



Inteligencia Artificial para profesionales TIC

Principios de ML – Deducción vs Inducción



01. Deducción vs. Inducción

02. Introducción al aprendizaje automático (ML)

03. Diferentes enfoques de ML



1. Deducción vs Inducción

Un cambio de paradigma

La **IA simbólica** se ha centrado en un **enfoque deductivo**.

Obtener el conocimiento de un experto, formalizar este conocimiento de alguna manera (por ejemplo, una colección de reglas), introducir un motor de inferencia (generalmente basado en algún tipo de lógica), introducir nuevos hechos y dejar que el sistema deduzca nuevos hechos.

La búsqueda en espacios enormes (exponenciales) es frecuente, por lo que la **heurística** es necesaria

¿Es posible ir más allá del conocimiento experto?

El aprendizaje automático como **enfoque inductivo**

Bien, cambiemos la estrategia

El problema es complejo, pero tenemos muchos datos (ejemplos), entonces, **¿podemos aprender de ellos?**

1. Deducción vs Inducción

Un cambio de paradigma

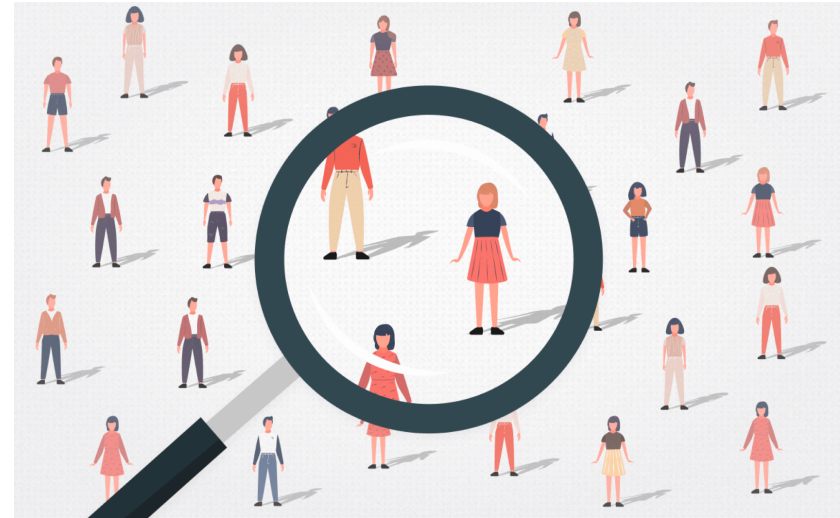


INDUCTIVE

specific examples \longrightarrow general rule

DEDUCTIVE

general rule \longrightarrow specific examples



2. Introducción al aprendizaje automático



Un cambio de paradigma

El aprendizaje automático es el campo de estudio que le da a las computadoras la capacidad de aprender sin ser programado explícitamente.

—Arthur Samuel, 1959

$$y = f(x)$$

Estamos tratando de estimar $f()$ a partir de los datos



2. Introducción al aprendizaje automático



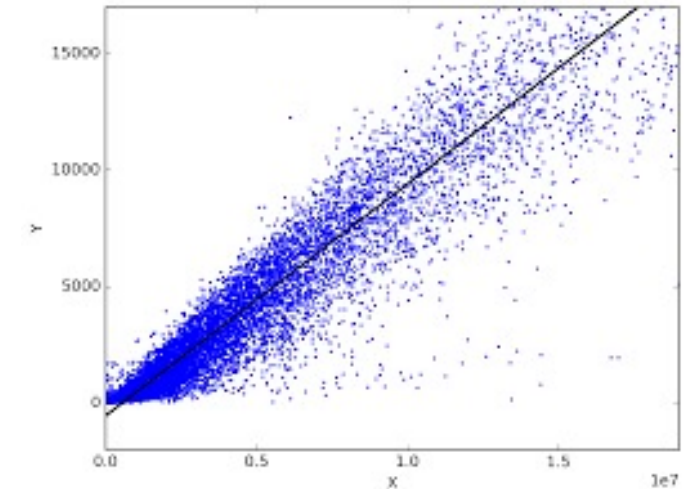
Una técnica simple de ML: la regresión lineal

La regresión lineal puede proporcionarnos buenos resultados frente a relaciones simples (lineales) entre entradas y salidas:

$$y = f(x)$$

Estamos tratando de estimar **un f óptimo** a partir de los datos:

- La regresión lineal simple es el modelo de aprendizaje automático más simple
- A partir de una colección de ejemplos (datos) se estima una relación lineal entre las variables independientes de entrada y la salida
- Cuando tenemos este modelo de regresión lineal, podemos intentar estimar una salida para una entrada totalmente nueva



2. Introducción al aprendizaje automático



ML solo cuando sea necesario

Algo muy importante a tener en cuenta es que el ML no es para todo:

- Si tenemos una función conocida, incluso realmente compleja, o un método o algoritmo (incluso complejo) que nos puede dar el resultado de la entrada, no uses ML. **¿Por qué si ya tienes una solución?**
- Debes tener un **buen conjunto de datos**. Bueno en **calidad** y **cantidad**. Si no, no creas en los milagros.
- Debe haber un **patrón en los datos de entrada** para obtener la salida, incluso si este patrón es muy complejo, pero debe haber un patrón. Si no es así, lo siento, ninguna técnica de ML te ayudará.
- El aprendizaje automático también es una buena estrategia para algunos problemas de espacio de búsqueda grandes
- Las técnicas de ML también nos permiten desarrollar soluciones altamente adaptativas y dinámicas (simplemente trabajando con nuevos datos y/o añadiendo nuevas funcionalidades, etc.)
- Una de las cosas más increíbles que estamos viviendo en la IA es que, utilizando algunas técnicas de ML, estamos **descubriendo patrones impredecibles en los datos de entrada y correlaciones antes no conocidas**, abriendo un mundo de posibilidades

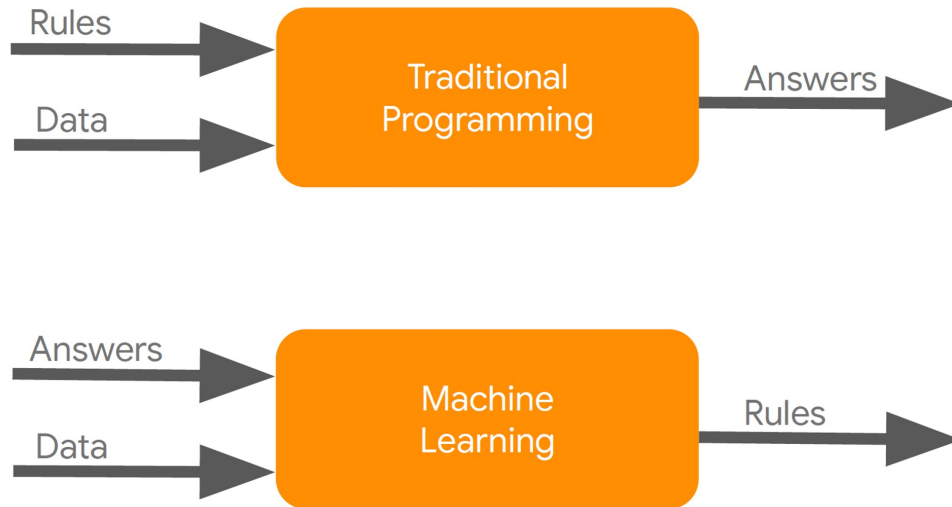
2. Introducción al aprendizaje automático



El ML como enfoque inductivo

Fácil. La inducción significa pasar de los ejemplos al conocimiento, utilizando técnicas capaces de encontrar los patrones y las reglas de las entradas a los resultados.

De la generación de respuestas a la generación de reglas (modelo):

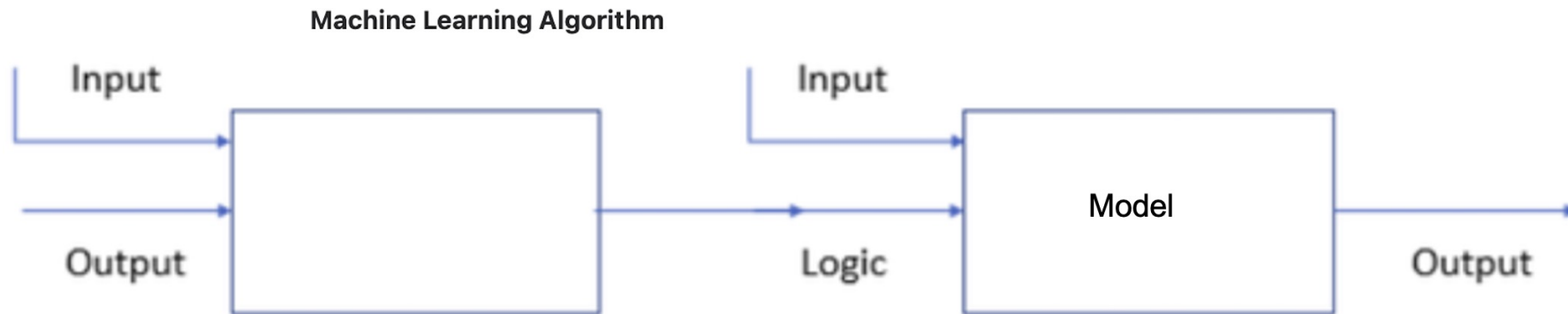


2. Introducción al aprendizaje automático



El ML como enfoque inductivo

Y cuando se crea el modelo, es hora de hacer predicciones.



