This Photo by Unknown author is licensed under CC BY-SA-NC.

Machine Learning

Agenda

El Problema del Aprendizaje

Estructura del Aprendizaje

Tipos de Aprendizaje

Machine Learning

El problema del Aprendizaje

El Problema del Aprendizaje

- ¿Qué elementos se encuentran presentes en la imagen?
- ¿Pueden dar una definición para cada elemento identificado?

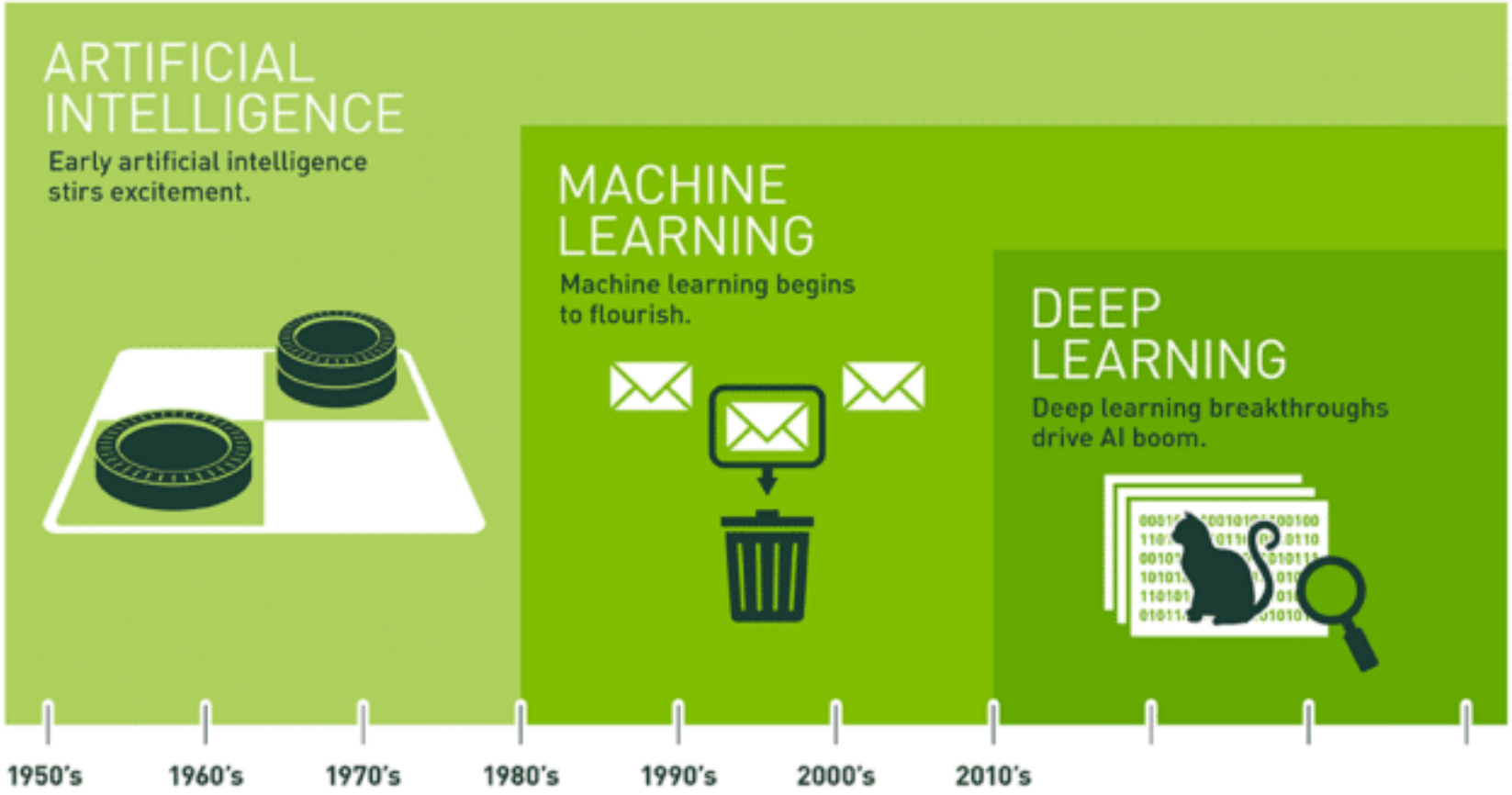
El Problema del Aprendizaje

- No aprendemos por medio de **definiciones rigurosas**.
- Aprendemos con **ejemplos**.
- Es decir, se **aprende por medio de datos o ejemplos** (*learn from data*).

El Problema del Aprendizaje

- Aprender de los datos es viable si no existe una solución analítica.
- Existen datos para aproximar una solución.
- Ciencia, ingeniería, economía, finanzas, etc.

Machine Learning



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

El Problema del Aprendizaje

Problema de sistemas de recomendación para películas

- ¿Cómo puede un sistema recomendar películas a los usuarios?
- Los criterios de cada persona son distintos y muy diversos, complejos.
- Modelarlo suena complicado, desde el punto de vista analítico.
- ¿Existe una solución empírica?

El Problema del Aprendizaje

Componentes del Aprendizaje

Componentes del Aprendizaje

Créditos bancarios

- No hay una fórmula mágica para indicar si un crédito es aprobado o no.
- ¡Es un candidato para aprender de los datos!

Componentes del Aprendizaje

- Cada dato se representa como una variable x (*la información del usuario que solicita el crédito*).
- Cada posible resultado de cada dato x se representa como y.
- La fórmula que nos permite determinar si se aprueba un crédito o no:

$$f: \chi \rightarrow \Upsilon$$

donde

- χ representa el espacio de los datos de entrada x.
  - $x \in \chi$
- Υ es el espacio de los resultados, en este caso sí (*es aprobado*) o no.
  - $y \in \Upsilon$

Componentes del Aprendizaje

El conjunto de datos $\mathfrak{D}$ recopila todos los datos x que tenemos a la mano, de la forma

$$(x_1, y_1), \ldots, (x_n, y_n)$$

donde

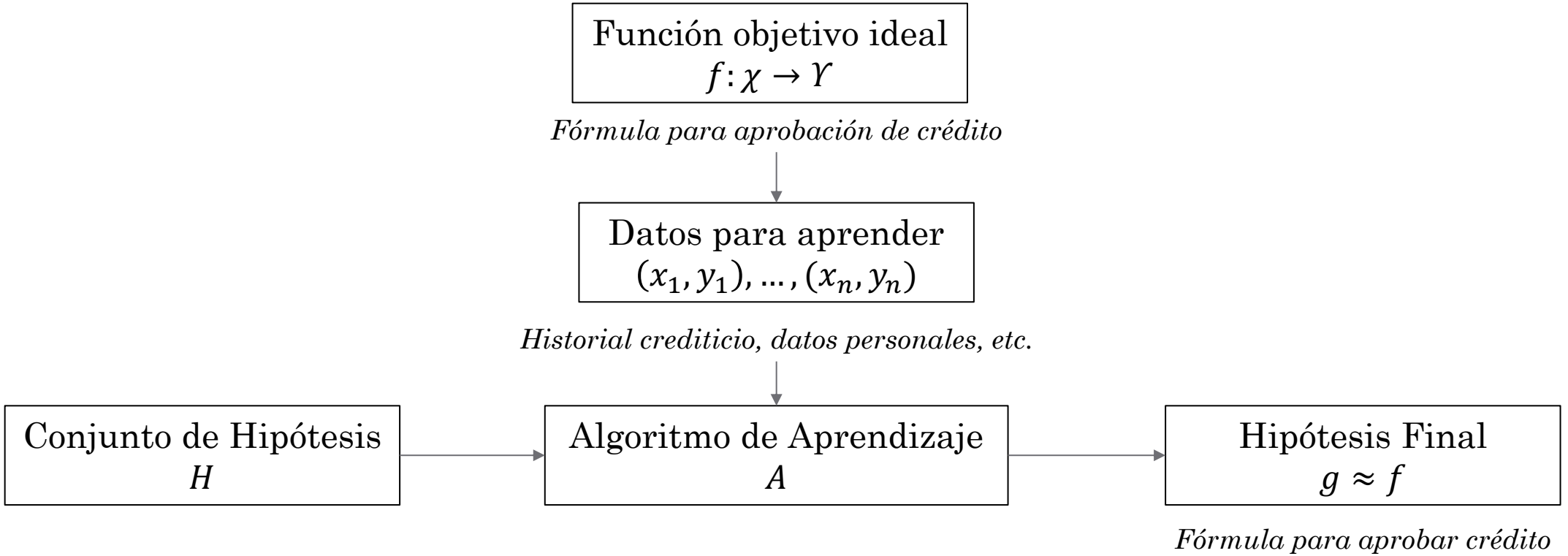
$$y_n = f(x_n)$$

para $n = 1, \ldots, N$

Componentes del Aprendizaje

- En práctica, es imposible determinar f, por lo que la única opción es acercarnos a ella.
- H es el espacio de todas las posibles funciones o reglas que se acercan a f. Unas se pueden acercar más que otras.
- Para encontrar $g \approx f$, utilizamos un algoritmo o método de aprendizaje que nos permite utilizar los datos para aprender esa regla de clasificación.

