



cursos d'estiu | cursos de verano
rafael
altamira

Introducción al *Machine Learning*

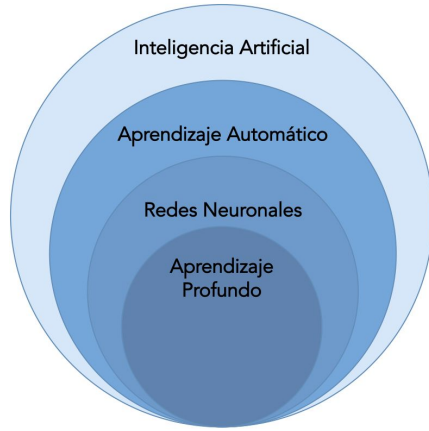
Conceptos básicos sobre el aprendizaje
automático

Profesor: Antonio Javier Gallego

Contenidos

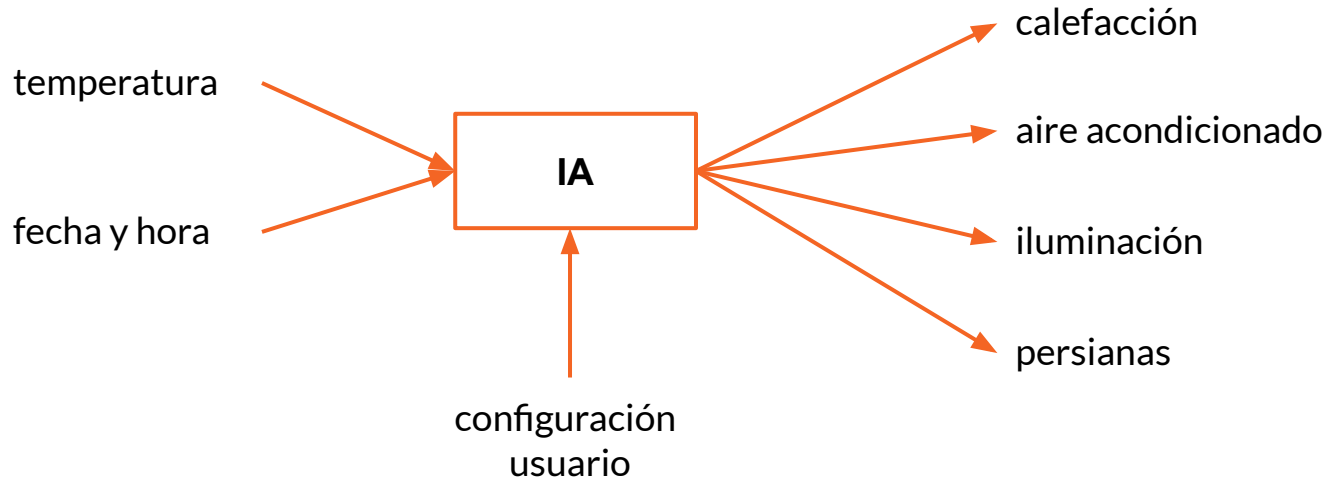
- Introducción
- Aprendizaje
- Datos
- Etapas
- Librerías

Introducción

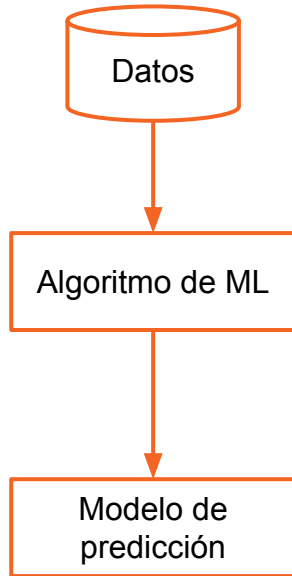


- **Inteligencia Artificial (IA):** Técnica que permita a un ordenador/máquina mostrar un comportamiento que imite la inteligencia humana.
- **Aprendizaje automático:** Subconjunto de técnicas de IA que utilizan métodos estadísticos para generar modelos de predicción a partir de datos.
- **Redes neuronales:** Subconjunto del ML basada en el desarrollo de modelos que imitan el funcionamiento del cerebro humano.
- **Aprendizaje profundo:** Subconjunto de las redes neuronales que emplean técnicas para extraer de manera automática las características descriptivas de la tarea a resolver.

