaje auto-supervisado

- Datos x_1, x_2, \dots, x_n
- Utilizar estructura interna y generar **tareas pretexto** (supervisión)
 - p.ej.: predecir siguiente elemento en una secuencia
- (pre)entrenar para aprender a representar bien los datos
- Adaptar a la tarea de interés (regresión, clasificación, ...)



Aprendizaje por refuerzo

- Dada una secuencia de estados y acciones con recompensa (*reward*), generar una secuencia de acciones (*policy*) que nos indique qué hacer ante un determinado estado observado



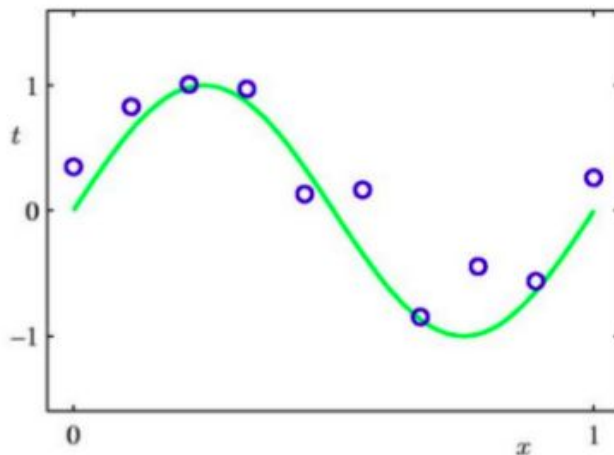
regresión

Regresión

- Disponemos de N pares de entrenamiento (observaciones)

$$\{(x_i, y_i)\}_{i=1}^N = \{(x_1, y_1), \dots, (x_N, y_N)\}$$

- El problema de regresión consiste en estimar $f(x)$ a partir de estos datos

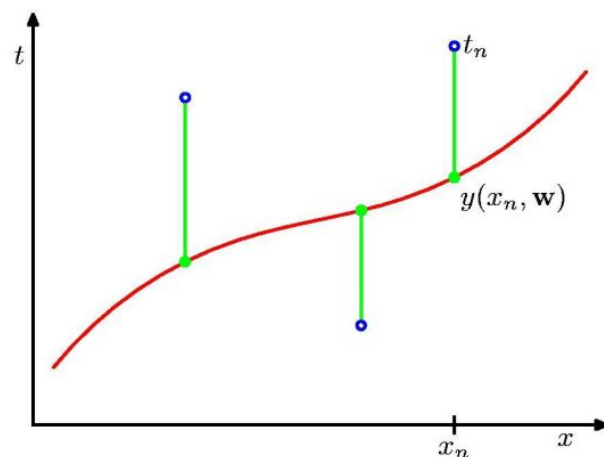
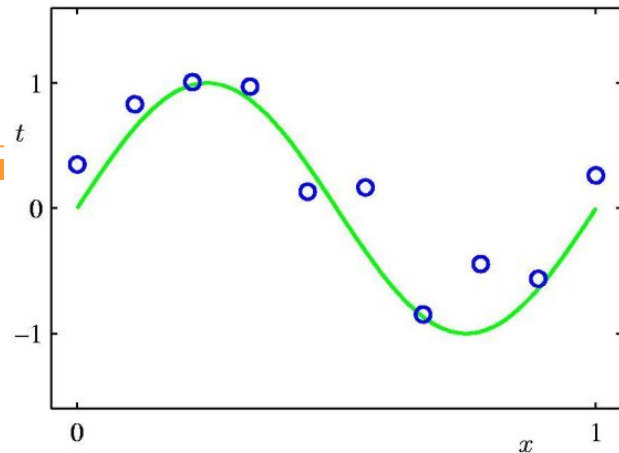


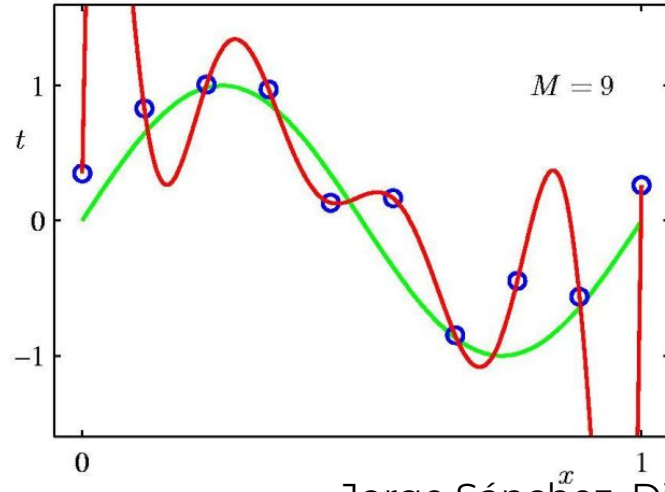
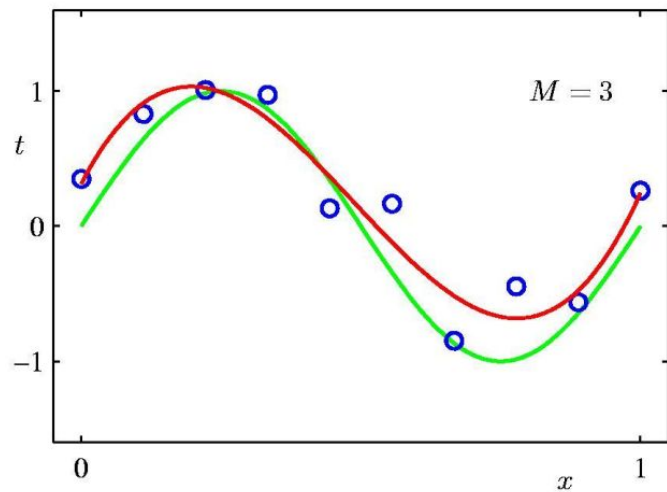
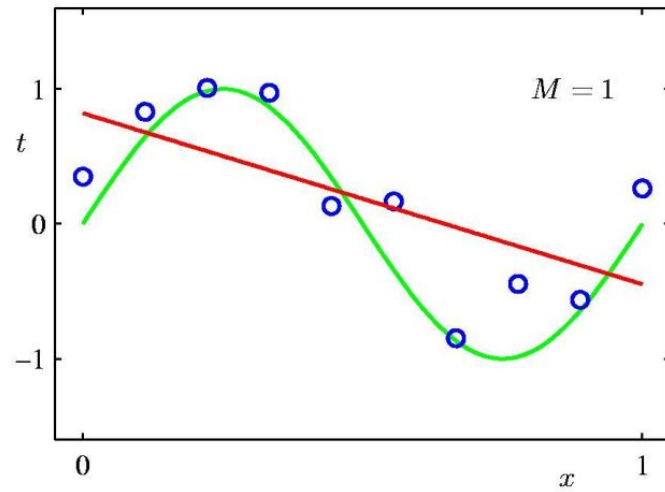
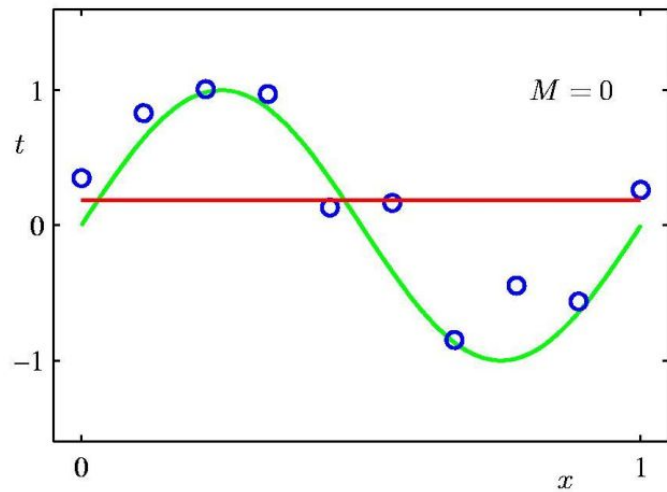
Regresión