Componentes del Aprendizaje



Componentes del Aprendizaje

Ejercicio #1:

Consideren el problema para determinar si un correo es spam o no.

1. ¿Cuáles son los datos de entrada? (χ)
2. ¿Cuáles son las posibles salidas? (Y)
3. ¿Qué características debe tener el conjunto de datos?

Componentes del Aprendizaje

Ejercicio #2:

Consideren el problema para determinar un diagnóstico medico.

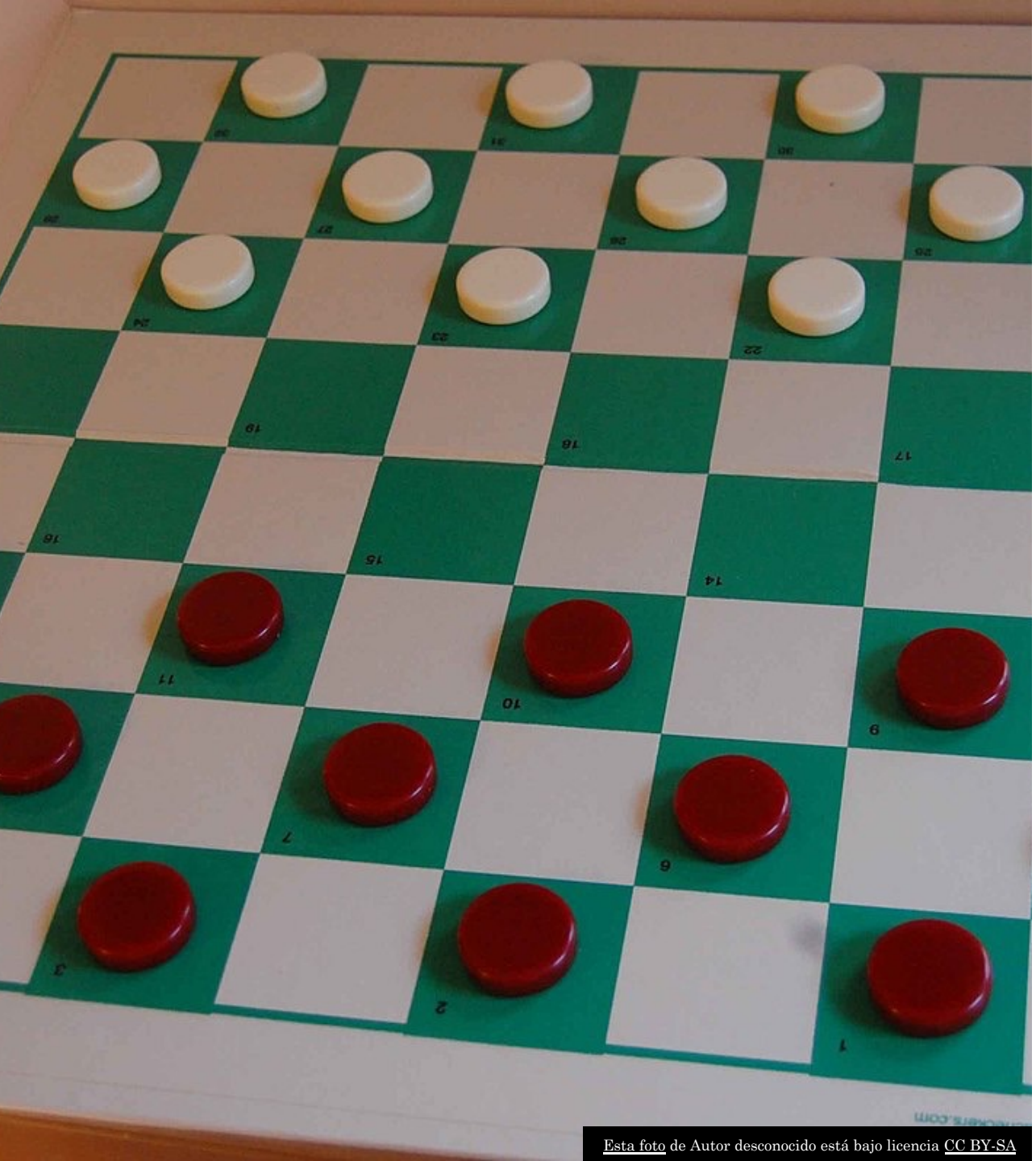
1. ¿Cuáles son los datos de entrada? (χ)
2. ¿Cuáles son las posibles salidas? (Y)
3. ¿Qué características debe tener el conjunto de datos?

Componentes del Aprendizaje

Ejercicio #3:

Consideren el problema para determinar la polaridad de opinión en un mensaje:

1. ¿Cuáles son los datos de entrada? (χ)
2. ¿Cuáles son las posibles salidas? (Y)
3. ¿Qué características debe tener el conjunto de datos?

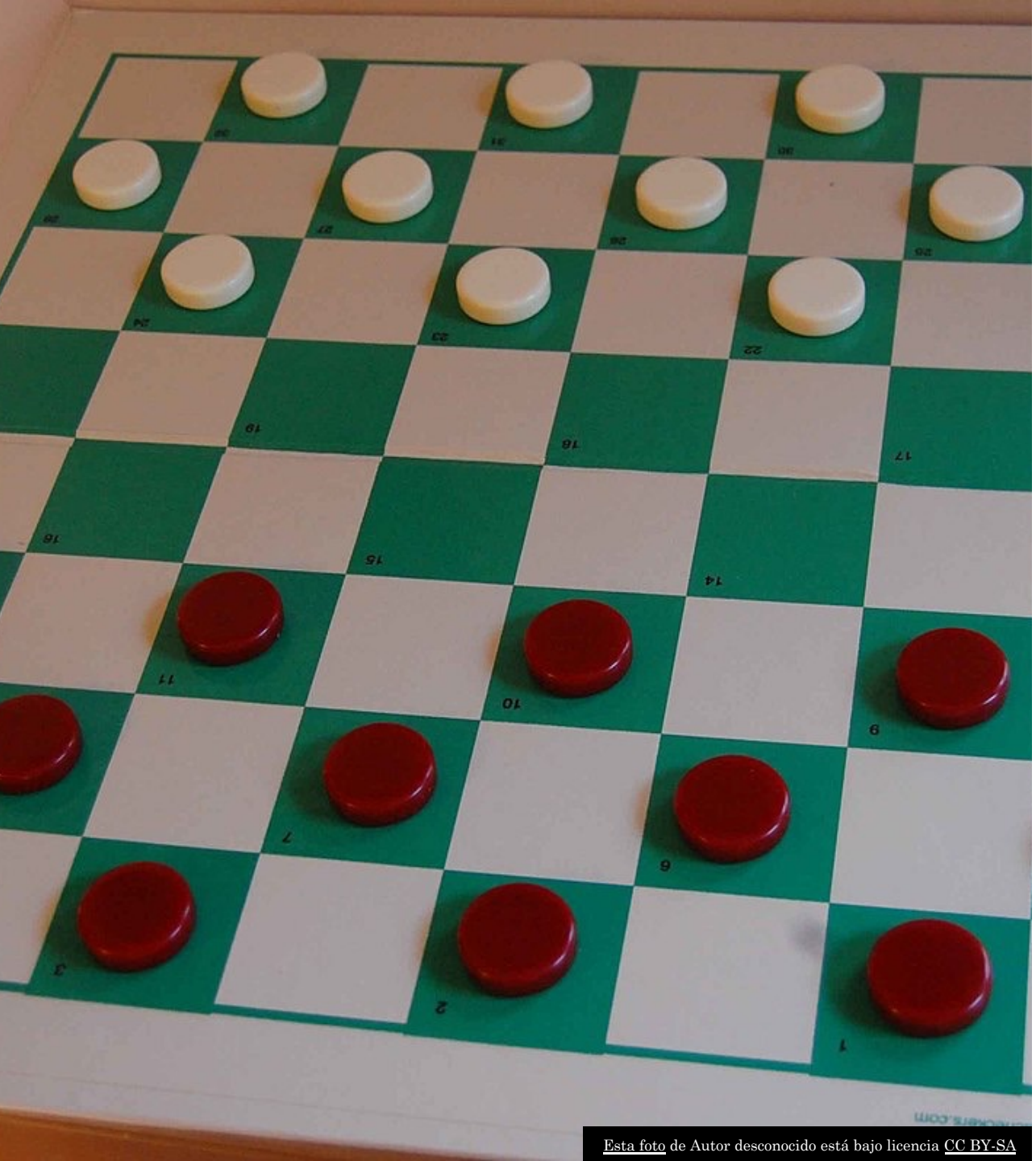


Esta foto de Autor desconocido está bajo licencia CC BY-SA

¿Qué es el Machine Learning? (informal)

