

L U M B R E



Qué es aprender de datos

Grado: 5

Materia: Fundamentos de Machine Learning · Machine Learning

11/10/2026

Antes de empezar

En lugar de escribir a mano las reglas, el aprendizaje automático las deduce de ejemplos. Ese cambio de paradigma —de programar a aprender— abre problemas que nadie sabría codificar de forma explícita, del reconocimiento de imagen al traductor automático.

QUÉ APRENDERÁS

- Definir el aprendizaje automático como la inducción de reglas a partir de ejemplos, y distinguirlo de la programación clásica y de sus tres grandes familias.

Qué es aprender de datos

De la regla al dato, y del dato a la regla

Programar es traducir una regla que ya entiendes a código. Aprender de datos es lo inverso: partes de ejemplos cuya regla desconoces y de ellos extraes la regla. Invertir la flecha cambia por completo cómo resolvemos problemas.

VIDEO

Machine learning explicado de forma sencilla

Vídeo en español que introduce qué es el aprendizaje automático y en qué se diferencia de programar a mano.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=n-Df1QldMuc>

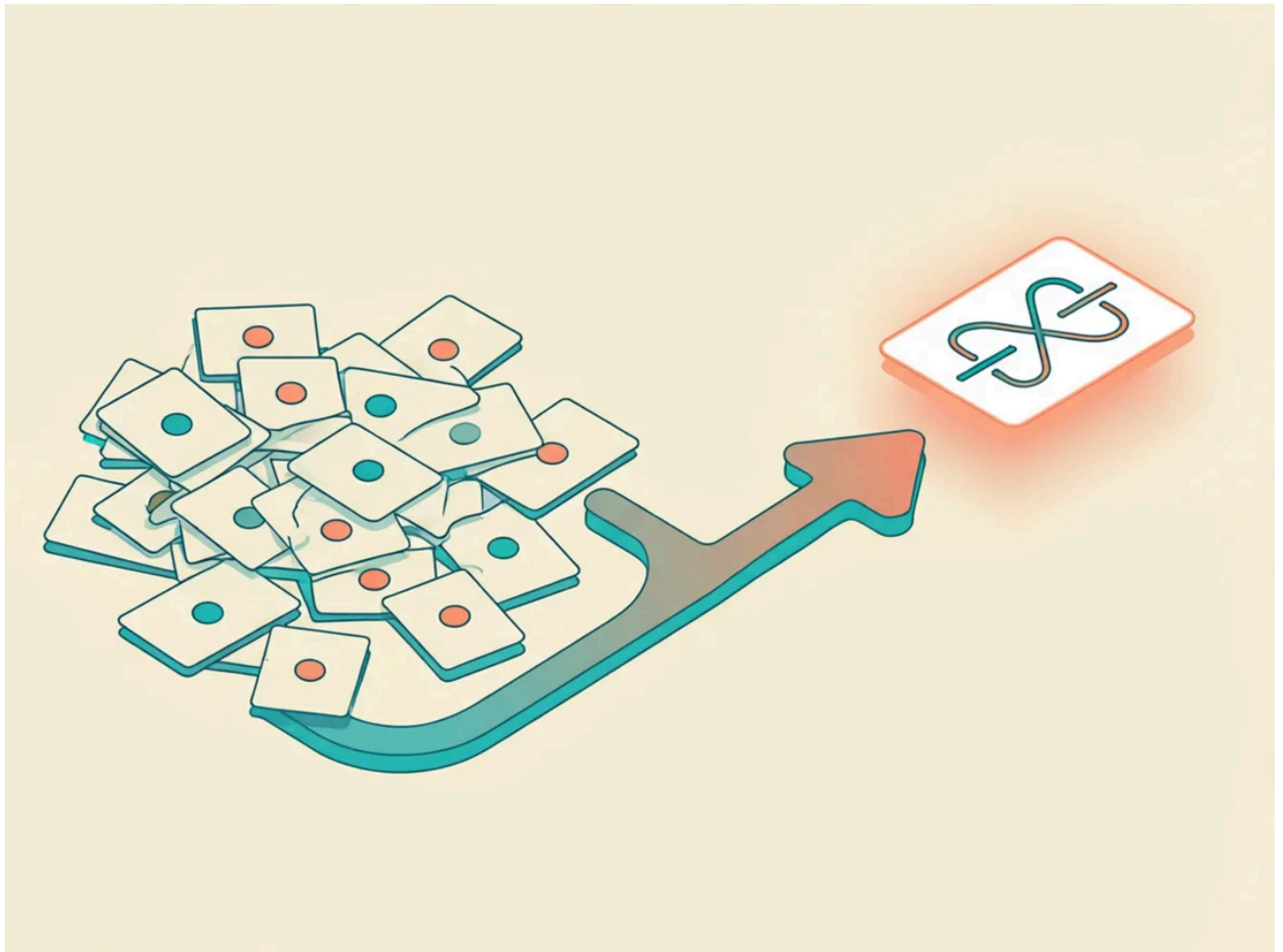
Programación clásica frente a aprendizaje

En la vía clásica damos reglas y datos al ordenador, y él produce respuestas. En aprendizaje automático damos respuestas y datos, y el sistema produce reglas: un

modelo. Cuando la regla es demasiado compleja o indefinible para un humano — reconocer una cara—, dejar que la descubra la máquina se vuelve la única salida.

La idea de generalizar

Aprender no es memorizar los ejemplos vistos, sino generalizar a ejemplos nuevos. Un modelo que reproduce de memoria el entrenamiento pero falla ante datos inéditos no ha aprendido nada útil. La generalización es el corazón del campo, y medirla, su obsesión.