- En verde se ilustra la función "verdadera" (inaccesible)
- Las muestras son uniformes en x y poseen ruido en y
- Utilizaremos una **función de costo** (error cuadrático) para medir el error en la predicción de y mediante $f(x)$

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2$$

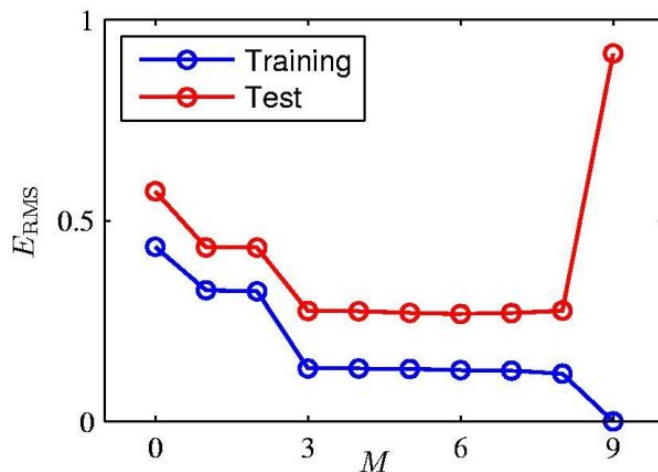
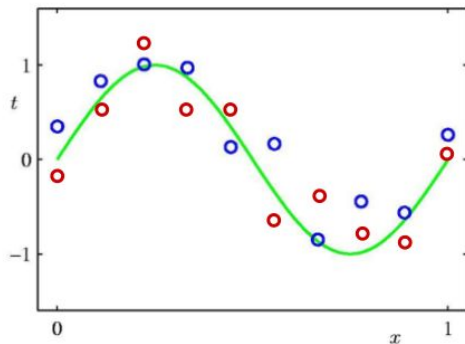
$$y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j$$





Sobreajuste (overfitting)

- Datos de test: otra muestra de los misma función subyacente
- El error de entrenamiento se hace cero, pero el de test crece con M



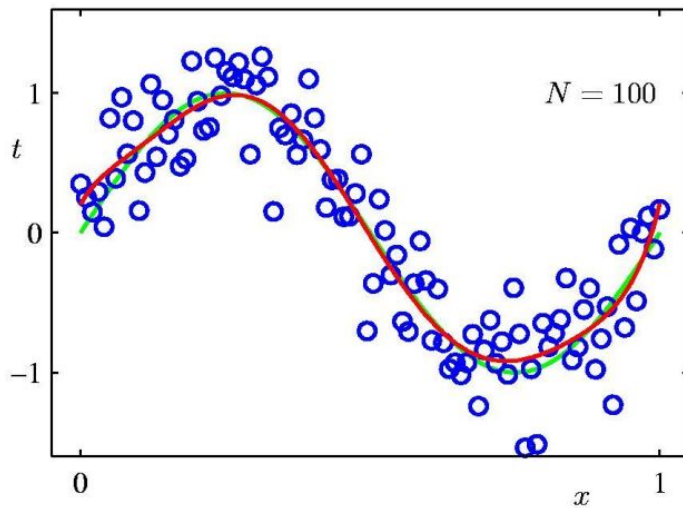
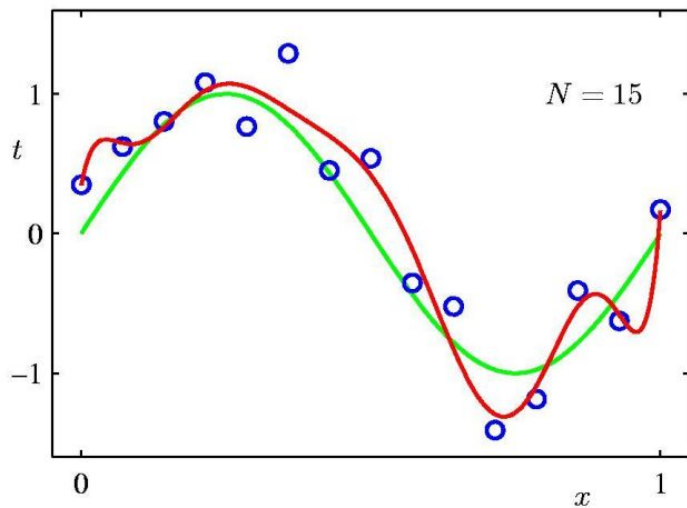
Root-Mean-Square (RMS) Error: $E_{\text{RMS}} = \sqrt{2E(\mathbf{w}^*)/N}$

Bondad de ajuste vs. complejidad de modelo

- Si el modelo tiene tantos grados de libertad como los presentes en los datos de entrenamiento, puede ajustarlos perfectamente
- El objetivo en aprendizaje automático no es el ajuste perfecto, sino la generalización a conjuntos nuevos (no vistos en entrenamiento)
- Podemos decir que un modelo generaliza, si puede explicar los datos empleando una complejidad acotada

Prevenir el sobreajuste

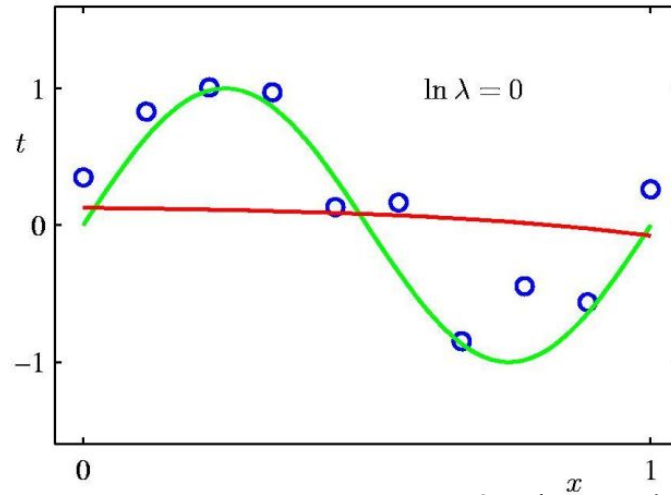
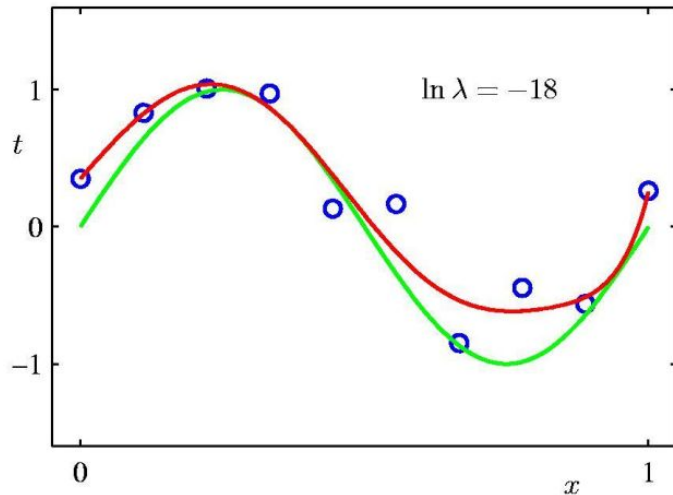
Agregar más datos (más que la "complejidad" del modelo)