Arthur Samuel (1959)

"Campo de estudio que permite que las computadoras sean capaces de aprender sin ser programadas explicitamente"



Esta foto de Autor desconocido está bajo licencia CC BY-SA

¿Qué es el Machine Learning? (informal)

- ¿Cómo mejoran su habilidad en un juego?
- Considerando la velocidad de aprendizaje, una computadora puede aprender más rápido que nosotros.
- Al final, puede resultar mejor que nosotros.

¿Qué es el Machine Learning? (formal)

Tom Mitchell (1998)

Un programa de computadora se dice que aprende de la experienca *E* relacionada a una tarea *T* y una medida de rendimiento *P*, si su rendimiento en *T*, medida por *P*, mejora con la experiencia *E*.

¿Qué es Machine Learning?

Ejercicio #4:

Consideren el problema de jugar Checkers:

1. ¿Qué sería *E*?
2. ¿Qué sería *T*?
3. ¿Qué sería P?

¿Qué es Machine Learning?

Ejercicio #5:

Consideren el problema de determinar si un correo es spam o no:

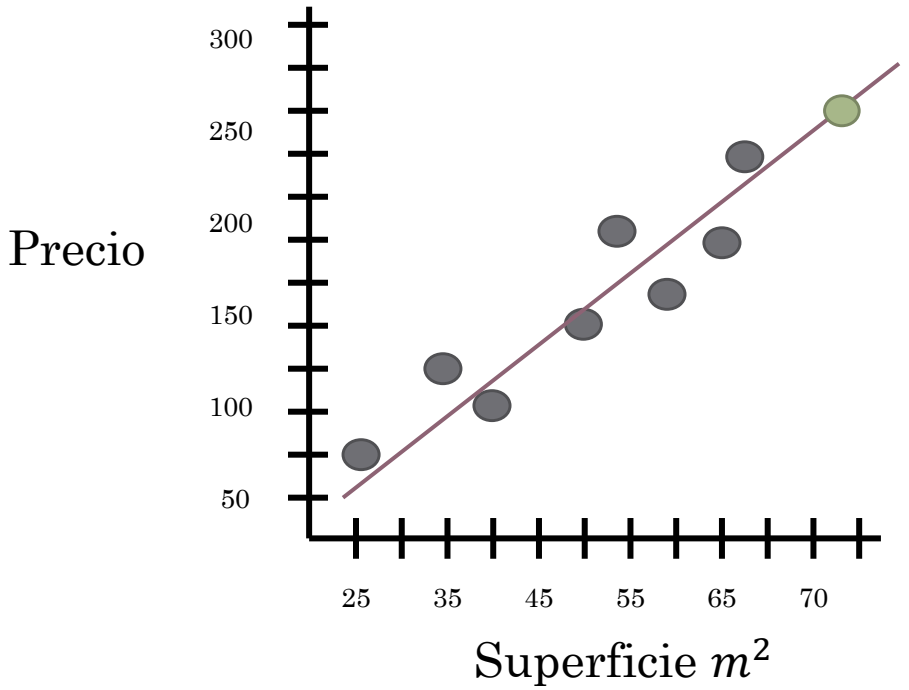
1. ¿Qué sería *E*?
2. ¿Qué sería *T*?
3. ¿Qué sería P?

Tipos de Aprendizaje

Tipos de Aprendizaje

- La premisa de aprender de los datos es **utilizar observaciones para descubrir** qué es lo que sucede en un proceso.
- ¡Es muy amplio!

Aprendizaje Supervisado



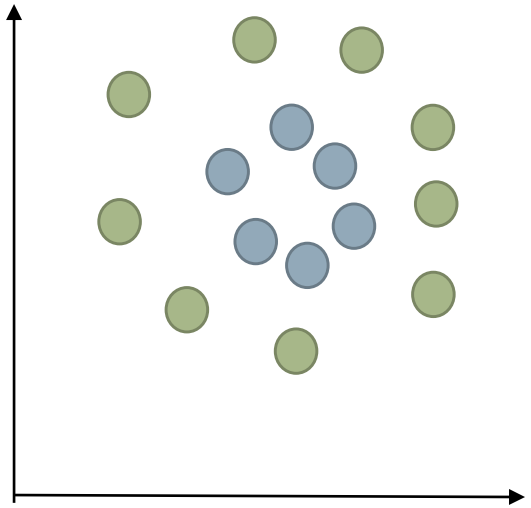
Supongamos que tenemos la siguiente información sobre el precio de la renta por metro cuadrado en una zona de la CDMX.

¿Cómo se podría determinar un nuevo valor considerando estos datos?

En este caso, estamos aprendiendo de los datos que tienen las «**respuestas correctas**».

Este es un problema de **regresión.**

Aprendizaje Supervisado



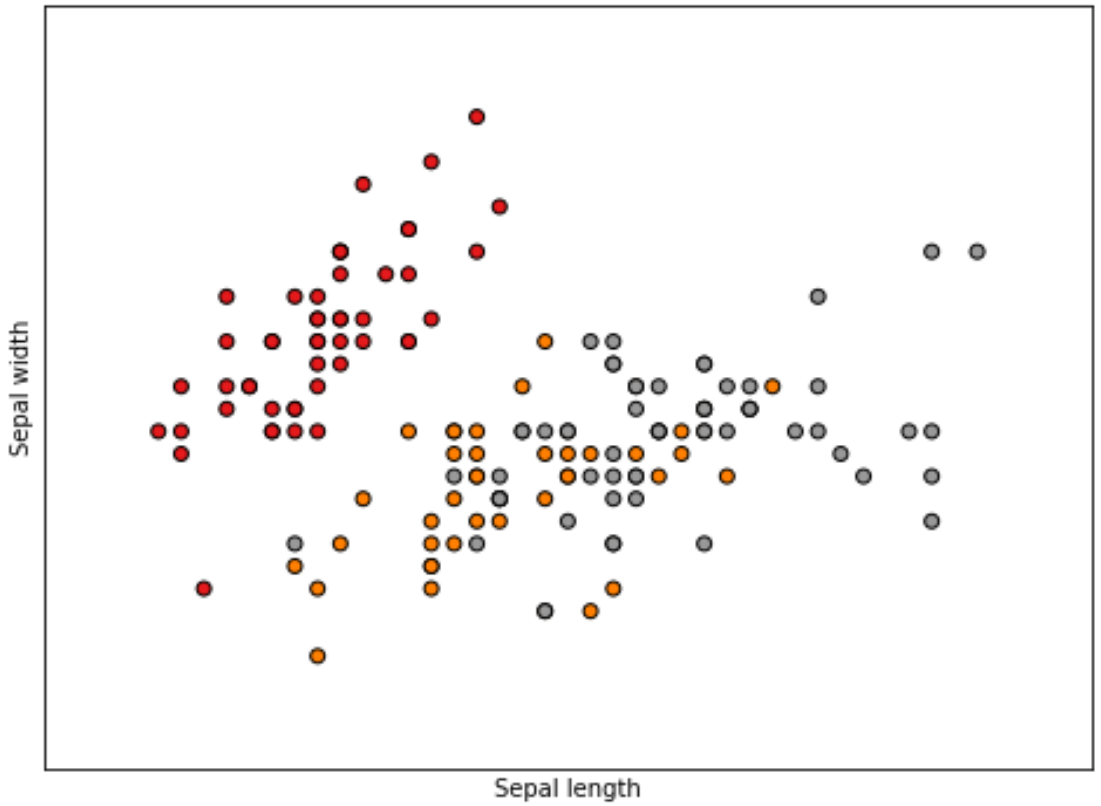
Consideremos ahora un problema de **clasificación**.