Inteligencia Artificial



Machine Learning



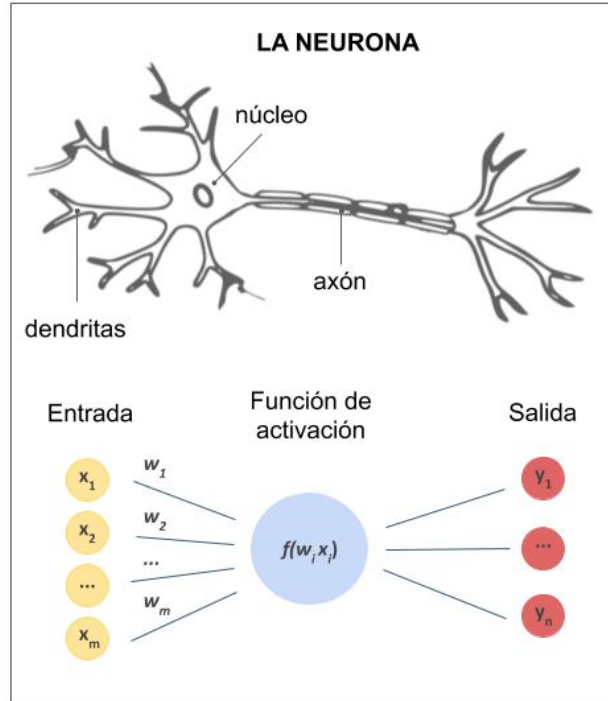
- Ciencia que permite que los ordenadores **aprendan** a resolver tareas a partir de ejemplos, es decir, sin instrucciones directas.
- Usa algoritmos para identificar **patrones en los datos** con los que generar un **modelo** (expresión matemática) que puede hacer predicciones sobre los mismos.
- Con más experiencia y datos, los resultados del aprendizaje automático son más precisos, de forma muy similar a cómo los humanos mejoran con más práctica o más ejemplos.
- El modelo se genera con los datos disponibles y luego se puede usar para hacer predicciones sobre datos nuevos, es decir, no utilizados durante el entrenamiento (ajuste o generación) del modelo.

IA vs. ML

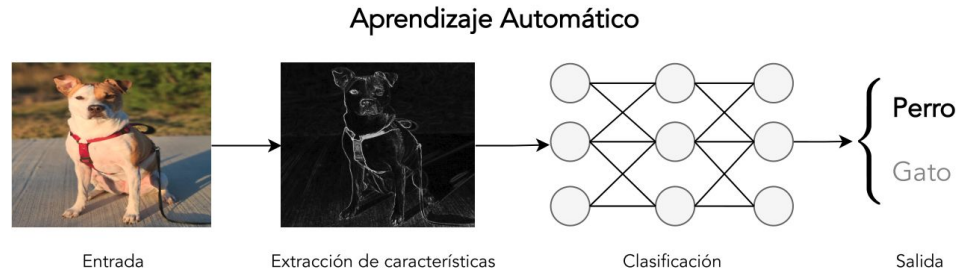


- **Ventajas:** La adaptabilidad del ML lo convierte en una excelente opción en escenarios en los que los **datos** siempre **cambian**, la naturaleza de la solicitud o la tarea siempre se transforma o la codificación de una solución sería realmente imposible.
- **Desventajas:** Necesita datos y un proceso de etiquetado.

