



# MACHINE LEARNING INTRODUCCIÓN

Dr. Jorge Hermosillo

Laboratorio de Semántica Computacional

[jhermosillo@uaem.mx](mailto:jhermosillo@uaem.mx)



# MACHINE LEARNING O APRENDIZAJE DE AUTOMÁTICO

- ▶ Arthur Samuel (1959). **Aprendizaje de Automático:** Campo de estudio que proporciona a las computadoras la habilidad de **aprender sin ser explícitamente programadas**.
- ▶ Tom Mitchell (1998). Se dice que **un programa de computadora aprende de la Experiencia E** con **respecto a alguna Tarea T** y una **medida de desempeño P**, si su desempeño sobre T, medido por P, mejora con la experiencia E.

# EL PROBLEMA DEL APRENDIZAJE

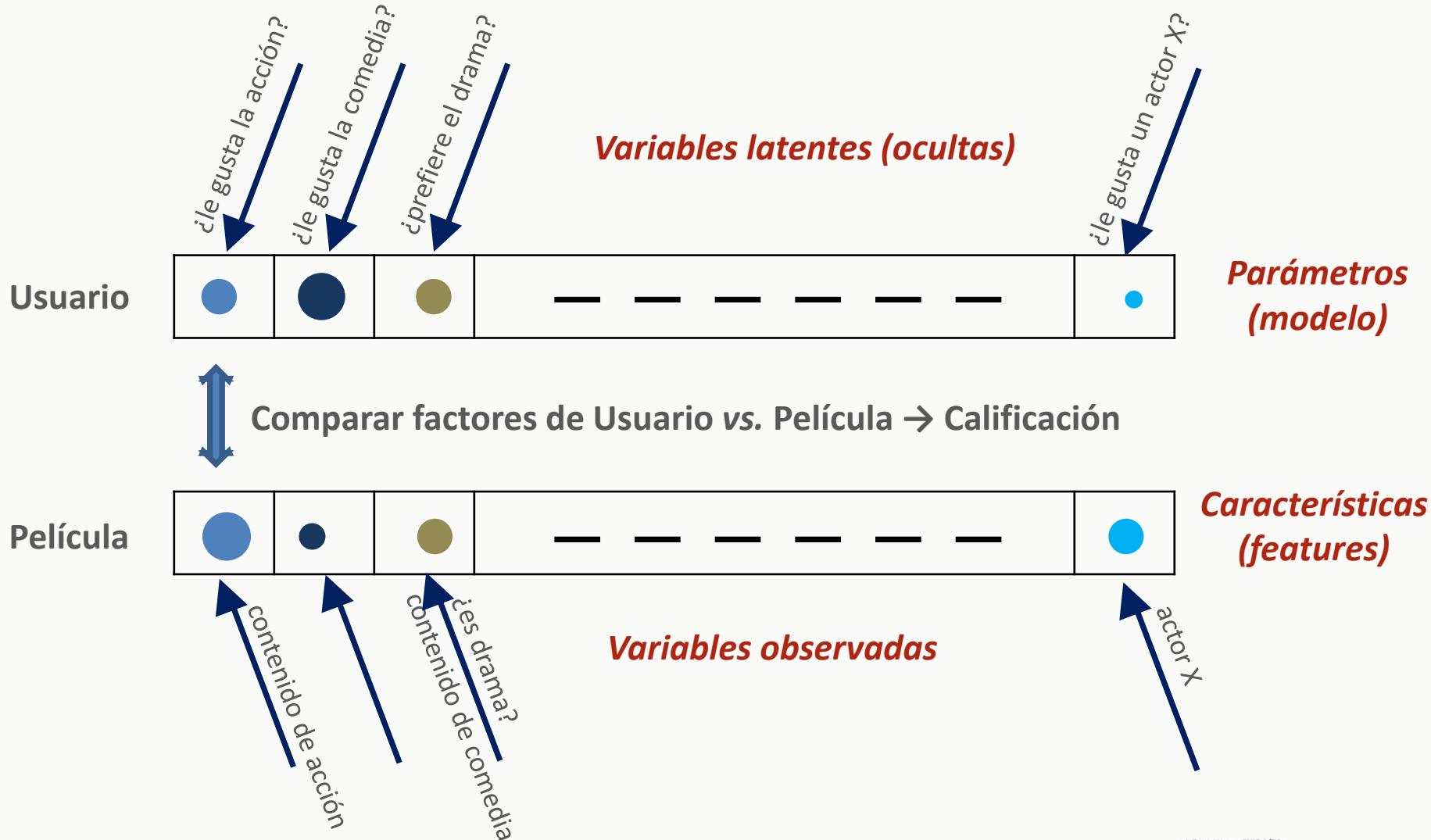
## ► Ejemplo:

*Predecir cómo un espectador calificará una película*
















## ► La esencia del ML:

- *Existe un patrón*
- *No se puede plantear matemáticamente una solución*
- *Tenemos datos sobre el problema*