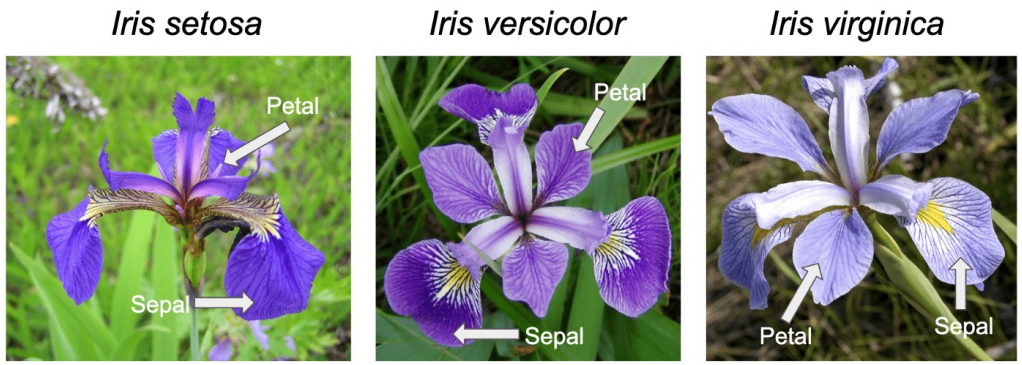
Se tienen dos (o más) clases de objetos a los cuales pertenecen cada elemento del conjunto de datos.

Usualmente se etiquetan con valores numéricos:

- 1 y 0
- 1 y -1

Aprendizaje Supervisado





Cuatro características:

- Ancho y largo del pétalo
- Ancho y largo del sépalo

¿Qué es Machine Learning?

Ejercicio #6:

Consideren el problema de determinar si un correo es spam o no:

1. ¿Es un problema de regresión o clasificación?
2. ¿Cuáles serían las clases?

¿Qué es Machine Learning?

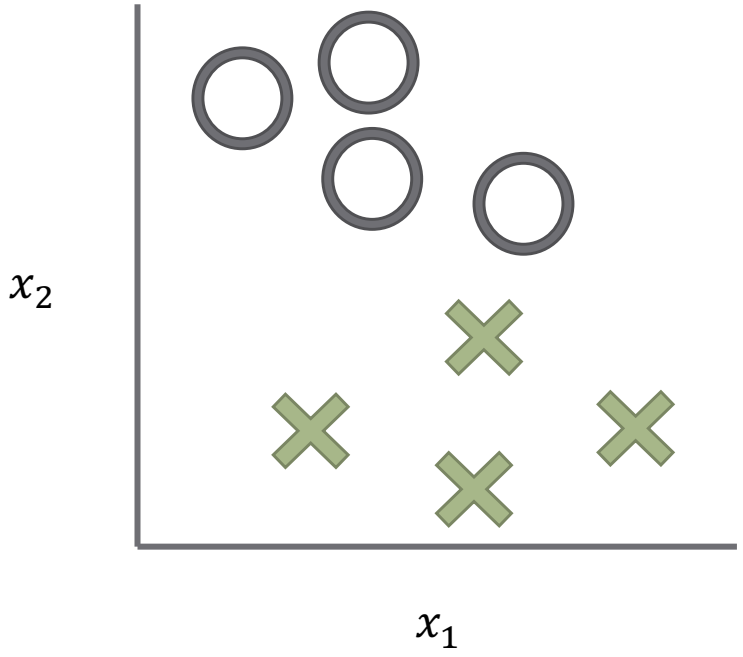
Ejercicio #7:

Consideren el problema de determinar el precio de un active financiero:

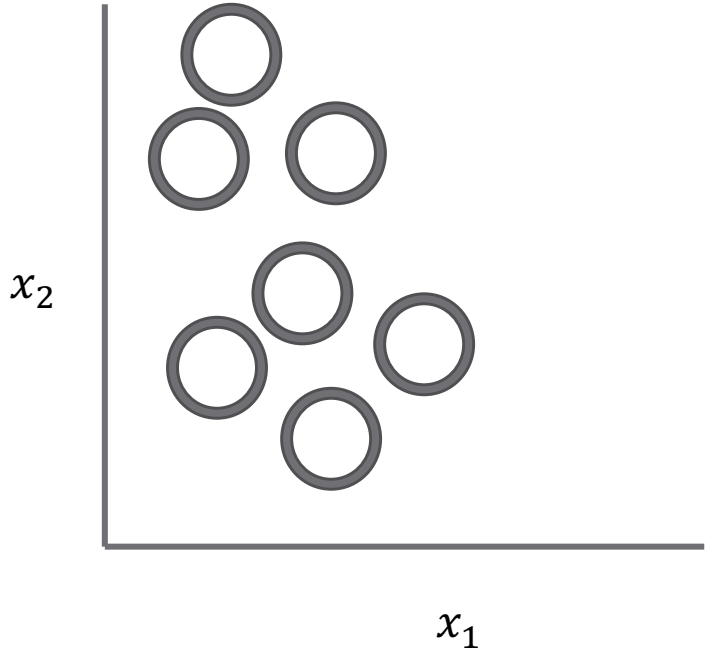
1. ¿Es un problema de regresión o clasificación?
2. ¿Cuáles serían los valores posibles para los precios?

Aprendizaje Supervisado

Aprendizaje No Supervisado



Aprendizaje Supervisado



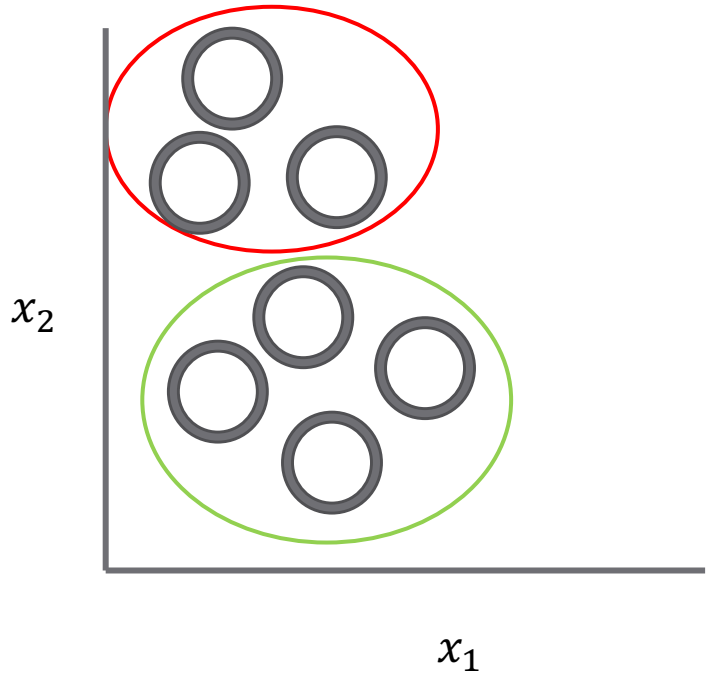
Aprendizaje No Supervisado

Aprendizaje No Supervisado

- En el aprendizaje no supervisado no se dan las clases o valores correctos de los datos.
- ¿Por qué? No siempre es posible determinar el número de clases de antemano, o es caro o difícil determinarlas.
- Aquí la tarea es encontrar estructuras o patrones en los datos.