Prevenir el sobreajuste

Regularización: penalizar valores grandes de los coeficientes

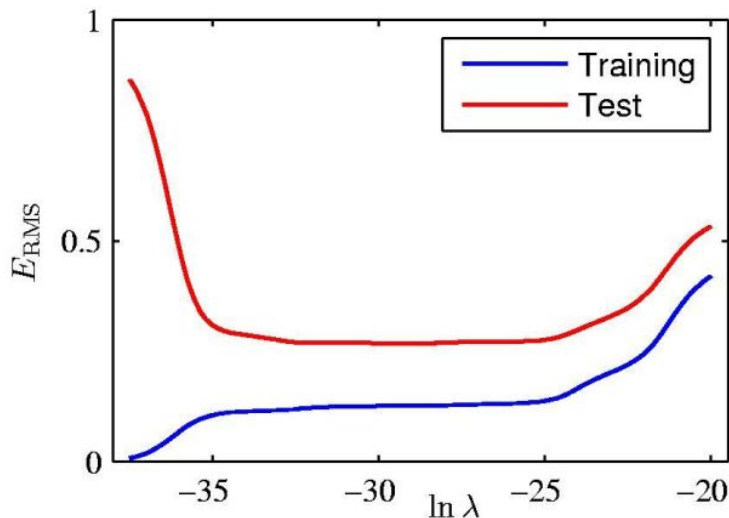
$$\tilde{E}(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2$$



Prevenir el sobreajuste

Regularización: penalizar valores grandes de los coeficientes

$$\tilde{E}(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2$$



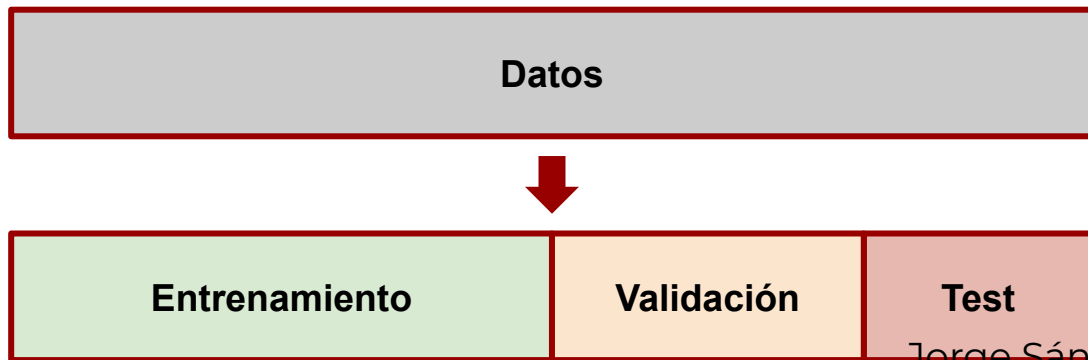
**Término de
regularización
(ridge)**

λ = hiperparámetro

Elección de hiperparámetros

Dividir el conjunto total de ejemplos en tres subconjuntos

- **Entrenamiento:** aprendizaje de variables del modelo
- **Validación:** ajuste/elección de hiperparámetros
- **Test:** estimación final de la performance del modelo entrenado (y con hiperparámetros elegidos adecuadamente)



clasificación

Clasificación binaria

- Disponemos de N pares de entrenamiento (observaciones)

$$\{(x_i, y_i)\}_{i=1}^N = \{(x_1, y_1), \dots, (x_N, y_N)\}$$

con $x_i \in \mathbb{R}^n$, $y_i \in \{-1, +1\}$.

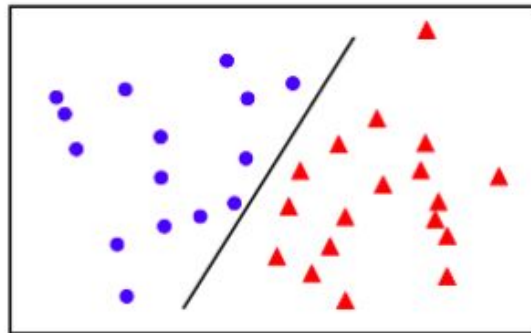
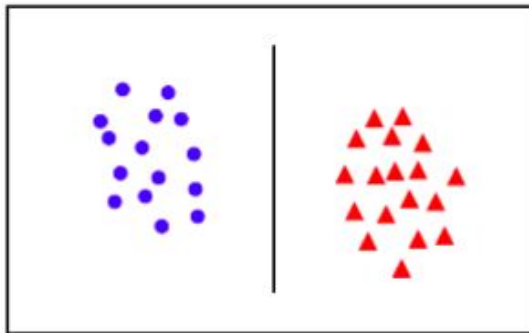
- Aprender una $f(x)$ tal que

$$f(\mathbf{x}_i) \begin{cases} \geq 0 & y_i = +1 \\ < 0 & y_i = -1 \end{cases}$$

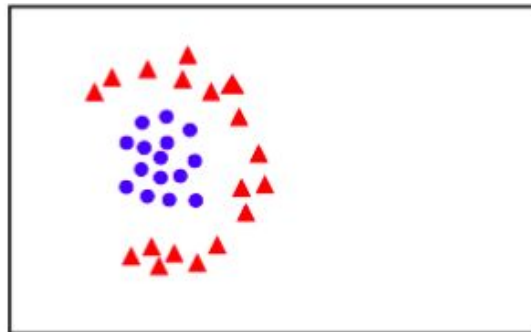
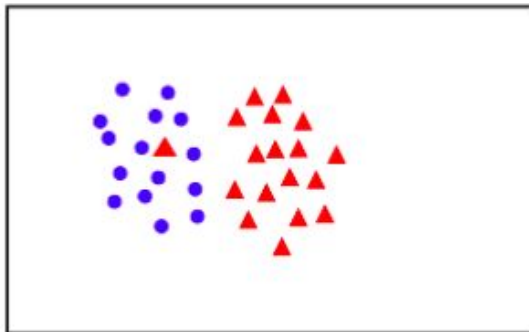
es decir: $y_i f(x_i) > 0$ para una clasificación correcta.