# UN ENFOQUE DE SOLUCIÓN



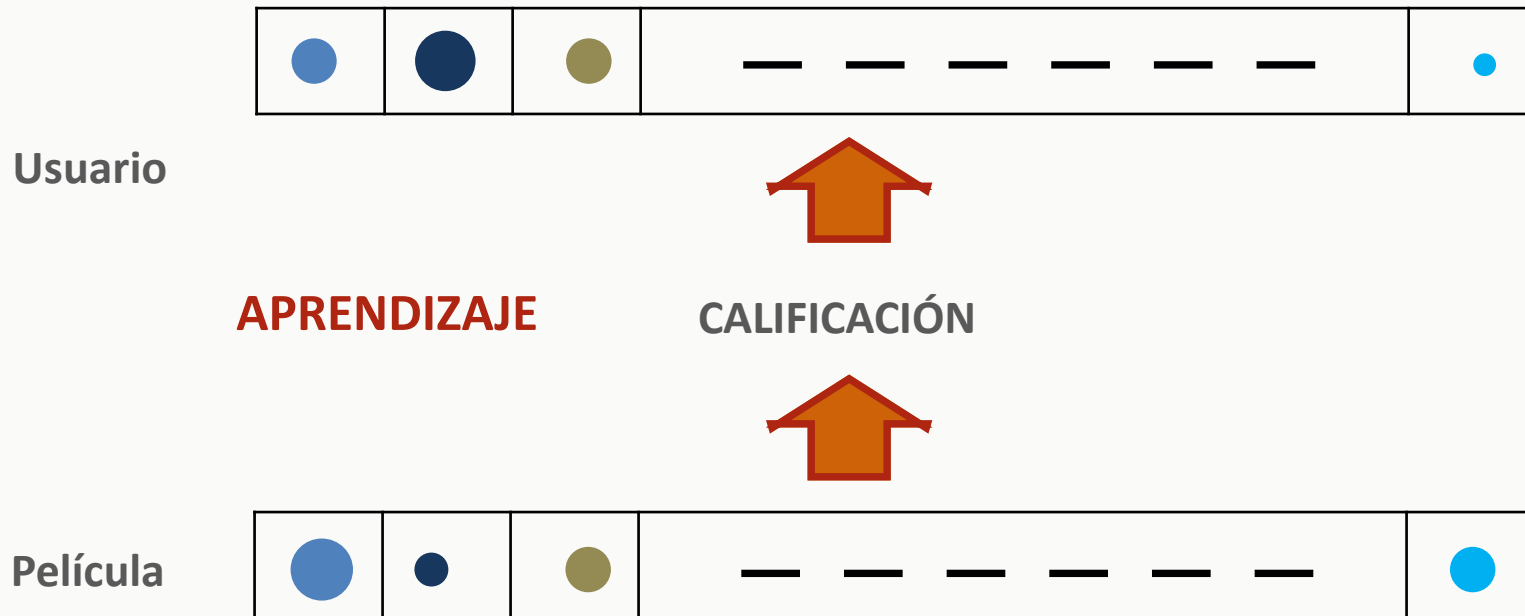
# EL ENFOQUE DEL APRENDIZAJE

Usuario				     	
Película				     	



CALIFICACIÓN

# EL ENFOQUE DEL APRENDIZAJE



# COMPONENTES DEL APRENDIZAJE

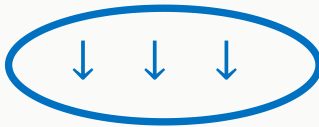
## Ejemplo: *Aprobación de Tarjeta de Crédito*

CONCEPTO	DATO	TIPO DE DATO
EDAD	25	Numérico (entero)
GÉNERO	M	Binario
SALARIO	B	Categorico
TIEMPO DE RESIDENCIA	5.5	Numérico (real)
DEUDA	[1.00 – 19,999.99]	Numérico (rango) ≈ Categorico

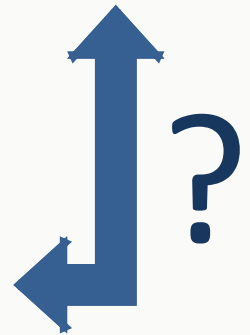
Otros tipos de datos pueden aparecer: tablas, arreglos, imágenes, texto.

# COMPONENTES DEL APRENDIZAJE

- ENTRADA  $x$  (información del solicitante)
- SALIDA  $y$  (cliente bueno o malo: +1 o -1) **DESCONOCIDA**
- Función OBJETIVO:  $f: \mathcal{X} \rightarrow \mathcal{Y}$  (fórmula ideal de aprobación de crédito)
- DATOS:  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  (registros históricos)