Aprendizaje No Supervisado

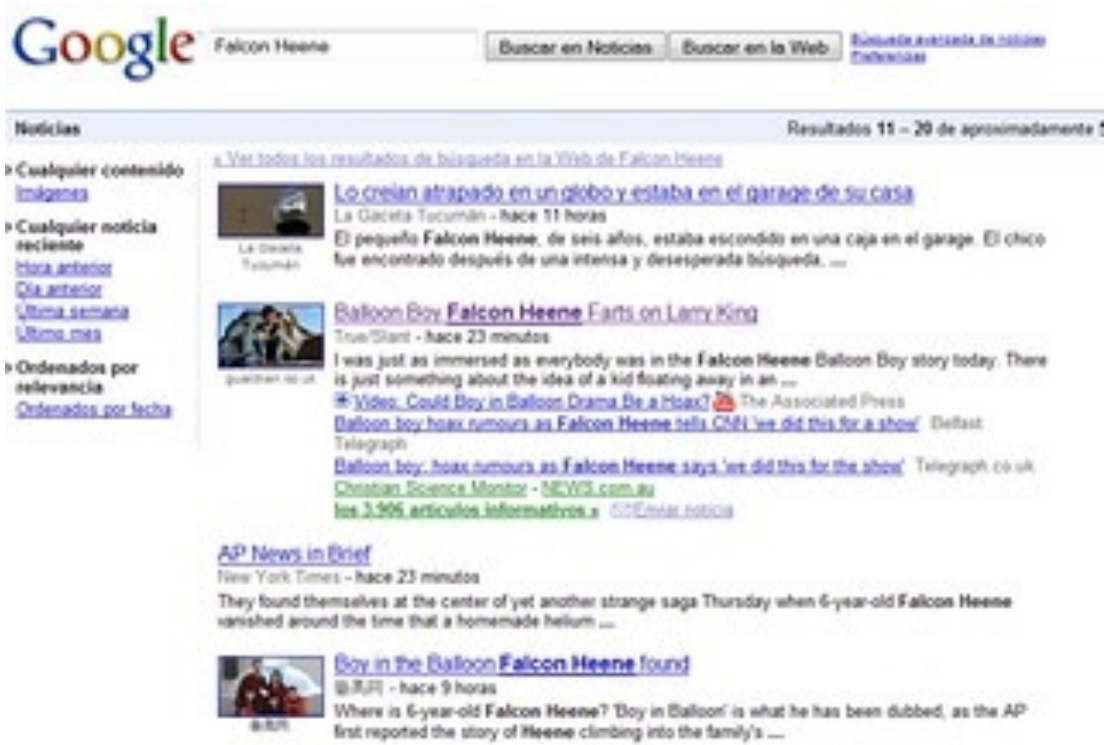
- En el aprendizaje no supervisado no se dan las clases o valores correctos de los datos.
- ¿Por qué? No siempre es posible determinar el número de clases de antemano, o es caro o difícil determinarlas.
- Aquí la tarea es encontrar estructuras o patrones en los datos.



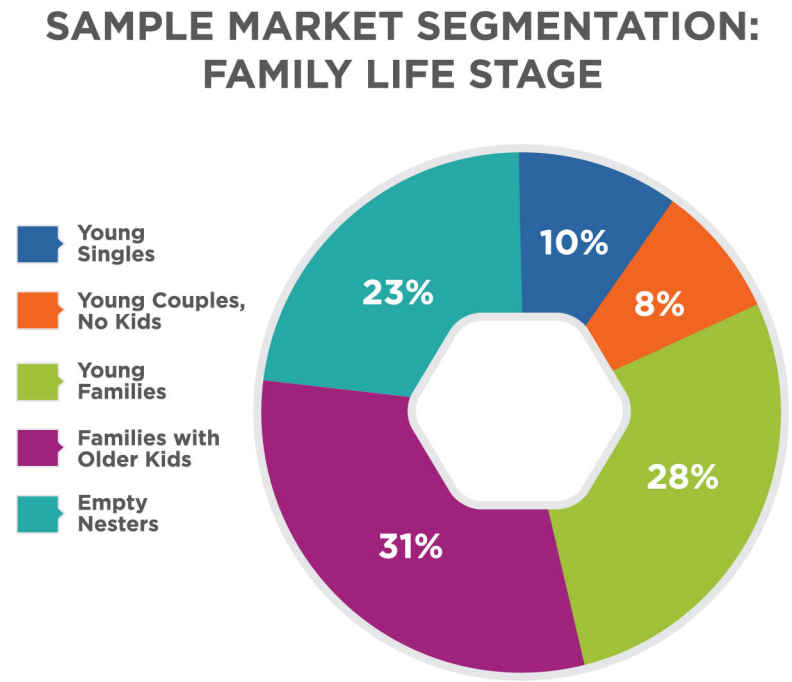
Aprendizaje No Supervisado

Una tarea común es en la clasificación de noticias:

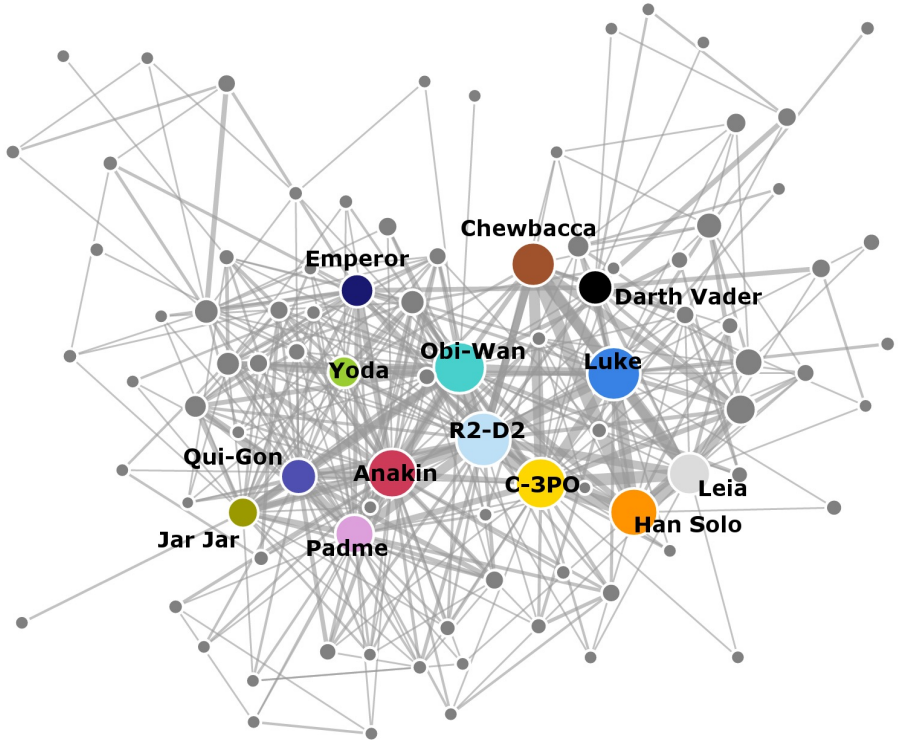
- No sabemos en cuantas clases separar cada noticias. E.g., deportes, sociales, nacional, internacional, etc.
- Una forma es determinar clústeres por medio de similitud en cuanto a temas o palabras.