Separabilidad lineal

linealmente
separable



no
linealmente
separable



Clasificadores lineales

- La entrada es un vector \mathbf{x}_i de dimensionalidad n
- La salida es una etiqueta $y_i \in \{-1, +1\}$
- Clasificador = función de predicción + función de decisión

$$g(f(\mathbf{x})) \rightarrow \{-1, +1\}$$

- Función de predicción **lineal**

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0$$

- Función de decisión

$$g(z) = \text{sign}(z)$$

$$g(f(\mathbf{x})) = \text{sign}(\mathbf{w}^T \mathbf{x} + w_0)$$

$$\mathbf{a}^T \mathbf{b} \equiv \langle \mathbf{a}, \mathbf{b} \rangle = \sum_{i=1}^D a_i b_i$$

cómo se representa un ejemplo

	bolsa	cancha	comer	comprarme	delicioso	ensuciar	fútbol	ganar	jugar	juro	noche	plato	plástico	suadero	taco	vida
Document 1	0	0	1	0	1	0	0	0	0	1	0	0	0	3	1	1
Document 2	1	0	0	0	1	1	0	0	0	0	0	1	1	1	1	0
Document 3	0	1	0	1	0	0	1	1	1	0	1	0	0	0	1	0

visualización del proceso

<http://www.r2d3.us/visual-intro-to-machine-learning-part-1/>

<http://www.r2d3.us/visual-intro-to-machine-learning-part-2/>

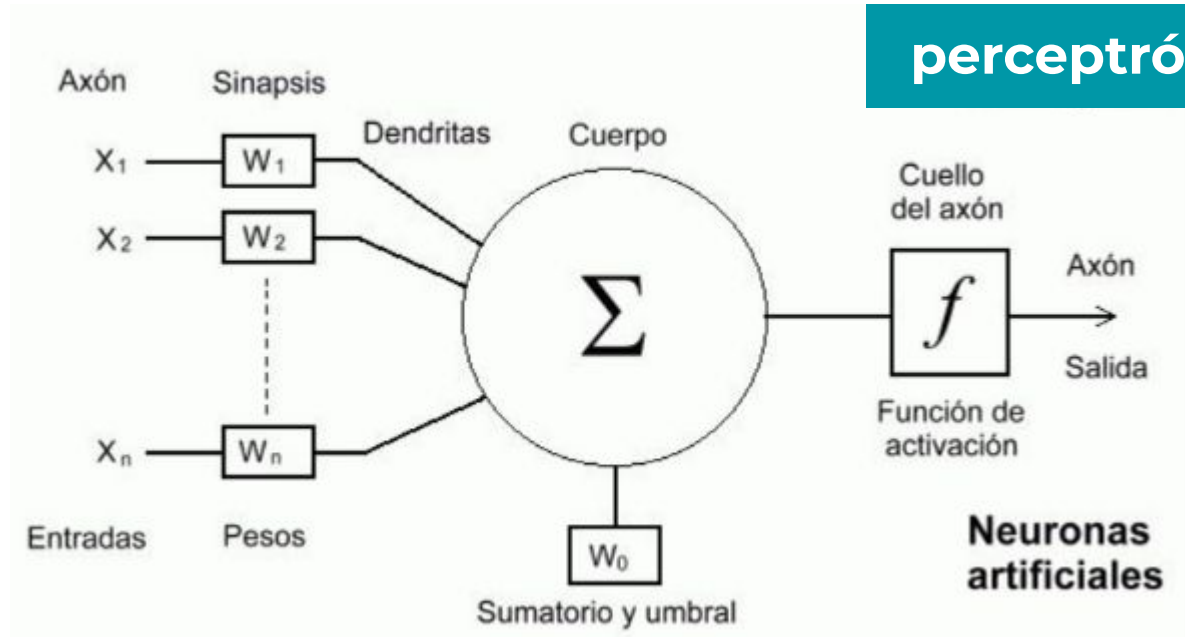
*redes
neuronales
artificiales*

múltiples clasificadores muy simples,
conectados entre ellos

redes neuronales artificiales

múltiples clasificadores muy simples,
conectados entre ellos

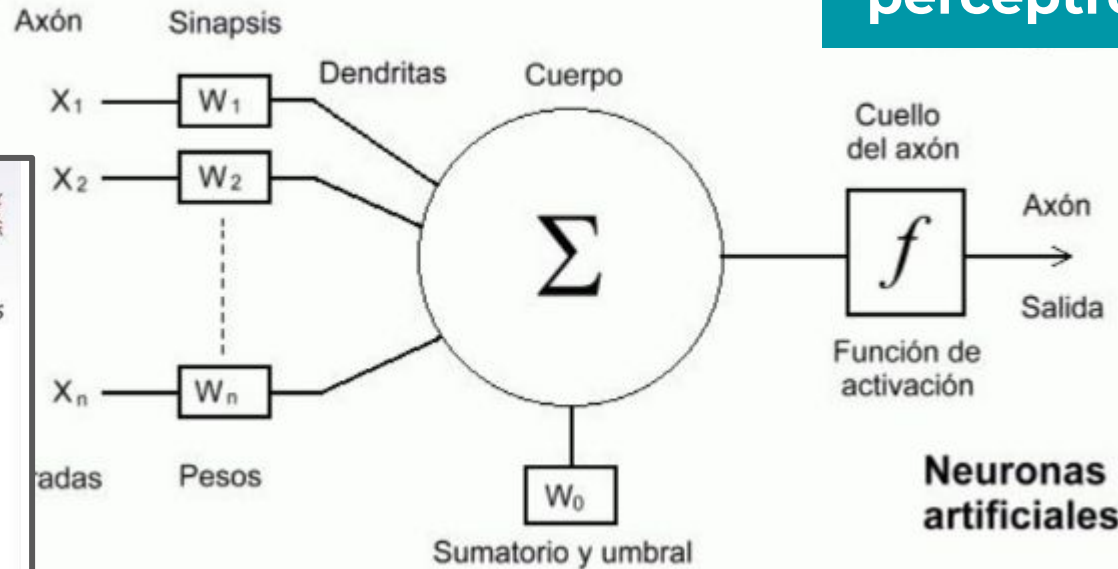
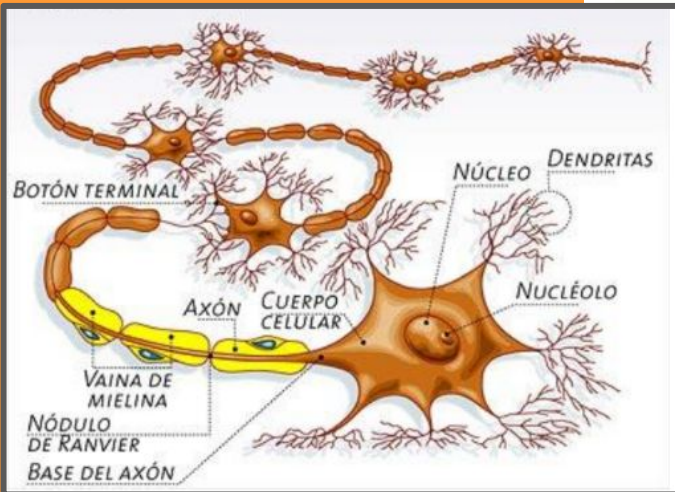
perceptrón