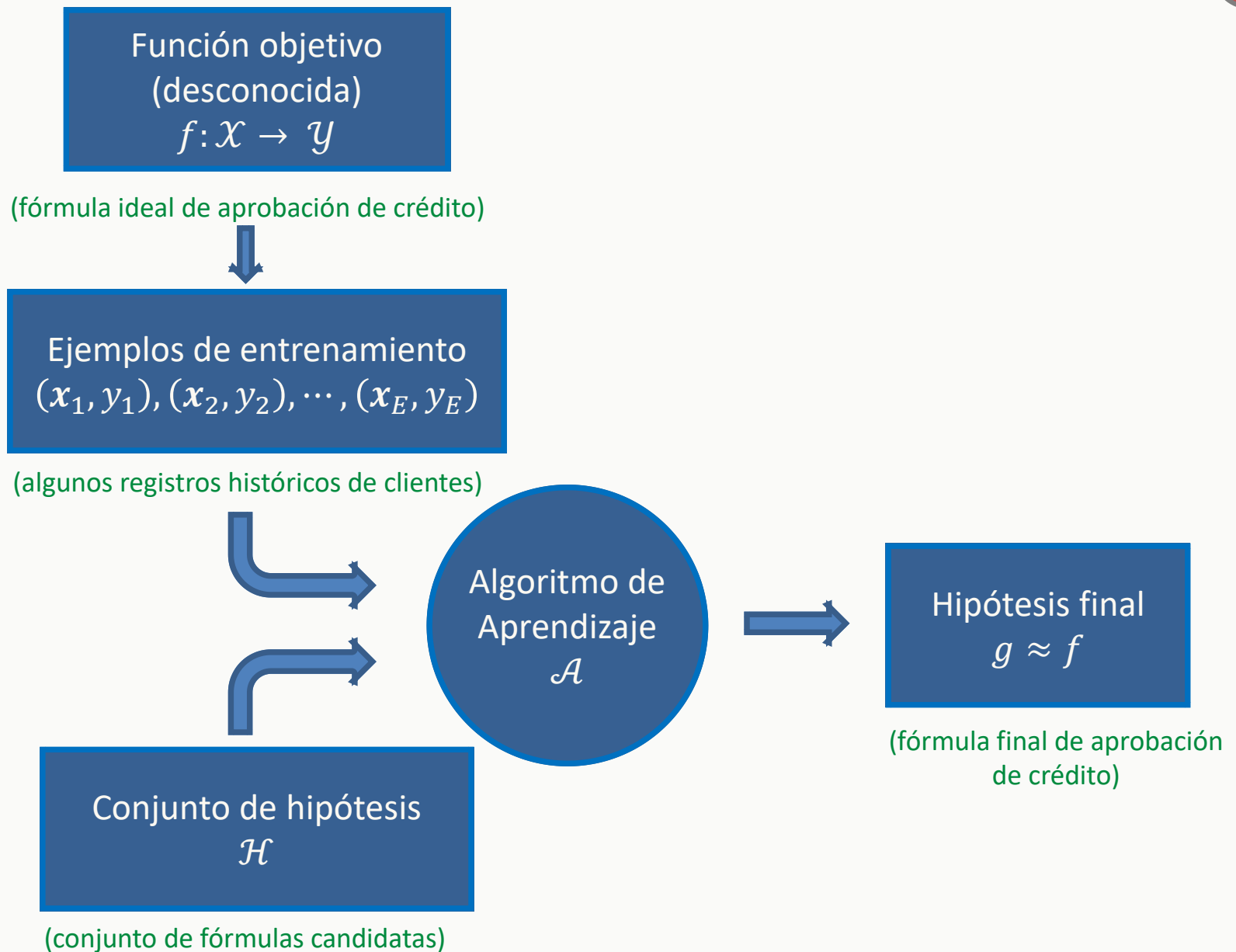


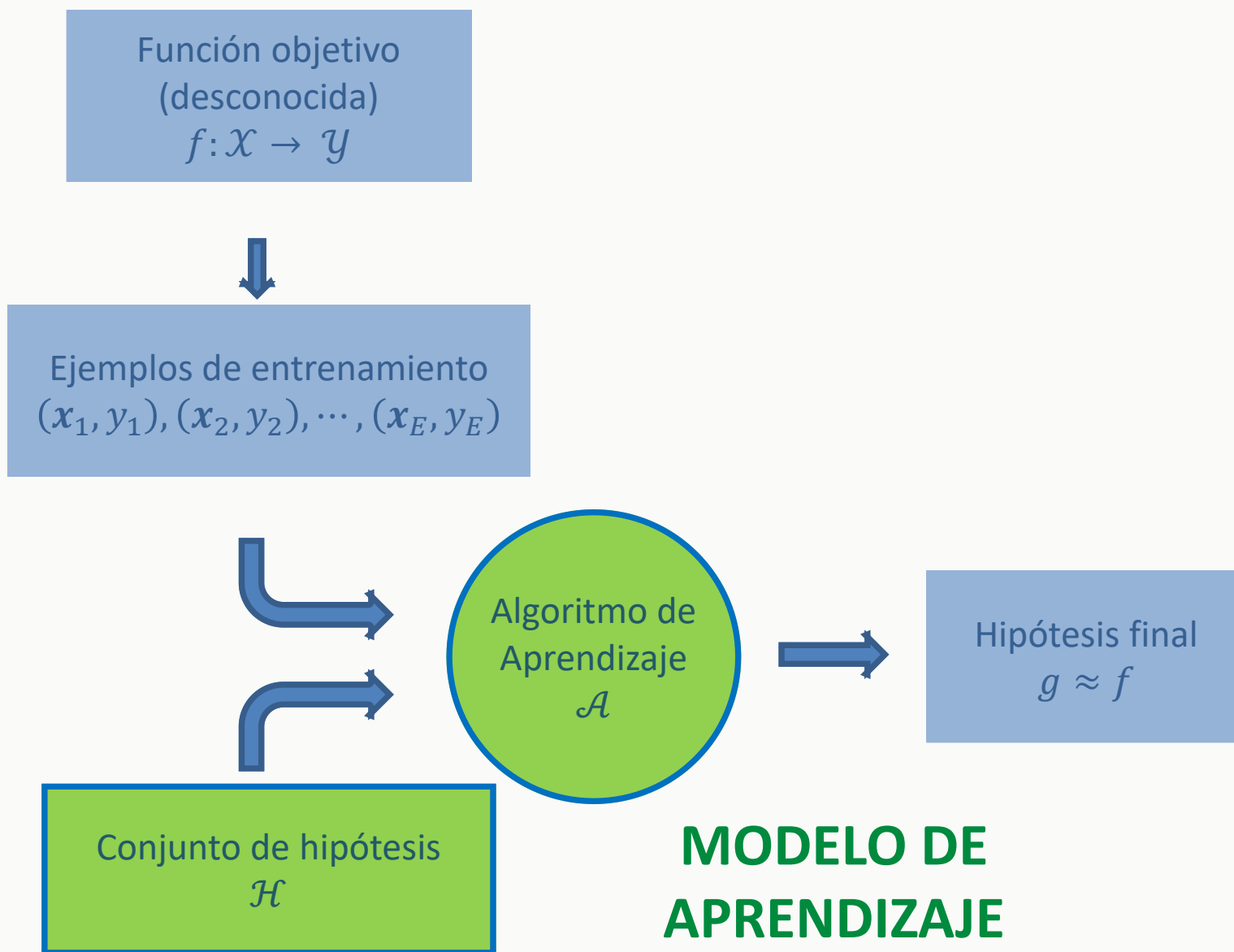
- HIPÓTESIS:  $g: \mathcal{X} \rightarrow \mathcal{Y}$  (fórmula final de aprobación de crédito)