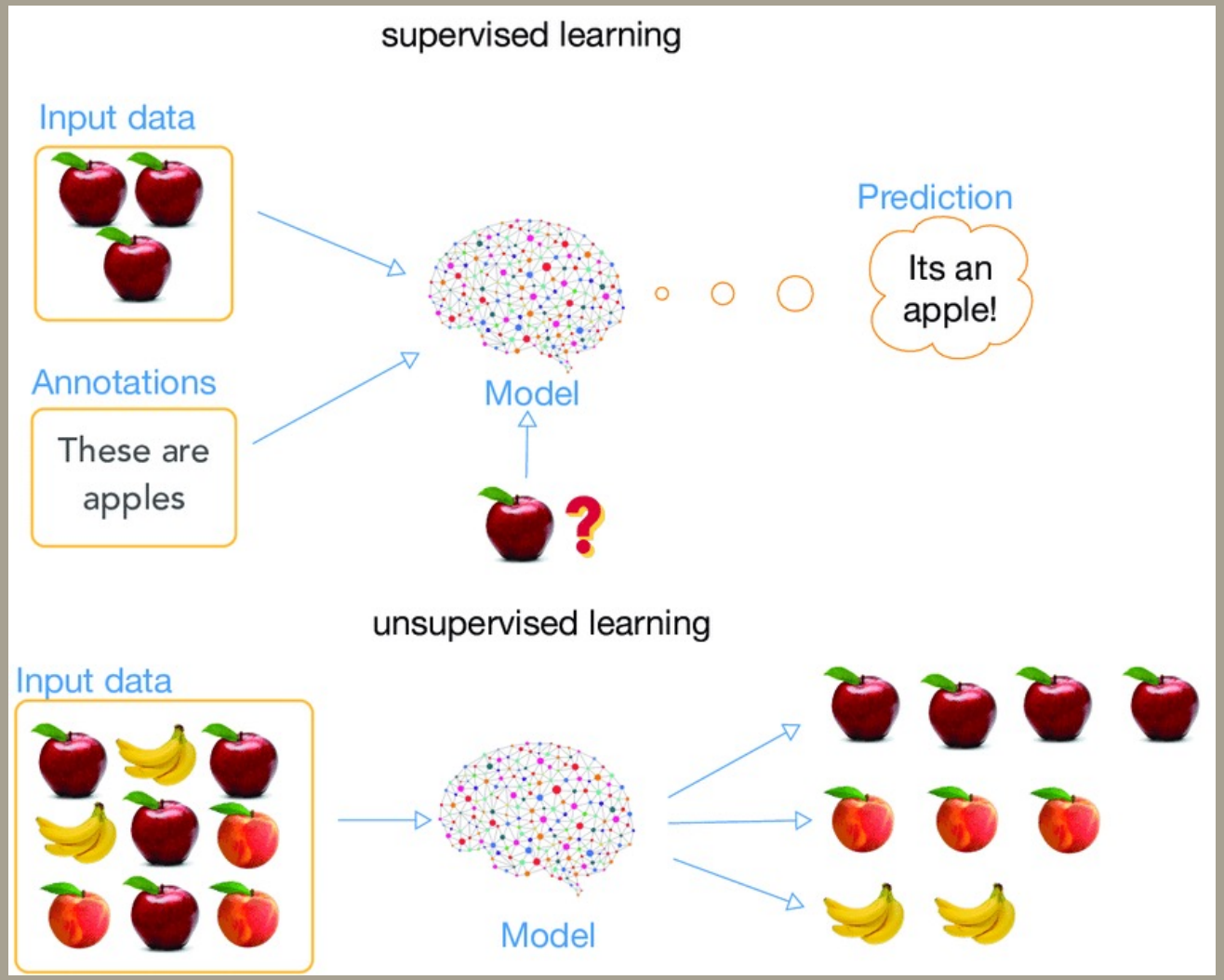
Aprendizaje No Supervisado



Esta foto de Autor desconocido está bajo licencia CC BY



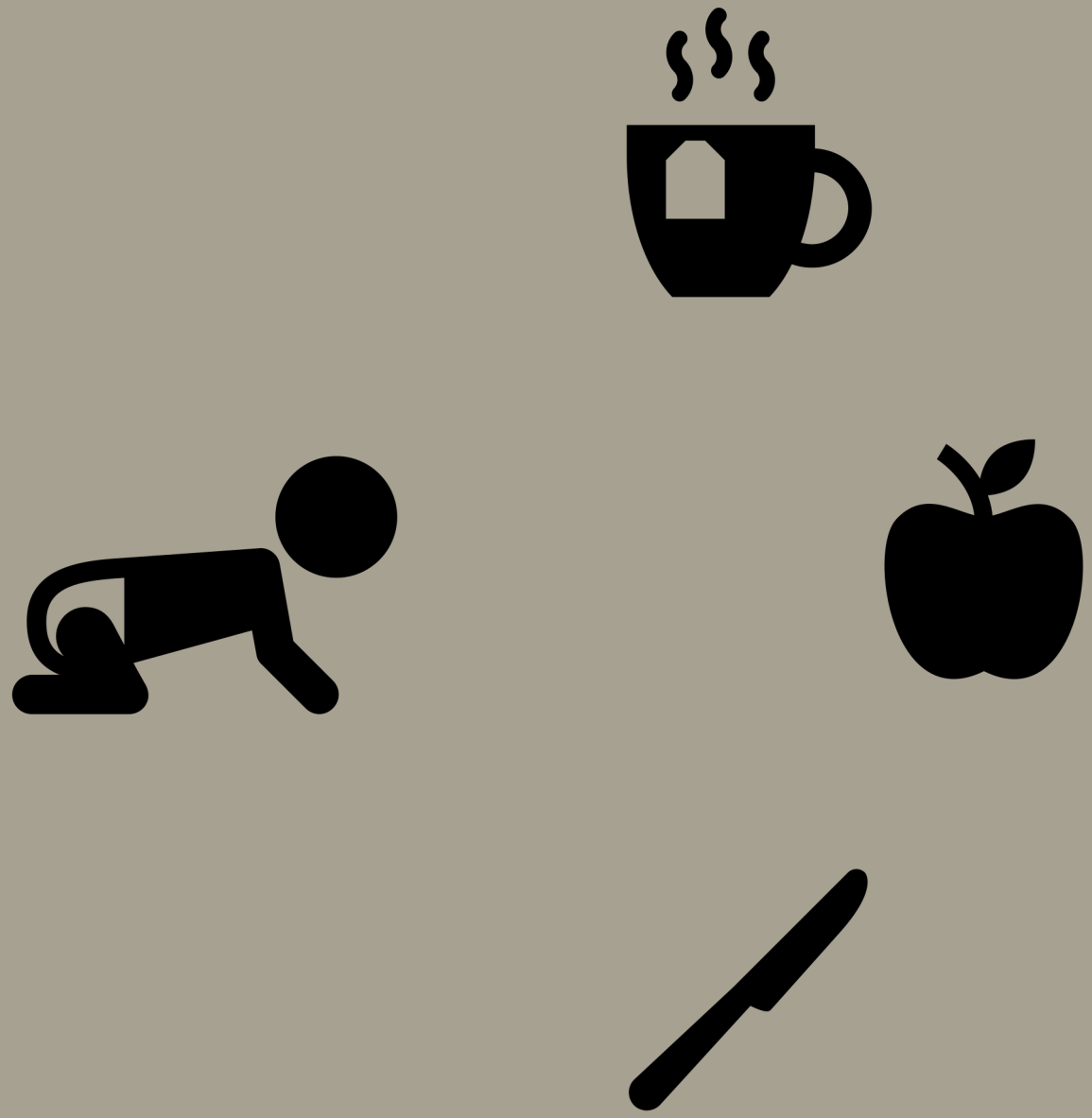
Esta foto de Autor desconocido está bajo licencia CC BY-SA



Aprendizaje Supervisado

vs

Aprendizaje No Supervisado



Aprendizaje por Refuerzo

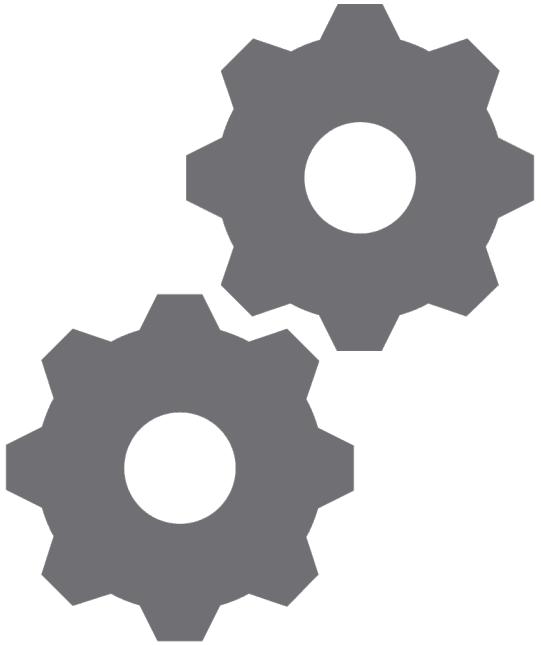
Machine Learning

Machine Learning

- Es mejor el **aprendizaje supervisado** (dar ejemplos de datos y su valor).
- Esto requiere que **se armen conjuntos de datos "grandes" y su anotación**, usualmente de forma manual.
- El científico de datos **debe proponer las características** para ayudar a dar forma a las reglas que permiten descifrar la estructura de cada dato.
  - Debe ser experto en el tema o área de aplicación.
  - Trabajar en conjunto con un experto.
- El Machine Learning **consta de métodos de aprendizaje** no tan complicados como el Deep Learning.
- Aprender es **optimizar**.