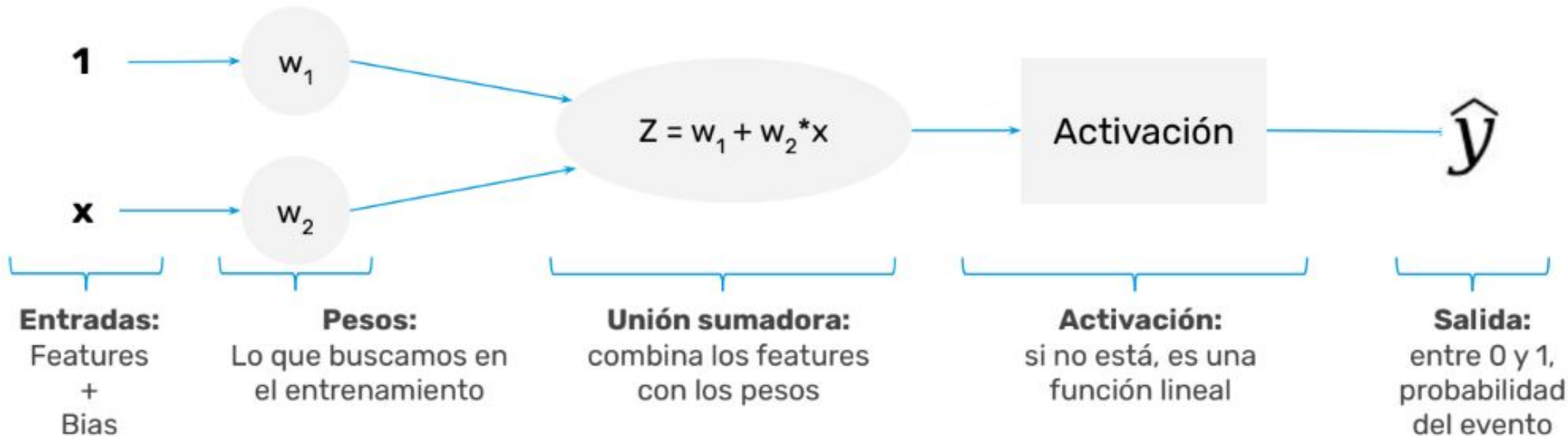
redes neuronales artificiales

múltiples clasificadores muy simples,
conectados entre ellos

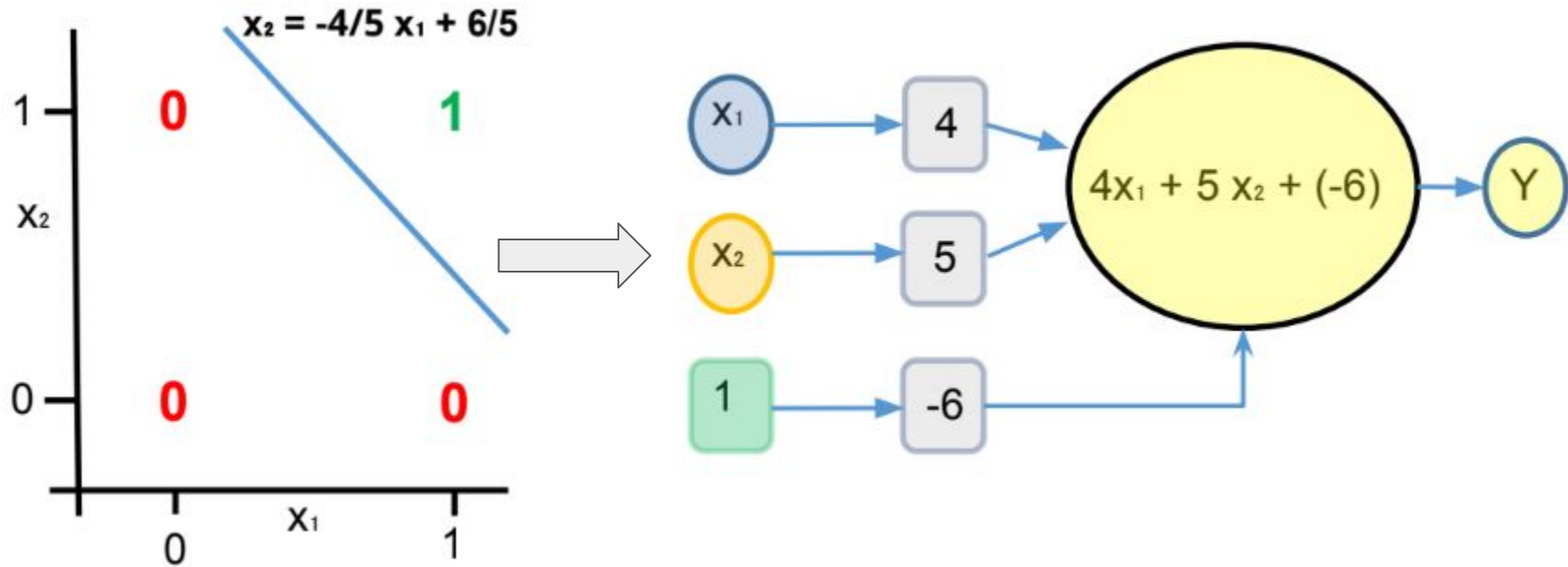
perceptrón



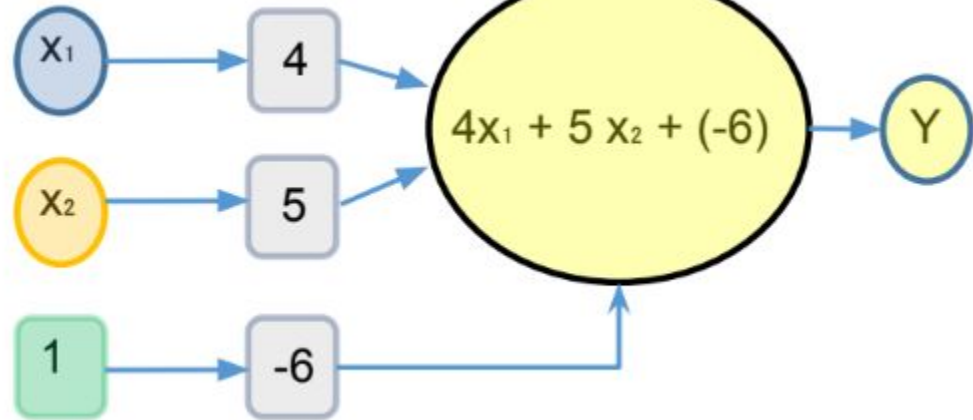
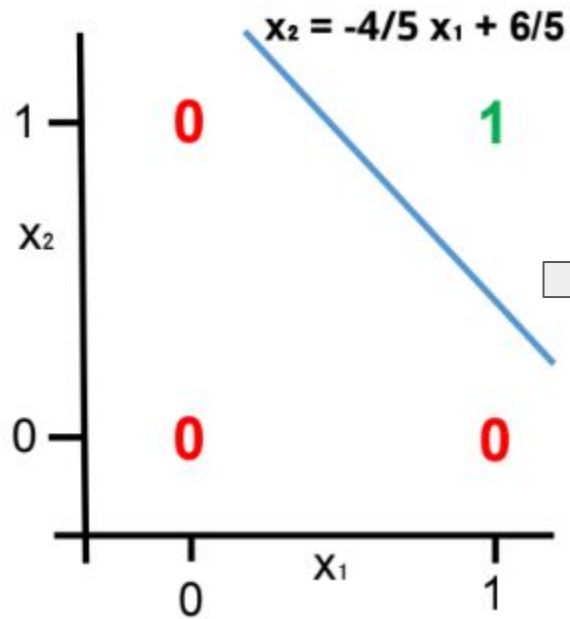
¿Cómo funciona un perceptrón?