Por Machine Learning (ML) entendemos una familia de algoritmos y técnicas que tratan de **aprender automáticamente a partir de los datos proporcionados**

Su objetivo es **generalizar comportamientos** a partir de los datos proporcionados como ejemplo

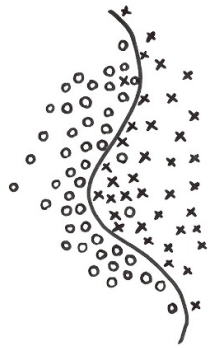
Es un proceso de inducción de conocimiento

2. Introducción al aprendizaje automático

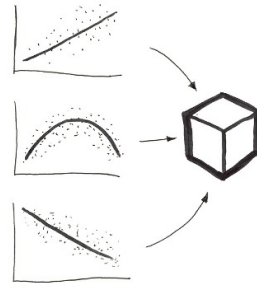


El ML como enfoque inductivo

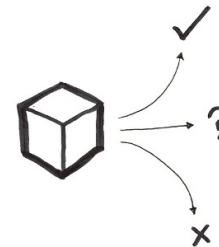
Los patrones en los datos son necesarios, ¡no hay magia!



Patterns
from Data



Models
from Patterns



Predictions
from Models



3. Diferentes enfoques de ML

Datos, pero ¿qué datos? ¿No hay datos pero sí experiencias?

El aprendizaje automático se puede dividir básicamente en **aprendizaje supervisado**, **no supervisado** y **de refuerzo** (y otras posibilidades como la detección de valores atípicos/anomalías y algunas en definición constante, como el aprendizaje semisupervisado, contrastivo, autosupervisado, etc.)

En el **aprendizaje supervisado** somos capaces de aportar datos a modo de ejemplos, donde conocemos la clase o valor de cada ejemplo que introducimos. Puede ser un problema de clasificación (clases o categorías) o un problema de regresión (un valor cuantitativo o numérico)

En el **aprendizaje no supervisado**, los datos no tienen etiquetas, por lo que tenemos que descubrir patrones ocultos, clases, anomalías, etc.