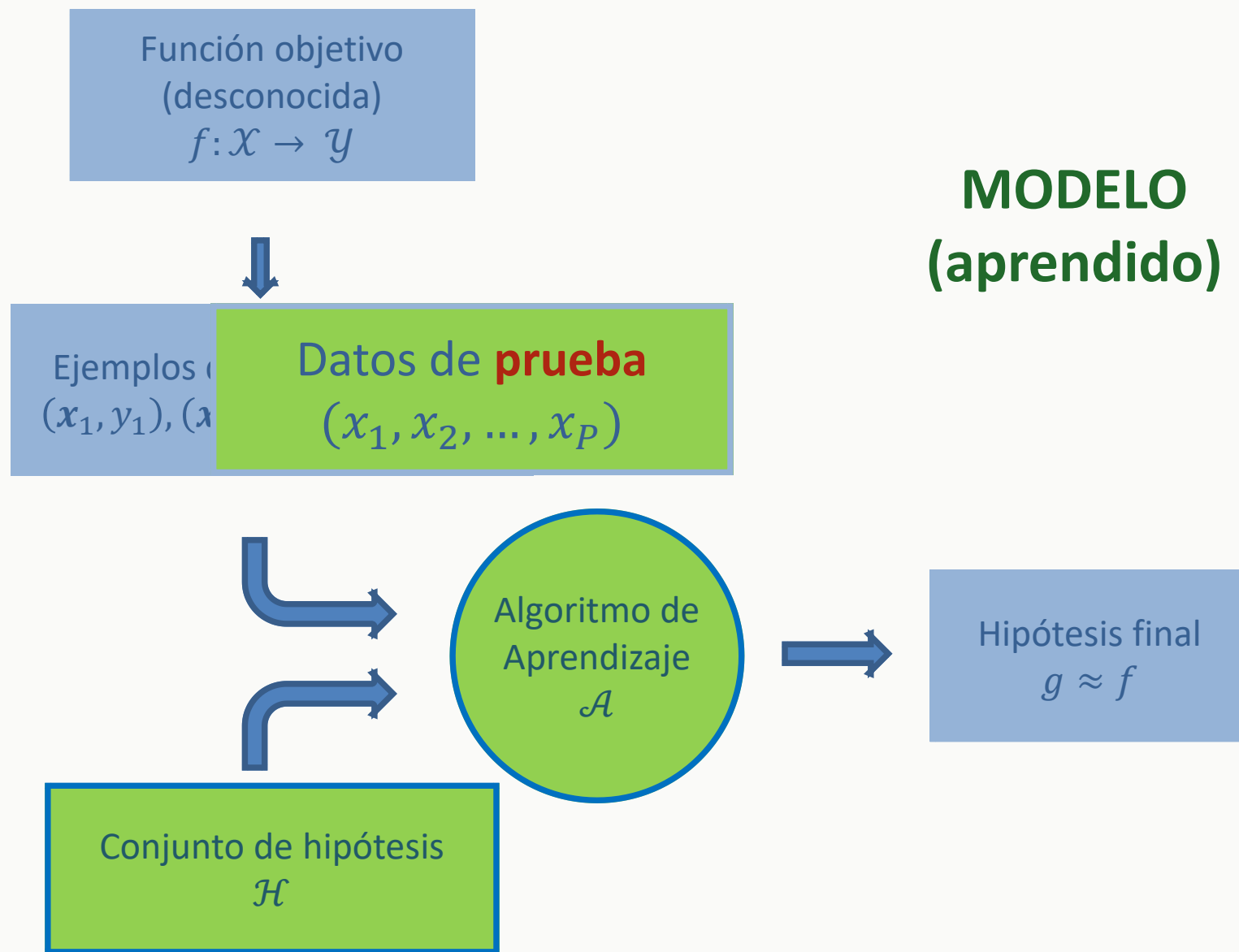


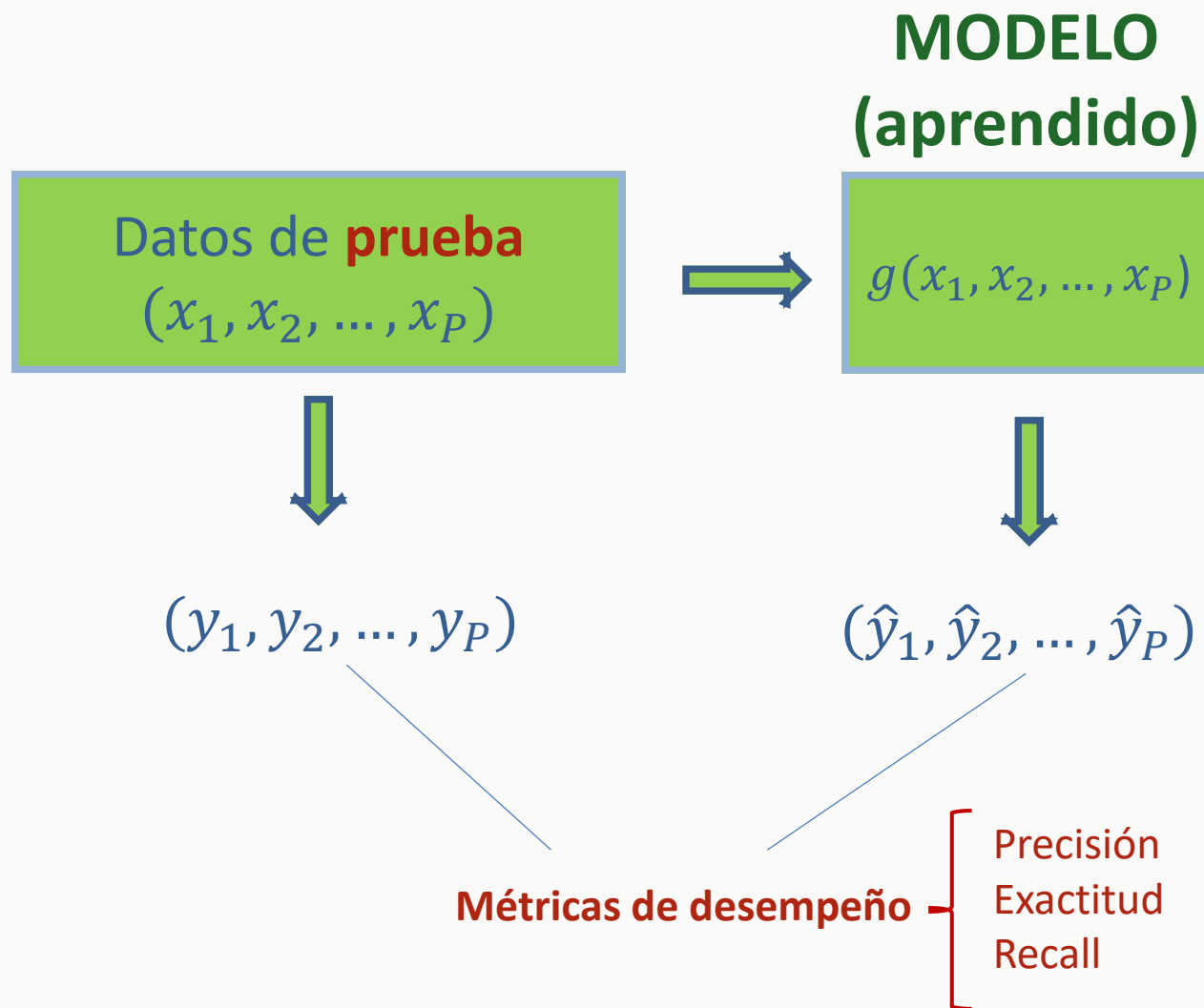
Teoría de la Probabilidad + Teoría de la Decisión + Teoría de la Información