Separar los ejemplos en clases

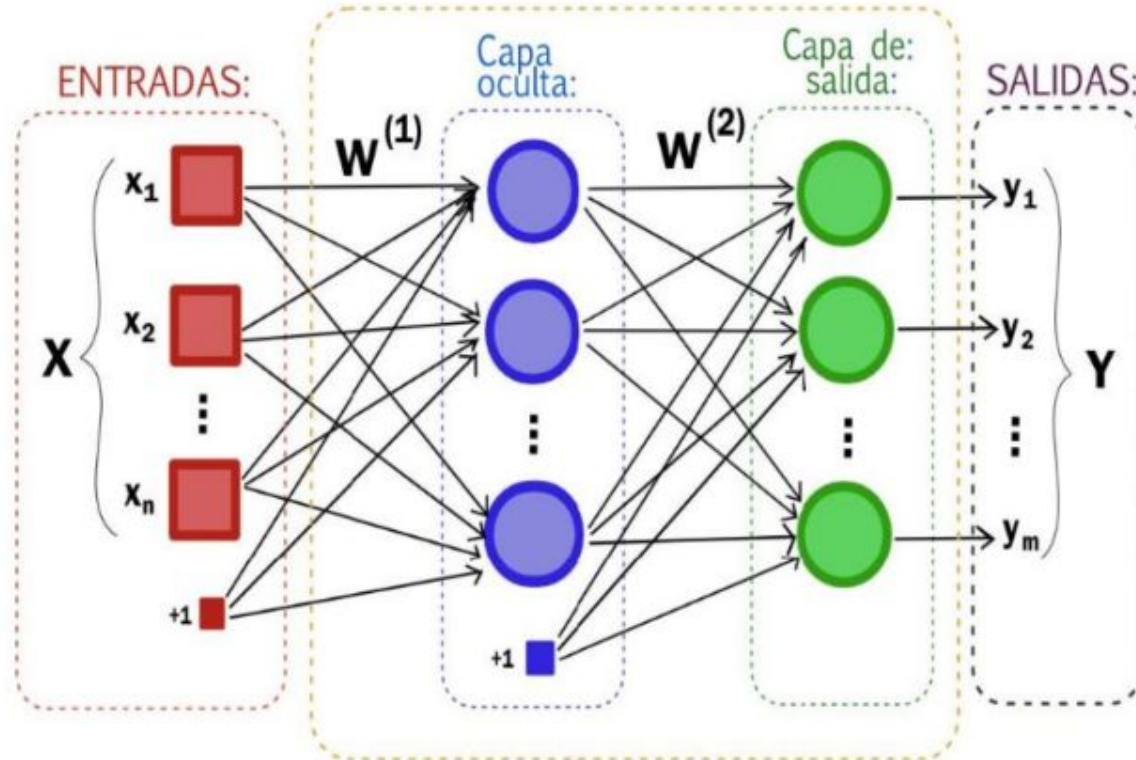


Separar los ejemplos en clases



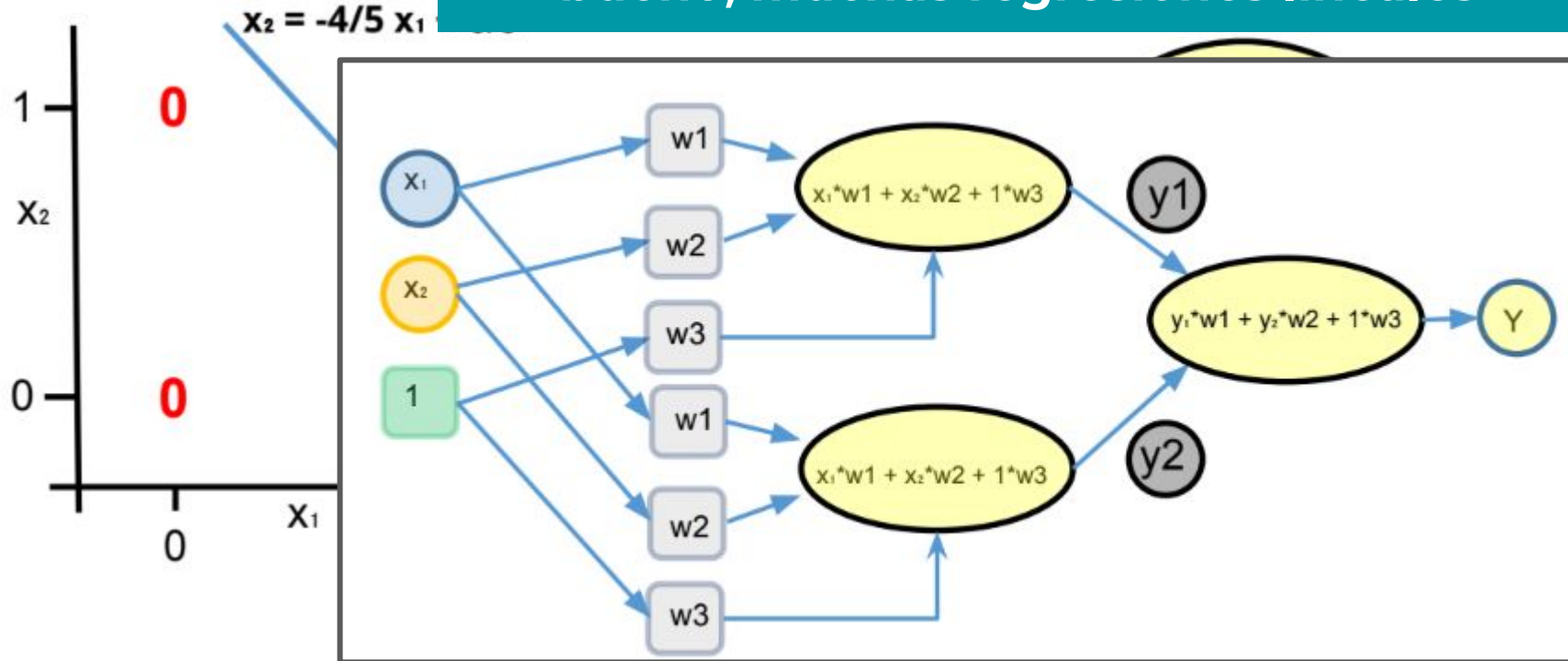
¡ una regresión lineal !

¿Cómo funciona una red neuronal?