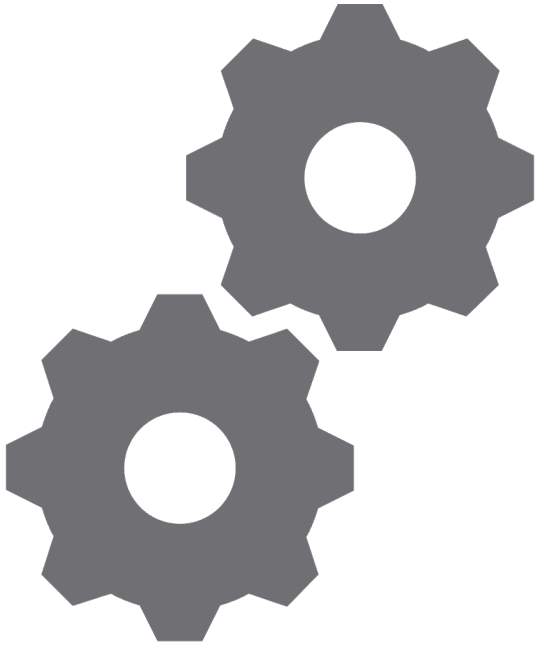
Algoritmos de Aprendizaje Supervisado

- Regresión Lineal, Polinomial y Logística
- Naïve Bayes
- Redes Neuronales (Perceptrón Multicapa)

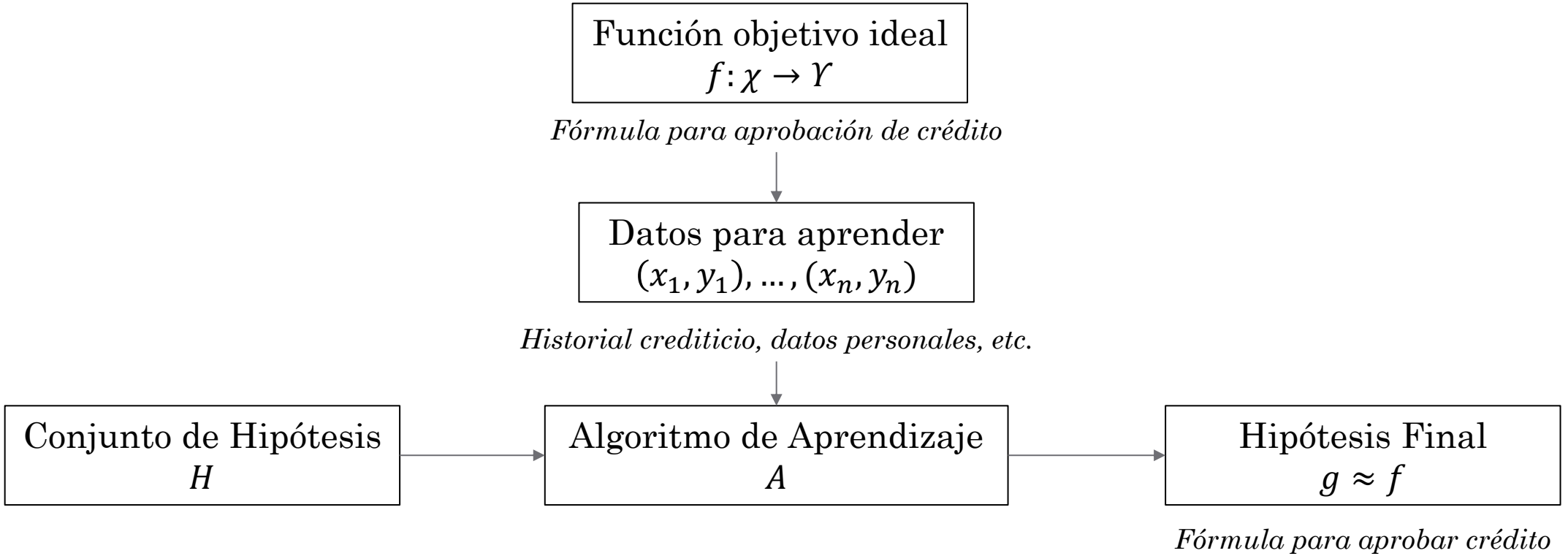


Algoritmos de Aprendizaje No Supervisado

- Métodos de clústering
  - K-Means
  - PCA



Componentes del Aprendizaje



Segunda Vuelta: Machine Learning

- Dado un problema específico, la función objetivo y los datos para entrenar los dicta el problema.
- Sin embargo, el algoritmo de aprendizaje y las hipótesis no.
- Lo anterior son **herramientas** que nosotros elegimos.
- Vamos a refrescar las cosas con un modelo muy sencillo.

Segunda Vuelta: Machine Learning

Sea $\mathrm{X} = \mathbb{R}^d$ el espacio de las entradas, y sea $y = \{-1, +1\}$ el espacio de las salidas, lo cual denota un espacio de decisión binario (sí/no, positivo/negativo, encendido/apagado, etc.).

Vamos a ponerlo en el contexto del problema para aceptar créditos bancarios:

- El vector $\boldsymbol{x} \in \mathbb{R}^d$ denota las características que se utilizan para modelar: salario, sexo, deudas actuales, estado en buró de crédito, etc.
- La variable de salida y correspondiente a cada vector $\boldsymbol{x}$ representa si se aceptó o se denegó el crédito.

Segunda Vuelta: Machine Learning

- Vamos a especificar el conjunto de hipótesis H mediane una forma funcional que todas las hipótesis $h \in H$ comparten, la cual asigna diferentes pesos a cada entrada de $\boldsymbol{x}$, lo cual refleja su importancia o contribución relativa para la decisión final.
- La ponderación se combina para determinar una calificación crediticia y se compara con un valor límite.
- Si pasa el límite, se acepta el crédito, si no, se niega:

$$\text{Se aprueba si} \quad \sum_{i=1}^{d} w_i x_i > \text{límite}$$

$$\text{Se niega si} \quad \sum_{i=1}^{d} w_i x_i < \text{límite}$$

Segunda Vuelta: Machine Learning

Ejercicio:

¿Cómo se puede escribir esto de forma compacta (i.e., en una sola expresión)?

Se aprueba si $\sum_{i=1}^{d} w_i x_i > -b$

Se niega si $\sum_{i=1}^{d} w_i x_i < -b$

Solución:

$$h(\boldsymbol{x}) = sign\left(\sum_{i=1}^{d} w_i x_i + b\right)$$

Segunda Vuelta: Machine Learning

$$h(\boldsymbol{x}) = sign\left(\sum_{i=1}^{d} w_i x_i + b\right)$$

- En este modelo, w_i son los pesos asociados a x_i.
- b es el límite que debe rebasar la ponderación, también llamado *bias*.
- $h(\boldsymbol{x}) = +1$ indica que el crédito es aprobado y $h(\boldsymbol{x}) = -1$ indica que es denegado.
- Este modelo de H se conoce como perceptrón.

Segunda Vuelta: Machine Learning

$$h(\boldsymbol{x}) = sign\left(\sum_{i=1}^{d} w_i x_i + b\right)$$

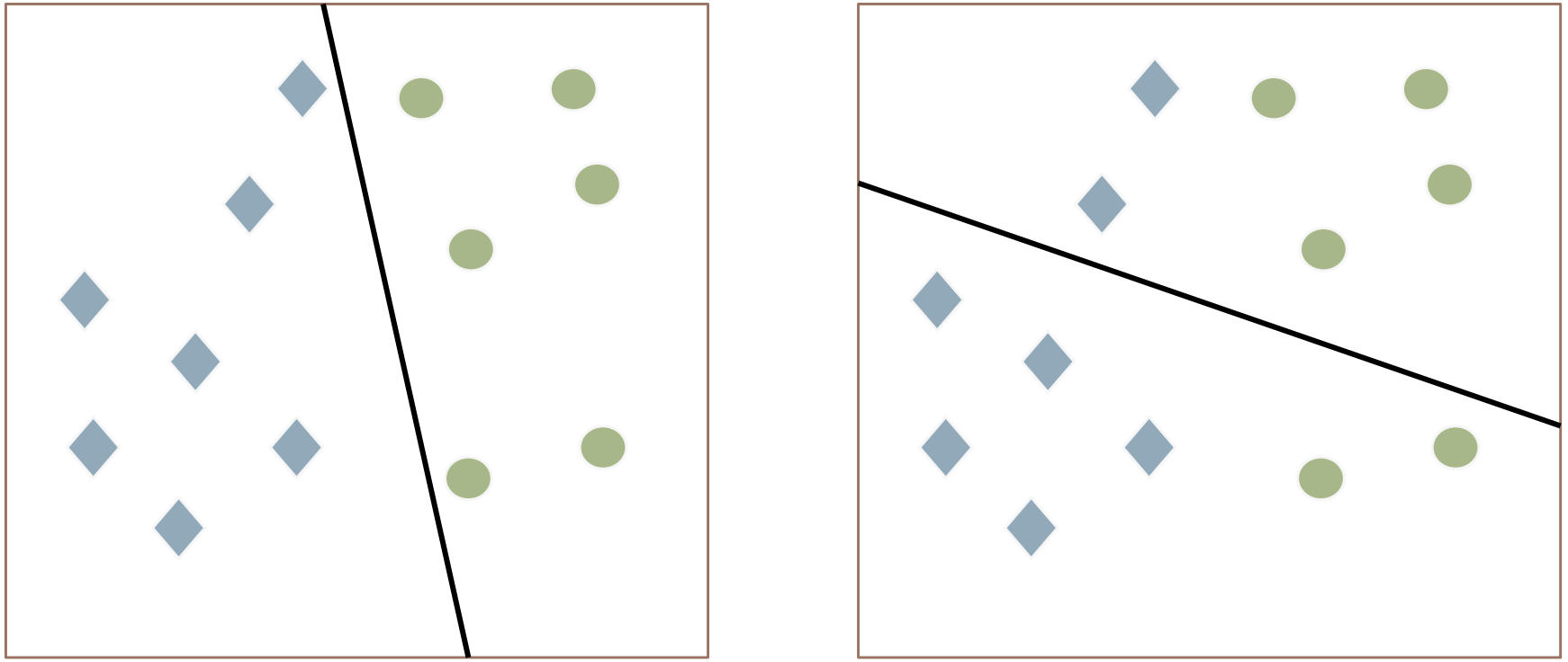
- El algoritmo de aprendizaje busca en H la mejor hipótesis mediante la comparación de pesos y bias que mejor desempeño tengan con el conjunto de datos dado.
- Los pesos w_i no deben ser necesariamente positivos. Si existe alguno negativo, indica que esa característica tiene un efecto adverso en el proceso de aprobación de crédito.
- El valor del bias puede ser grande o pequeño, depende de la actitud histórica del banco para la aprobación de créditos.