# PREMISA BÁSICA DEL ML

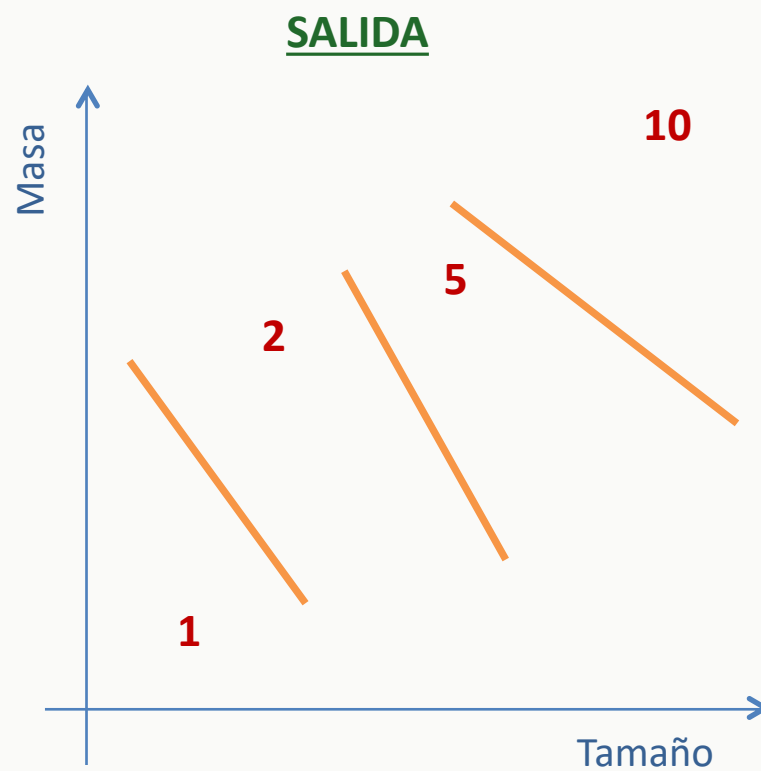
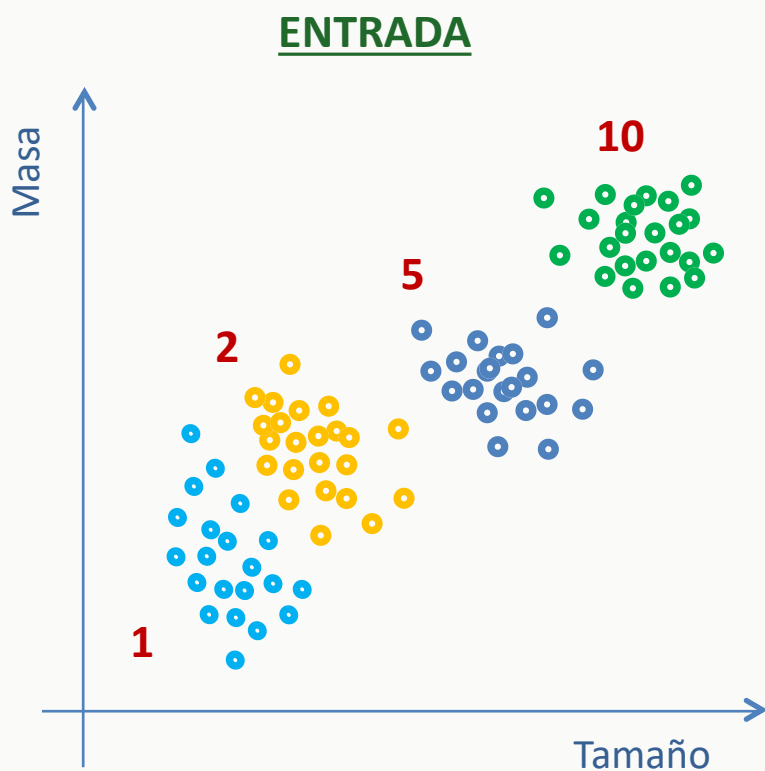
*“Utilizar un conjunto de observaciones para descubrir un proceso subyacente”*