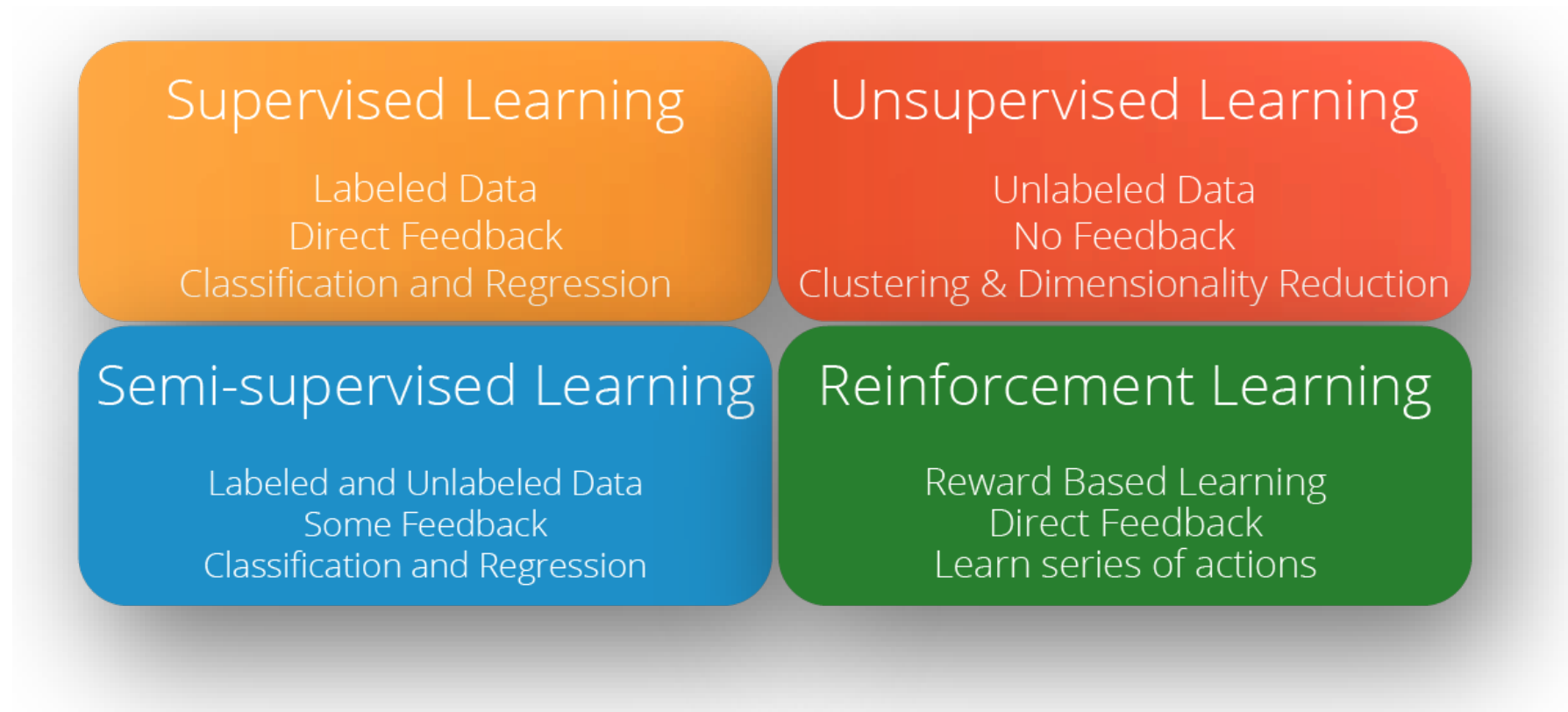
En el **aprendizaje por refuerzo**, el algoritmo prueba muchas cosas diferentes y es recompensado o penalizado dependiendo de si sus comportamientos le ayudan o dificultan alcanzar su objetivo

Otras técnicas, como la detección de valores atípicos, pueden compartir características o pertenecer a técnicas supervisadas y no supervisadas



3. Diferentes enfoques de ML

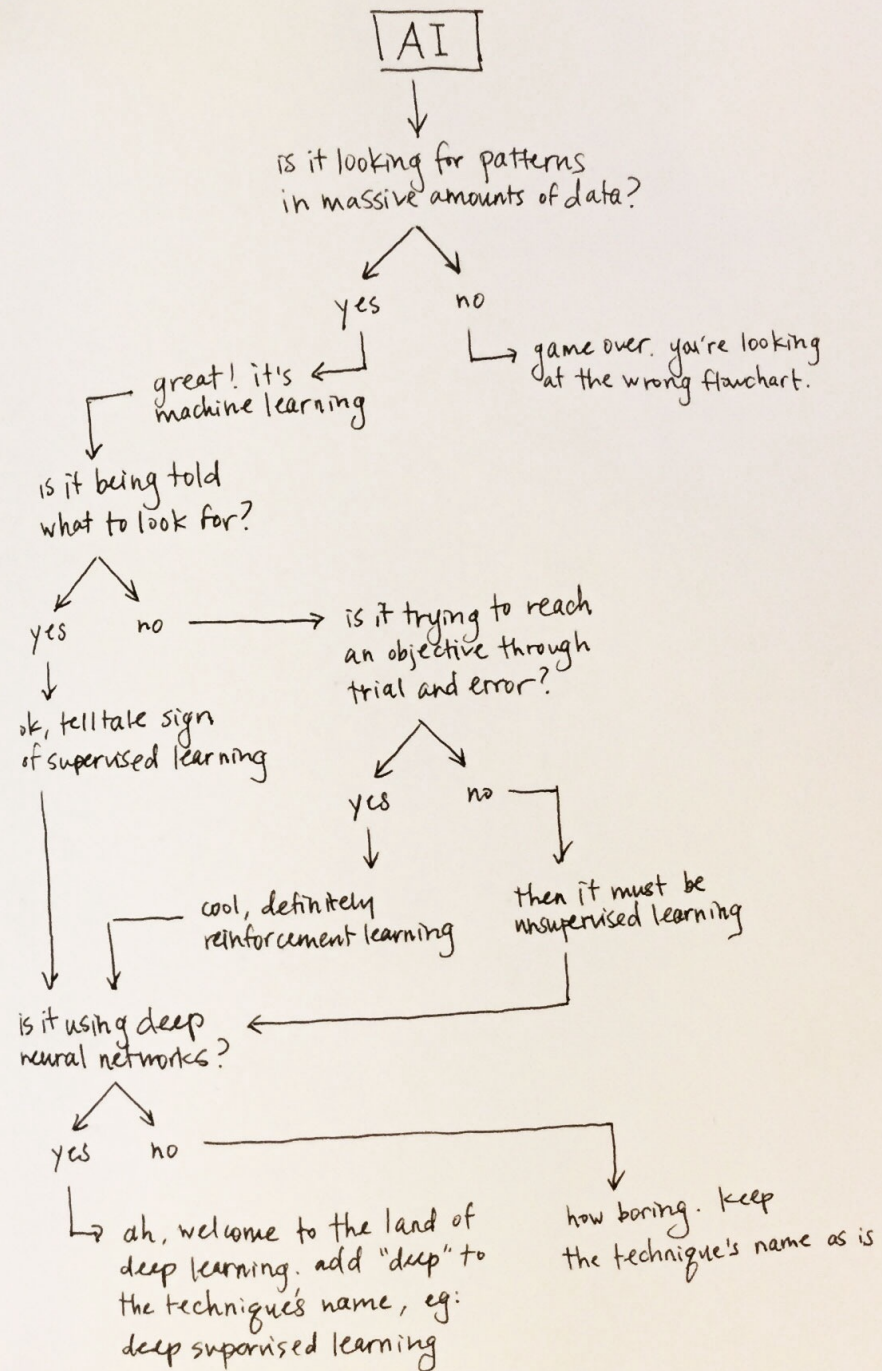
Datos, pero ¿qué datos? ¿No hay datos sino experiencias?