Segunda Vuelta: Machine Learning

$$h(\boldsymbol{x}) = sign\left(\sum_{i=1}^{d} w_i x_i + b\right)$$

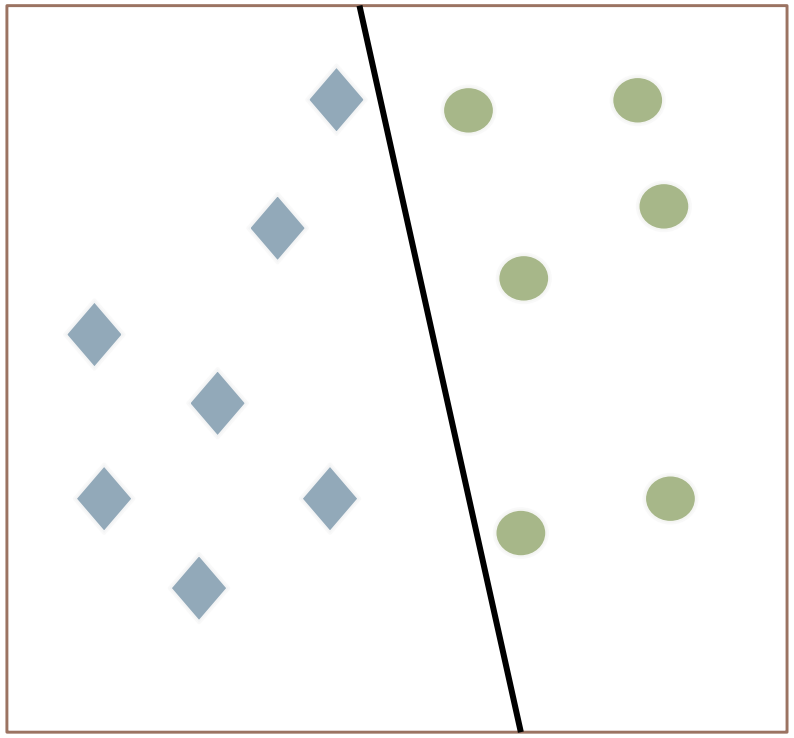
- La elección óptima de pesos y el sesgo definen la hipótesis final $g \in H$ que el algoritmo arroja como resultado.

Segunda Vuelta: Machine Learning



Clasificación realizada por el perceptrón para el caso datos linealmente separables en $\mathbb{R}^2$. Izquierda: separación perfecta. Derecha: Algunos puntos son clasificados mal.

Segunda Vuelta: Machine Learning



- El perceptrón parte el plano en dos regiones: la región +1 y la región -1.
- Valores distintos para w_1, w_2 y b dan como resultado distintas rectas $w_1 x_1 + w_2 x_2 + b = 0$.
- Si la información es linealmente separable, existen valores para los parámetros que separan perfectamente los puntos.

Segunda Vuelta: Machine Learning

Para simplificar la notación, vamos a tratar el bias b como un peso $w_0 = b$ y juntarlo con los otros pesos en un vector:

$$\boldsymbol{w} = (w_0, w_1, \dots, w_d) = (b, w_1, \dots, w_d)$$

y

$$\boldsymbol{x} = (x_0, x_1, \dots, x_d) = (1, x_1, \dots, x_d)$$

De esta manera, la fórmula se reduce a

$$h(x) = sign\left(\sum_{i=0}^{d} w_i x_i\right) = sign(\boldsymbol{w}^T \boldsymbol{x}).$$

Segunda Vuelta: Machine Learning

Con esto en mente, introducimos el **algoritmo de aprendizaje del perceptrón** (AAP), que determina el vector $\boldsymbol{w}$ basado en los datos disponibles.

Supuesto: Los datos o puntos de entrenamiento son linealmente separabales.

Segunda Vuelta: Machine Learning

El AAP busca determinar el vector $\boldsymbol{w}$ usando un método iterativo:

- En la t-ésima iteración, $t = 1,2, \ldots$, se tiene un valor de $\boldsymbol{w}$ en el paso t, llamémoslo $\boldsymbol{w}(t)$.
- El algoritmo elige un punto $\boldsymbol{x}$ mal clasificado, llamémoslo $(\boldsymbol{x}(t), y(t))$, para actualizar el peso $\boldsymbol{w}(t)$.
- Ya que el punto se encuentra mal clasificado,

$$y(t) \neq sign\left(\boldsymbol{w}^T(t)\boldsymbol{x}(t)\right)$$

- La regla de actualización es

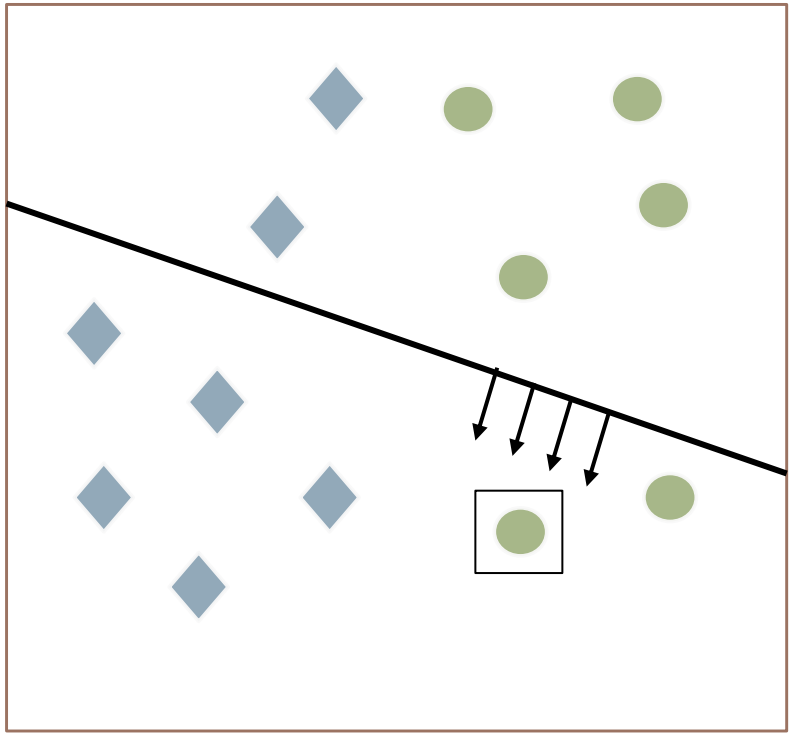
$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) + y(t)\boldsymbol{x}(t)$$

Segunda Vuelta: Machine Learning

- La regla de actualización
  $$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) + y(t)\boldsymbol{x}(t)$$

tiene el efecto de mover el borde o hiperplano de separación en la dirección de $\boldsymbol{x}(t)$ de tal manera que se clasifica correctamente.

- El algoritmo continua con iteraciones subsecuentes hasta que no existen puntos mal clasificados.



Segunda Vuelta: Machine Learning

Tarea:

- Leer del capítulo 1 del libro *Learning From Data* de Mostafa et al., las secciones 1.1 y 1.2.
- Realizar el ejercicio 1.3, página 8 del libro *Learning From Data* de Mostafa et al.
- Realizar el ejercico 1.4, página 8 del libro *Learning From Data* de Mostafa et al.

Fin de la presentación

¡Gracias por su tiempo!