TIPO DE APRENDIZAJE	TAREAS	ENFOQUES COMUNES
Supervisado	Regresión	Lineal/Polinomial/Bayesiana/Redes Neuronales
	Clasificación	Lineal/Basada en Kernels/Bayesiana/Redes Neuronales
No Supervisado	Agrupamiento (clustering)	Particiones/Jerárquico/Probabilista/Espectral/Gravitacional/Topológico*
Por Refuerzo	Diseño de agentes/MDP's	Programación dinámica/Recocido simulado/Algoritmos genéticos

\* <https://www.hindawi.com/journals/mpe/2019/4540731/>

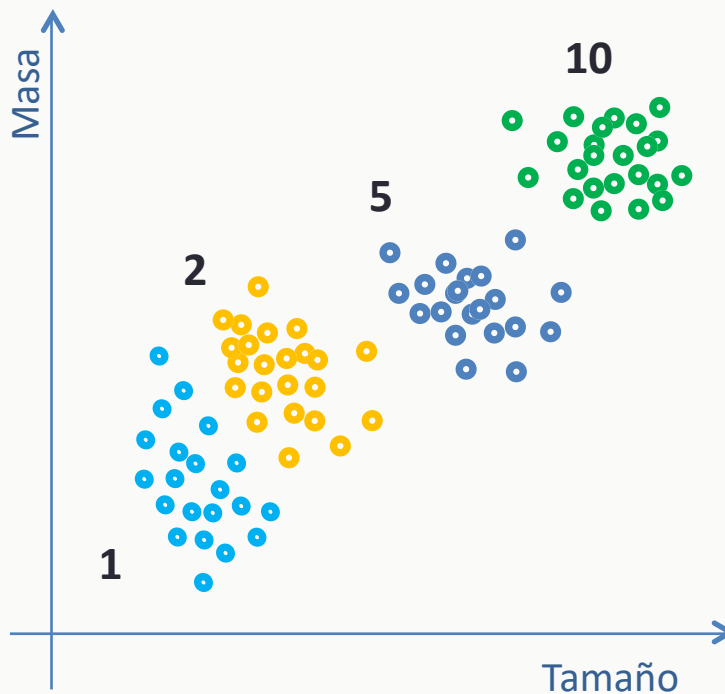
# APRENDIZAJE SUPERVISADO

Ejemplo de una máquina de expendio: **reconocimiento de monedas**



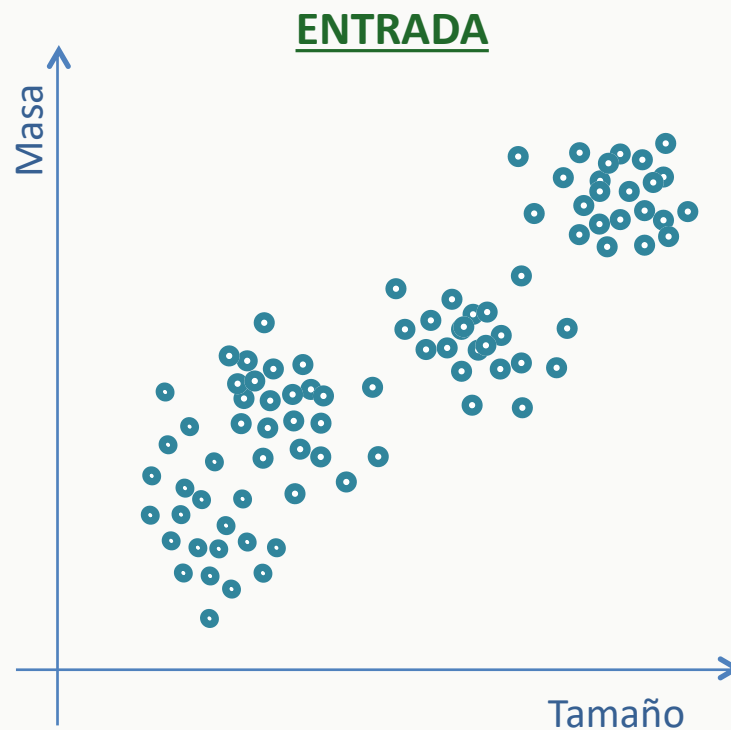
# APRENDIZAJE NO SUPERVISADO

En lugar de (**entrada, salida correcta**),



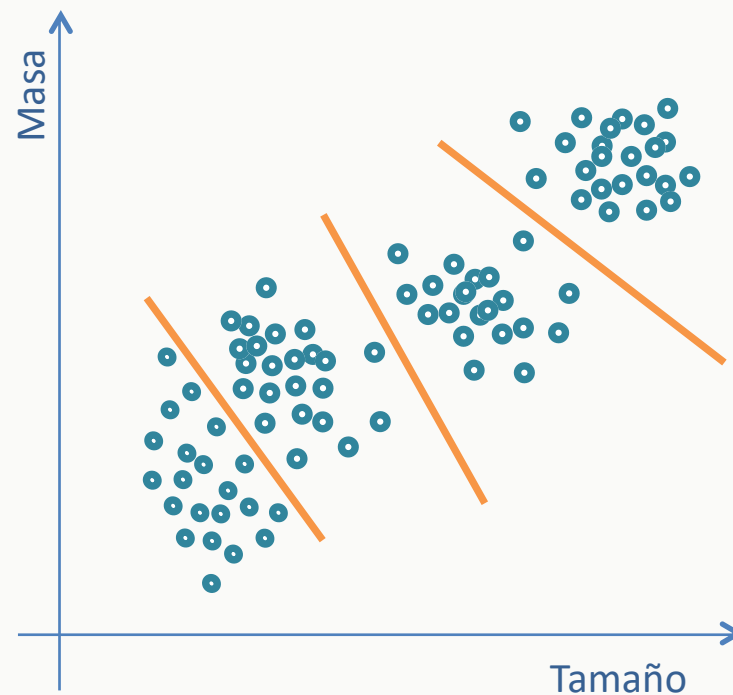
# APRENDIZAJE NO SUPERVISADO

En lugar de (**entrada, salida correcta**), tenemos (**entrada, ?**)

