



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.



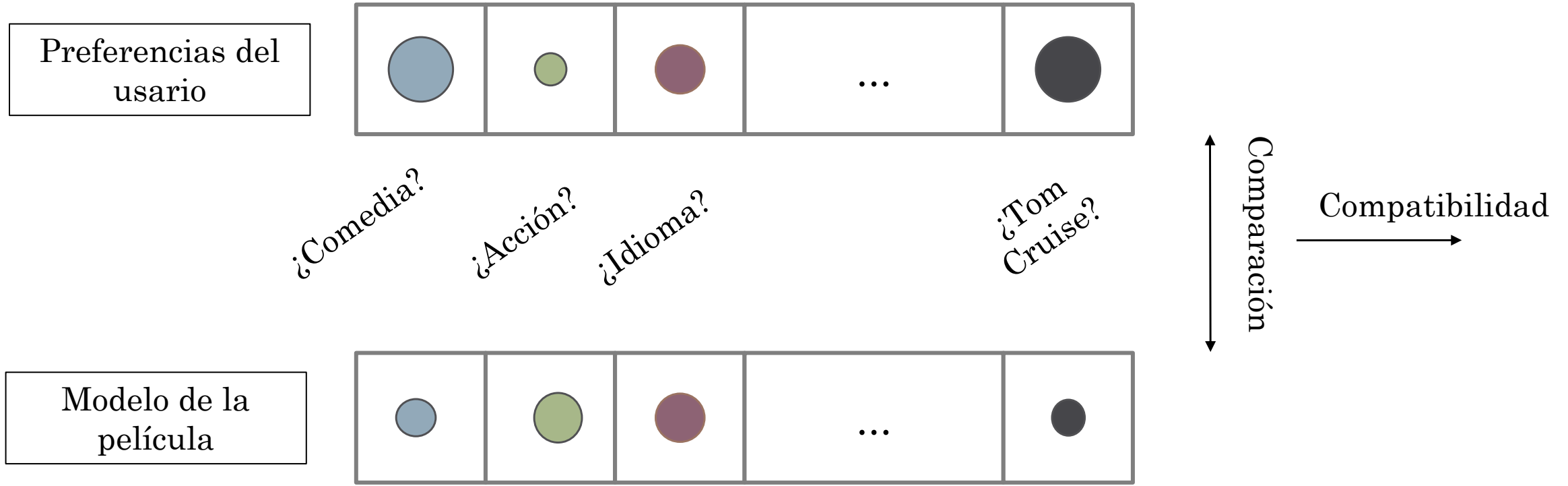
El Problema del Aprendizaje

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:

$$f: \chi \rightarrow Y$$

donde

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

Componentes del Aprendizaje

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

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

donde

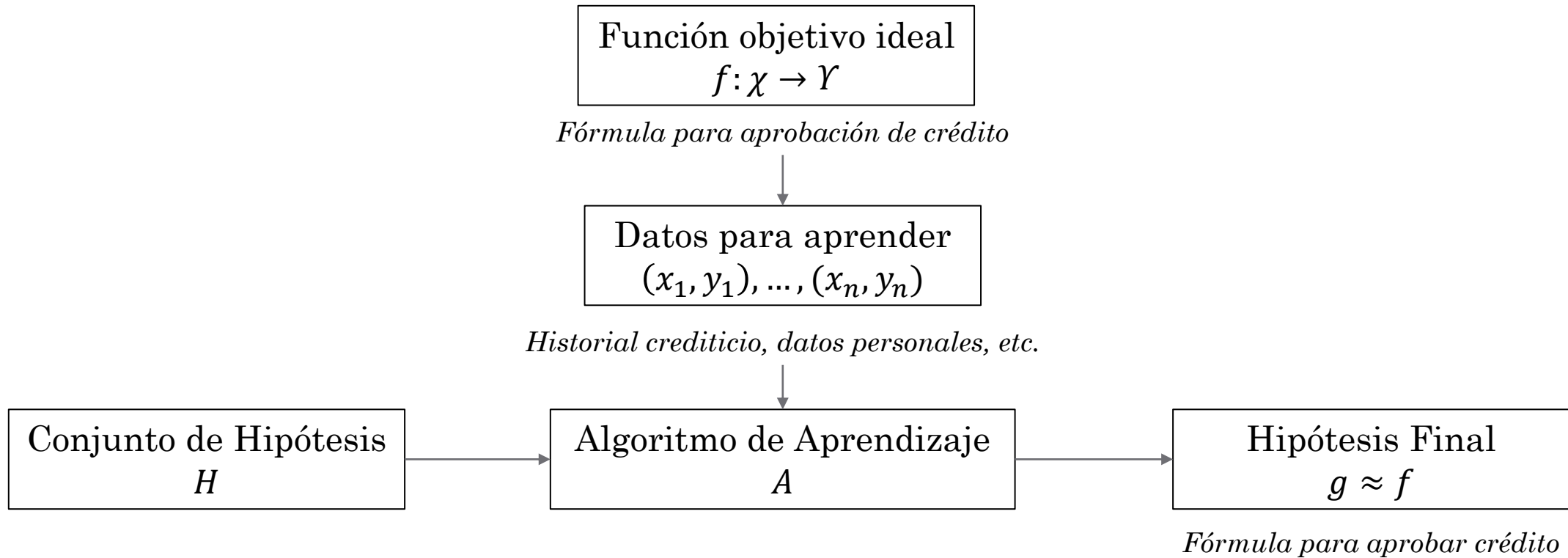
$$y_n = f(x_n)$$

para $n = 1, \dots, N$

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? (\mathcal{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? (\mathcal{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? (Υ)
3. ¿Qué características debe tener el conjunto de datos?

Tipos de Aprendizaje

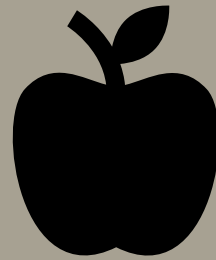
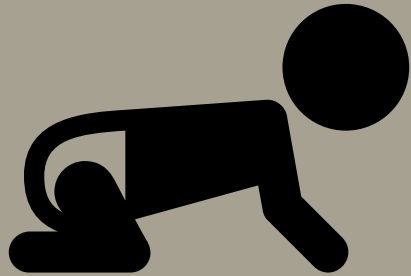


Tipos de Aprendizaje

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



Aprendizaje
Supervisado



Aprendizaje
por Refuerzo

supervised learning

Input data



Annotations

These are apples



Model

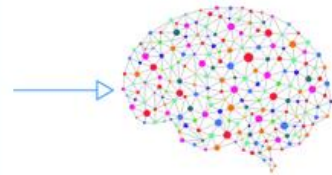


Prediction

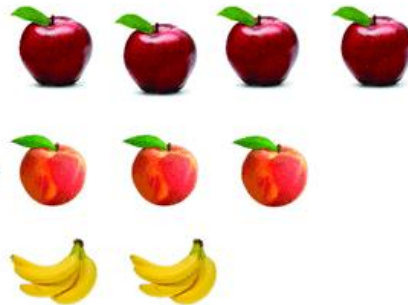
Its an apple!

unsupervised learning

Input data



Model



Aprendizaje Supervisado

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

- Máquinas de Vectores de Soporte (Support Vector Machines)
- Regresión Lineal, Polinomial y Logística
- Árboles de Decisión
- Redes Neuronales (Perceptrón Multicapa)
- Métodos de Ensamble