El cambio de paradigma: en lugar de escribir la regla a mano, se ofrecen ejemplos y el sistema deduce la regla. La flecha va de los datos a la ley, no de la ley a los datos.

Inducir desde lo particular

El aprendizaje es razonamiento inductivo: de muchos casos concretos extraemos una regularidad general. Esa generalización nunca está garantizada del todo —es el problema de la inducción—, por eso exigimos validar con datos que el modelo no usó para entrenar.

Aprendizaje supervisado

Cuando cada ejemplo trae la respuesta correcta —una etiqueta o un número—, el sistema aprende a copiar esa correspondencia: correo basura o no, precio según características. Es la familia más extendida, y su materia prima son las etiquetas, caras de conseguir.

Aprendizaje no supervisado

Sin respuestas, el sistema busca estructura por sí solo: agrupa clientes similares, comprime dimensiones, detecta anomalías. No hay un «correcto» externo que guíe; el criterio lo pone la propia estructura de los datos y la intención del analista.

Aprendizaje por refuerzo

Aquí un agente actúa en un entorno y recibe recompensas o castigos según sus decisiones. Aprende una estrategia que maximice la recompensa acumulada, probando y ajustando. Es el paradigma del juego, la robótica y la conducción autónoma.

¿Cuándo conviene?

- Cuando la regla es imposible o carísima de escribir a mano.
- Cuando patrón varía con datos nuevos y hay que adaptarse.
- Cuando tenemos muchos ejemplos de lo que queremos reproducir.

El coste oculto

Aprender de datos exige datos: representativos, limpios y etiquetados. Un modelo no supera a la calidad de su experiencia. Basura dentro, basura fuera; y unos datos sesgados producen un modelo sesgado con aires de objetividad matemática.

Errores frecuentes

- Confundir memorizar los ejemplos con aprender a generalizar.
- Creer que el modelo es objetivo cuando heredó el sesgo de los datos.

Ejemplo resuelto: decidir si un correo es spam

Nadie sabe escribir todas las reglas del spam. Se plantea como aprendizaje supervisado:

1. Reunir ejemplos: miles de correos ya etiquetados como basura o legítimos por personas.
2. Elegir familia: un clasificador que asocie texto con una de las dos etiquetas.

3. Entrenar: el sistema induce qué señales predicen spam en lugar de que se las dictemos.
4. Validar: medir con correos no vistos para comprobar que generaliza y no memorizó.
5. Vigilar el sesgo: si los ejemplos solo cubren un idioma o un tipo, el modelo fallará fuera de él.

¿Para qué sirve en la realidad?

Todo problema cuya regla es borrosa o cambia con el tiempo —detección de fraude, traducción, recomendación— es candidato a aprenderse de datos. Reconocer cuándo el aprendizaje encaja, y cuándo una regla escrita a mano bastaría, es la primera decisión del oficio.

1. En el aprendizaje automático, a diferencia de la programación clásica, el sistema obtiene como principal resultado:

- ☐ Respuestas a partir de reglas escritas por un humano
- ☐ Reglas o un modelo a partir de ejemplos con respuestas
- ☐ Un conjunto de datos más limpio
- ☐ Una interfaz de usuario más rápida

2. El objetivo del aprendizaje automático es reproducir lo mejor posible los datos de entrenamiento.

- ☐ Verdadero
- ☐ Falso

Aprender es comprimir

Un buen modelo resume millones de ejemplos en unos pocos parámetros que capturan la regularidad. Esa compresión con pérdida de detalle irrelevante es lo que le permite predecir: guarda la señal, descarta el ruido. Un modelo que guardara todo sería solo una base de datos más.

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. Fundamentos



4. Ordena el flujo inverso del aprendizaje frente a la programación:

Elementos:

- humano escribe la regla
- máquina aplica la regla a datos
- damos datos y respuestas
- máquina induce la regla

Programar	. . .
Aprender	. . .

A I Thinker [https://www.abhishek-tiwari.com/Reflexiones accesibles sobre qué es y qué no es aprender de máquinas.](https://www.abhishek-tiwari.com/Reflexiones%20accesibles%20sobre%20qué%20es%20y%20qué%20no%20es%20aprender%20de%20máquinas.)

5. Une cada familia de aprendizaje con su materia prima.

5.	ejemplos con la respuesta correcta	-----
5.	datos sin etiquetas, buscar estructura	-----
5.	acciones premiadas o castigadas en un entorno	-----
5.	rendimiento en datos nunca vistos	-----

Opciones: supervisado · no supervisado · refuerzo · generalización

Fundamentos del ML

Aprender de datos

Paradigma

de ejemplos a regla

invertir la flecha

Familias

supervisado

no supervisado

refuerzo

Clave

generalizar

validar

Riesgos

memorizar

sesgo de los datos

El mapa mental interactivo es interactivo en la versión online.



¿Qué es el machine learning? <https://www.oracle.com/latam/artificial-intelligence/machine-learning/what-is-machine-learning/> Introducción de referencia en español al aprendizaje automático.

6. Un hospital quiere detectar neumonía en radiografías. ¿Es un problema supervisado, no supervisado o por refuerzo? ¿Qué datos necesitaría y qué podría sesgar al modelo?

Referencia curricular

Grado:	5
Materia:	Fundamentos de Machine Learning · Machine Learning

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-mlf-001-que-es-aprender-de-datos>

Material educativo generado desde Lumbre · lumbre.lat