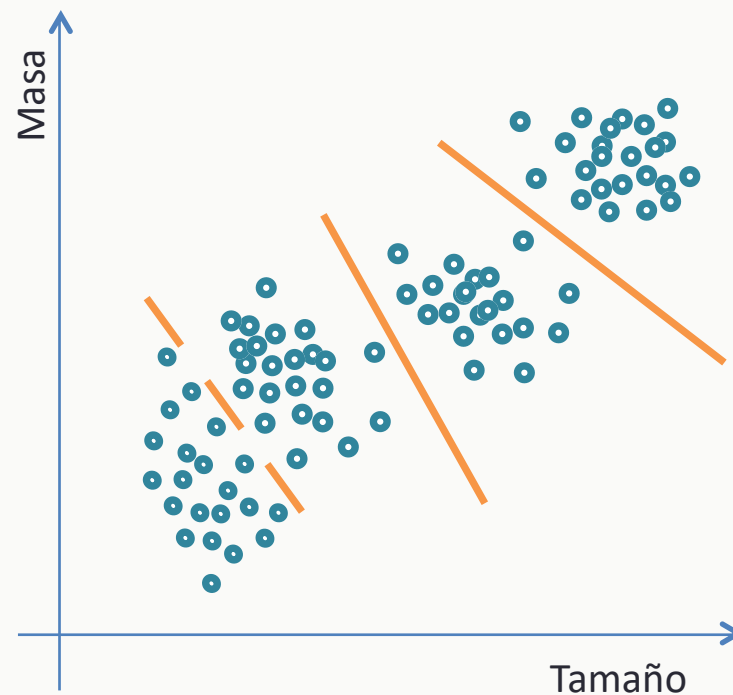




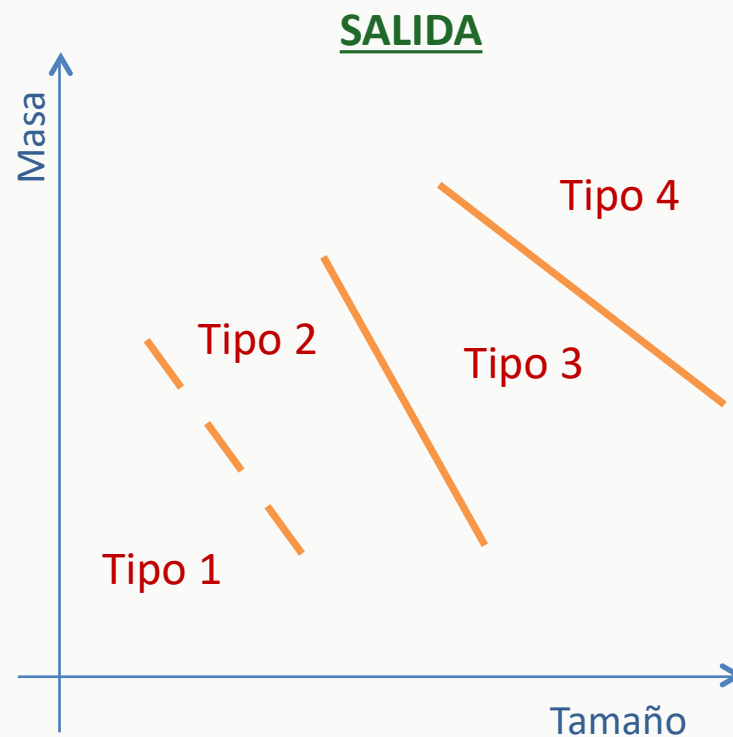
# APRENDIZAJE NO SUPERVISADO



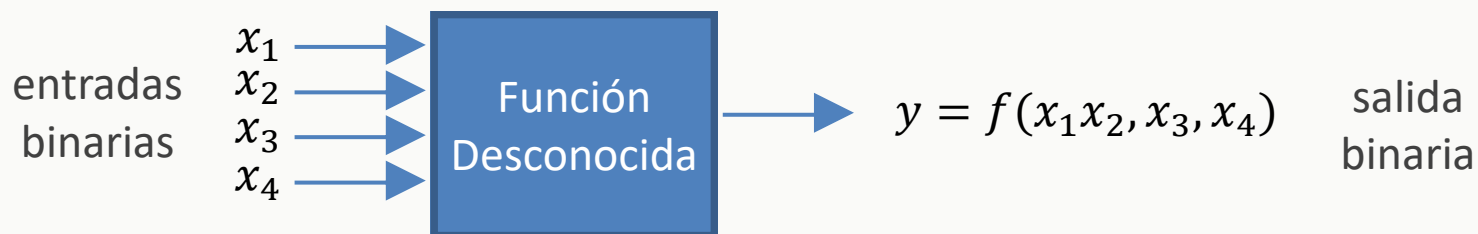
# APRENDIZAJE NO SUPERVISADO



# APRENDIZAJE NO SUPERVISADO



# LA MALDICIÓN DE LA DIMENSIONALIDAD



Ejemplo	$x_1$	$x_2$	$x_3$	$x_4$	$y$
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

¿Cuántas hipótesis podemos formular?

i.e.

¿De qué tamaño es el espacio de hipótesis?

i.e.

¿Cuántas funciones posibles podemos tener?

# APRENDER PARECE IMPOSIBLE

$2^{16} = 65536$  posibilidades.

Necesitaríamos conocer todas las instancias para poder determinar la función con certeza.

Después de 7 ejemplos, todavía tenemos  $2^9$  posibilidades

Ej	$x_1$	$x_2$	$x_3$	$x_4$	$y$
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0
8	x	x	x	x	1/0 ?



*Razonar con incertidumbre, elegir el enfoque de acuerdo a la complejidad del problema y la cantidad de datos disponibles.*

# DOS ESTRATEGIAS PARA ML

- ▶ Desarrollar lenguajes para expresar **conocimiento previo (a priori)**
  - *Gramáticas de reglas, modelos estocásticos, redes bayesianas.*
- ▶ Desarrollar espacios de hipótesis flexibles (**aproximación por error**)
  - *Colecciones anidadas de hipótesis: árboles de decisión, redes neuronales, SVMs*
- ▶ En cualquier caso debemos desarrollar algoritmos para encontrar una hipótesis que se ajuste a los datos.

# PREGUNTAS CLAVE EN ML

- ▶ ¿Cuáles son buenos espacios de hipótesis?
  - *¿Qué espacios que han sido útiles en aplicaciones prácticas?*
- ▶ ¿Qué algoritmos pueden trabajar con estos espacios?
  - *¿Existen principios generales de diseño para los algoritmos de aprendizaje?*
- ▶ ¿Cómo podemos optimizar la precisión en los puntos de datos futuros?
  - *Problema de "sobreajuste" (overfitting)*
- ▶ ¿Cómo podemos tener confianza en los resultados? (pregunta estadística)