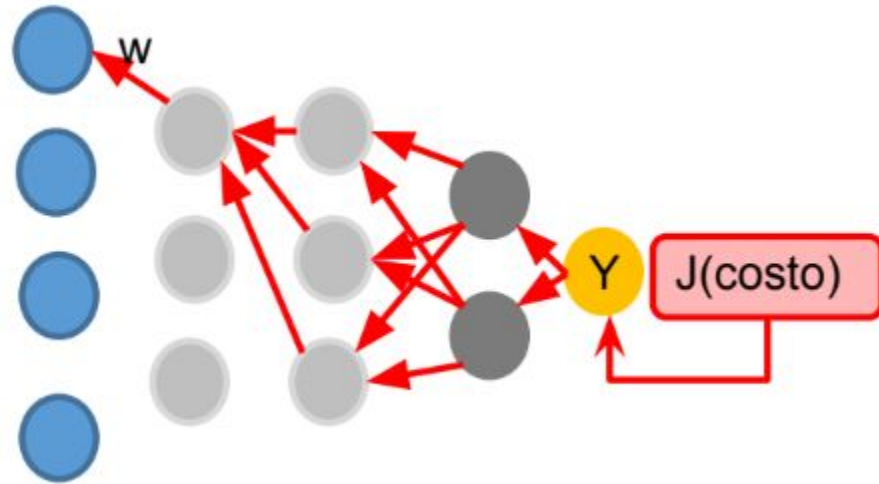
Separar los ejemplos en clases

bueno, muchas regresiones lineales



Cómo se entrena una red neuronal

algoritmo de retropropagación (*backpropagation*)



para jugar un rato

<https://playground.tensorflow.org/>

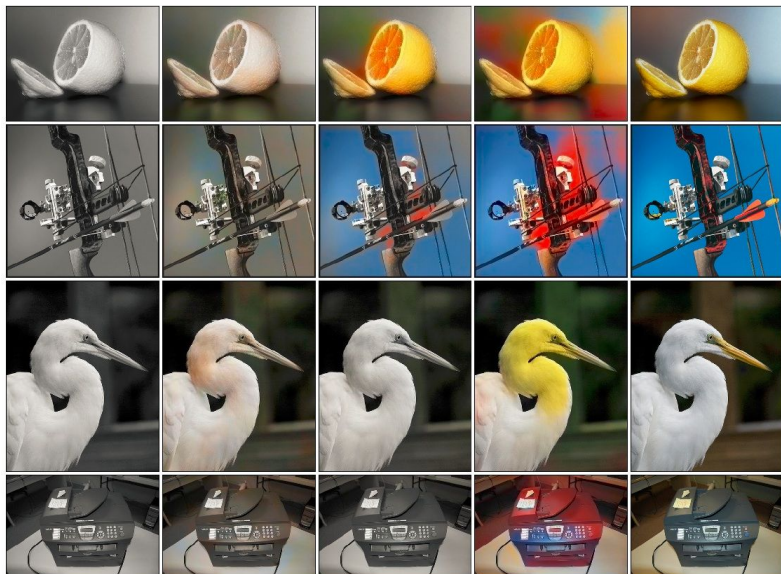
Embeddings neuronales

- Entrenar una red neuronal con una **tarea de pretexto** para la que tenemos muchos ejemplos naturalmente
 - Predecir una palabra dado su contexto, o un contexto dada una palabra
 - Reconstruir una imagen
- Eliminar la capa de predicción de la red
- La capa anterior a la de predicción es la nueva caracterización de los objetos
 - Menos características → acercándonos a las causas latentes!
- Se usa la red para convertir los objetos del espacio original al espacio de embeddings
- Es relativamente barato de obtener
- Ahora podemos caracterizar datos supervisados con información poblacional de grandes cantidades de datos no supervisados

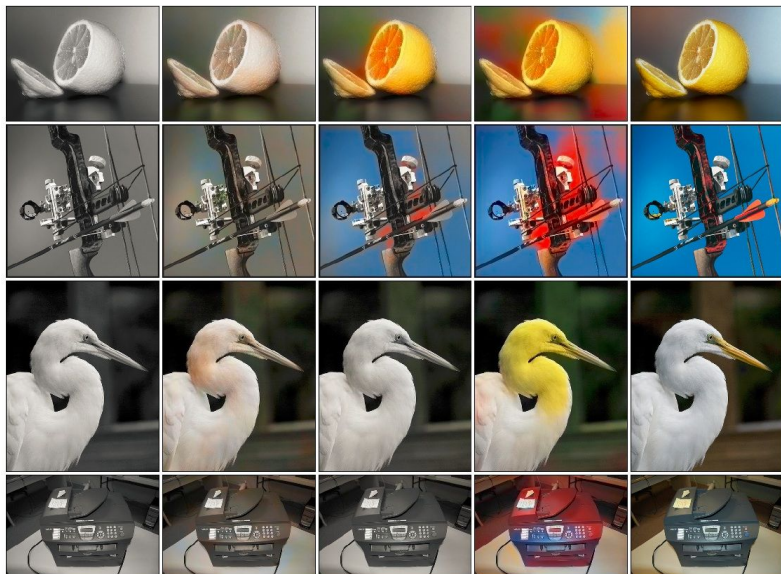
Tareas de pretexto



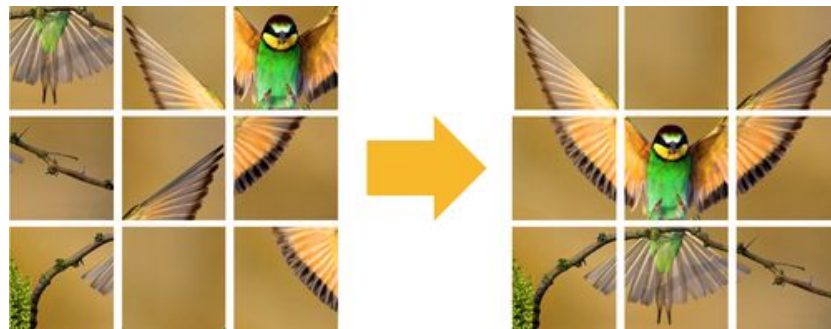
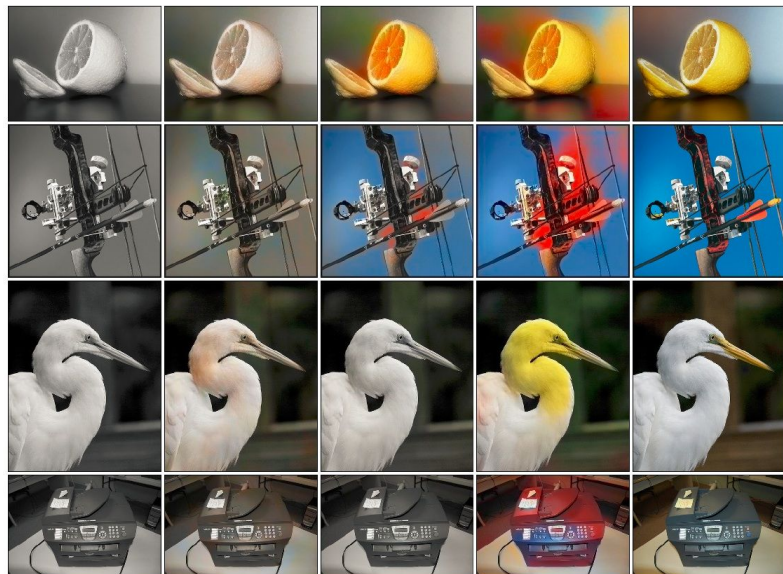
Tareas de pretexto