"What kind of machine learning is this?"

The Algorithm,
MIT Technology Review

By: Karen Hao





3. Diferentes enfoques de ML

Aprendizaje supervisado

Es el que nos puede dar **mejores resultados** (siempre dependemos de la calidad y cantidad de datos)

Sin embargo, **requiere ejemplos completos**, donde conocemos el valor de salida correcto para las entradas de cada ejemplo

Si el valor de **salida** conocido de los ejemplos es **numérico**, se trata de un problema de **regresión**

Si el resultado es una **categoría/etiqueta**, entonces es un problema de **clasificación**

En el **aprendizaje supervisado** intentamos crear un modelo capaz de recibir un **vector x de variables de entrada (llamadas variables independientes)** y generar **y**

$x \rightarrow \text{Modelo} \rightarrow y$

Si **y es una clase o etiqueta**, estamos modelando un problema de **clasificación**. Si **y es un valor numérico**, tenemos un problema de **regresión**.

El objetivo del aprendizaje supervisado es **predecir la salida de valores de entrada futuros**. En otras palabras, **generalizar los resultados frente a los datos de entrada no observados**, utilizando el conocimiento adquirido por los ejemplos anteriores

La clasificación y la regresión son posibles en muchas de las técnicas de ML, sin embargo, las métricas que usamos para medir qué tan bueno es el modelo son diferentes



3. Diferentes enfoques de ML

Aprendizaje supervisado

	Feature 1	Feature 2	Feature 3	Feature 4	Label
0	X11	X12	X13	X14	y1
1	X21	X22	X23	X24	y2
2	X31	X32	X33	X34	y3

Entradas = Características = Atributos = Variables independientes

Salidas = etiquetas = variable dependiente

$$y = f(x)$$

Número de entidades = Dimensión del conjunto de datos



3. Diferentes enfoques de ML

Clasificación vs regresión

- En la **clasificación**, el aprendizaje supervisado utiliza ejemplos conocidos para inferir una etiqueta (clasificar) de los vectores de entrada x eligiendo entre varias categorías o clases.
- Utilizamos ejemplos previamente etiquetados para "aprender" patrones con el fin de llevar a cabo la clasificación
- En la **regresión**, en lugar de una etiqueta o categoría por ejemplo, tenemos un valor numérico, y el objetivo es de nuevo predecir este valor a partir de un nuevo elemento de entrada o vector de características
- **Ejemplos:** A partir de un correo electrónico de entrada para averiguar si se trata de spam o no (**clasificación** binaria), diagnóstico médico (**clasificación** multiclase) a partir de un vector de características (sexo, presión arterial, colesterol, etc.), estimar el precio de una casa a partir de algunos datos de entrada (**regresión**)