  - *¿Cuántos datos de entrenamiento se requieren para encontrar hipótesis precisas?*
- ▶ ¿Son algunos problemas de aprendizaje computacionalmente intratables? (pregunta computacional)
- ▶ ¿Cómo podemos formular problemas del mundo como problemas de aprendizaje automático? (pregunta de ingeniería)

# RESUMEN

- ▶ **MACHINE LEARNING** es la combinación de la Probabilidad Bayesiana, Teoría de la Decisión y Teoría de la Información para realizar tareas de regresión o clasificación de datos, mediante un modelo de aprendizaje que toma datos históricos de entrada-salida y produce una hipótesis respecto de la función que relaciona estos datos.
- ▶ Los datos se representan como variables observadas y los parámetros del modelo como variables ocultas. Los parámetros se aprenden para producir una hipótesis final cuyo rendimiento o desempeño se evalúa en términos de métricas específicas como la precisión o la sensibilidad.
- ▶ El Aprendizaje Supervisado requiere de datos “etiquetados”, lo cual puede requerir la ayuda de humanos calificados. El Aprendizaje No Supervisado requiere de una selección cuidadosa del modelo de aprendizaje en función del conocimiento a priori que se tiene sobre los datos.
- ▶ El problema de la dimensionalidad nos lleva a elegir modelos de aprendizaje adaptados al problema y con una complejidad flexible.



# REFERENCIAS RELEVANTES

- ▶ **Bishop, Christopher M. *Pattern Recognition and Machine Learning*.** New York :Springer, 2006.  
<http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf>
- ▶ **Flach, Peter A. *Machine Learning : the Art and Science of Algorithms That Make Sense of Data*.** Cambridge ; New York :Cambridge University Press, 2012.  
<https://github.com/jprudhvi47/ML-Books/blob/master/Peter%20Flach-Machine%20Learning%20The%20Art%20and%20Science%20of%20Algorithms%20that%20Make%20Sense%20of%20Data.pdf>