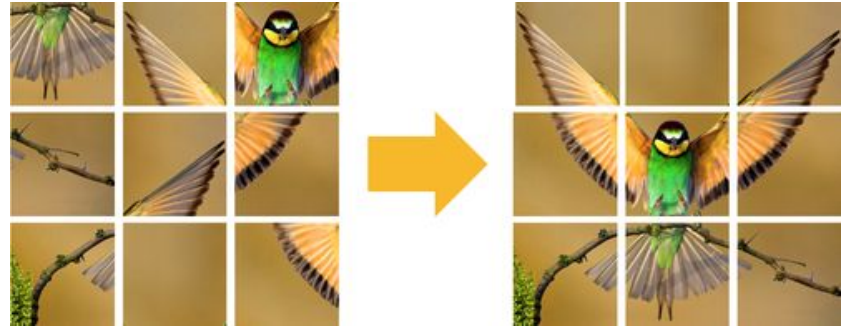
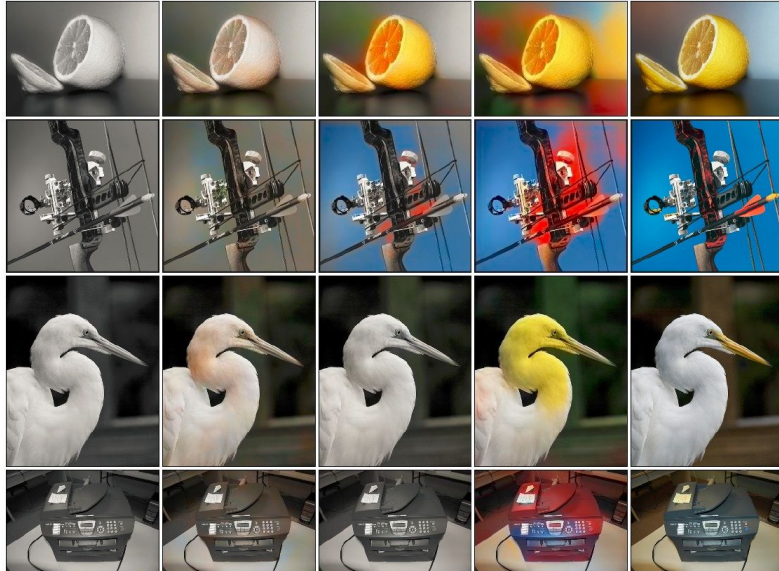
Tareas de pretexto



Tareas de pretexto

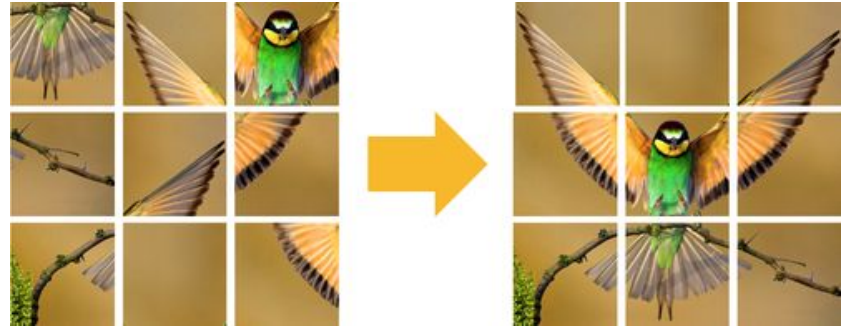
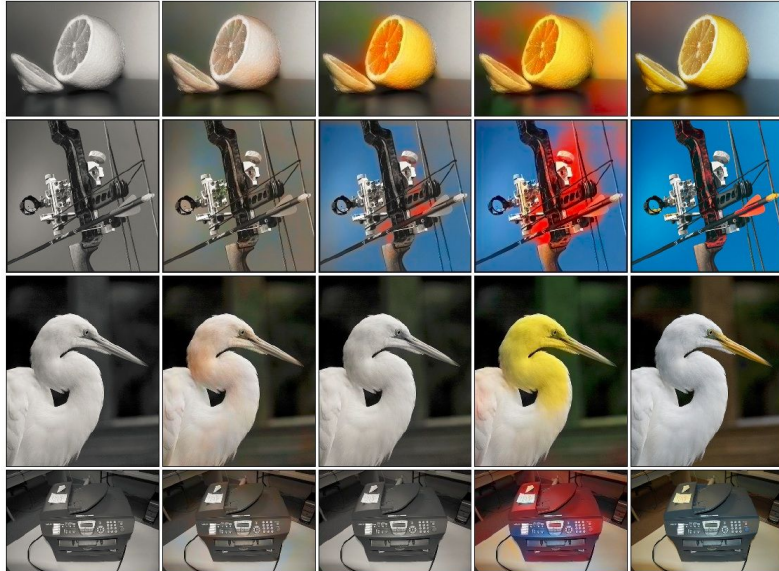


Tareas de pretexto



el gato come pescado

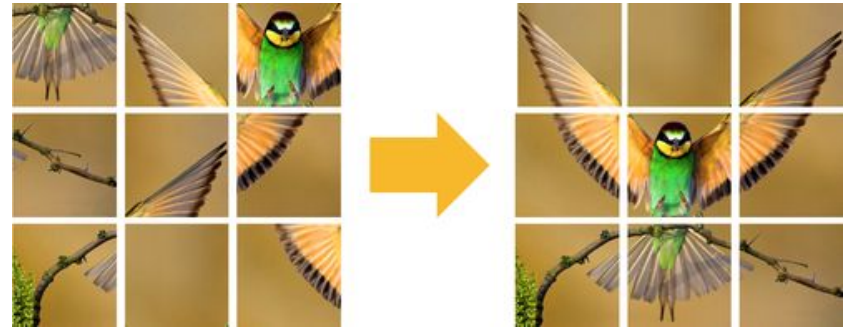
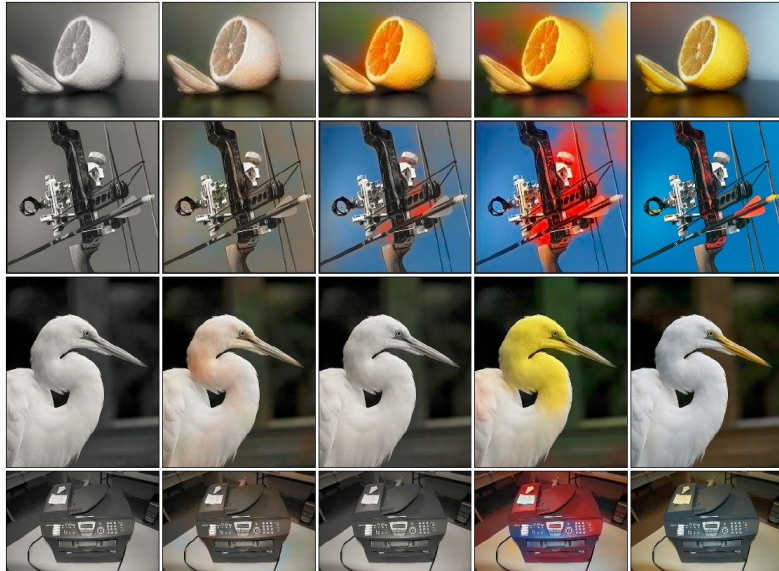
Tareas de pretexto



el gato come pescado

__	gato	come	pescado	?
el	__	come	pescado	?
el	gato	__	pescado	?
el	gato	come	__	?

mejores representaciones



el gato come pescado

__	gato	come	pescado	?	el
el	__	come	pescado	?	gato
el	gato	__	pescado	?	come
el	gato	come	__	?	pescado

¿y los modelos de lenguaje?