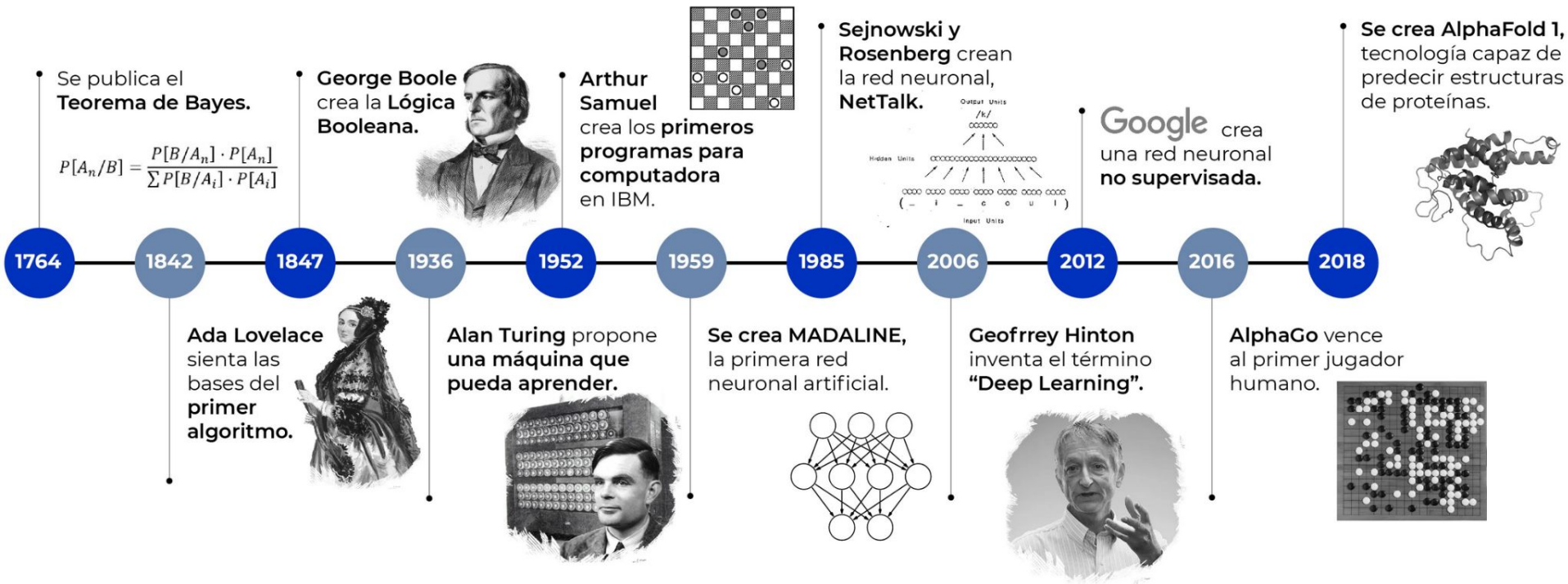
Redes neuronales y DL



ML vs. DL



Línea del tiempo del ML



Aplicaciones de la IA y el ML

Fabricación



- Mantenimiento predictivo
- Estimación de la reserva de garantía
- Tendencia a comprar
- Previsión de la demanda
- Optimización de procesos
- Telemática

Ventas



- Planificación predictiva de inventario
- Sistemas de recomendación
- Ventas y marketing multicanal
- Segmentación del mercado y las tendencias
- Identificación de perfiles de clientes

Salud



- Mantenimiento predictivo
- Estimación de la reserva de garantía
- Tendencia a comprar
- Previsión de la demanda
- Optimización de procesos
- Telemática

Viajes y alojamiento



- Planificación de vuelos
- Precios personalizados
- Medios de comunicación social - consumidor retroalimentación y análisis de la interacción
- Solución de quejas de clientes
- Gestión de tráfico

Servicios financieros



- Análisis de riesgo y regulación
- Segmentación de clientes
- Venta y marketing multicanal
- Gestión de campañas de venta y marketing
- Evaluación de riesgos de inversión

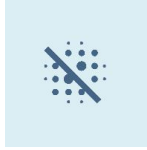
Energía



- Análisis de uso de energía
- Emisiones de carbono
- Precios específicos para clientes
- Gestión inteligente de la red eléctrica
- Optimización de la oferta y la demanda de energía

Aprendizaje

Tipos de aprendizaje



- **Aprendizaje supervisado:** El modelo se genera mediante datos etiquetados, lo cual significa que las respuestas o salidas correctas se proporcionan junto con los datos utilizados.
 - Ejemplo: detectar si un e-mail es spam o no a partir de un conjunto de datos que contenga e-mails convencionales y otros clasificados como spam.



- **Aprendizaje no supervisado:** El modelo se entrena con datos no etiquetados, lo cual significa que debe encontrar patrones y relaciones en los datos por sí mismo.
 - Ejemplo: agrupar documentos similares en función del texto que contienen o detectar segmentos concretos de usuarios que acceden a una web en función de sus comportamientos e interacciones.

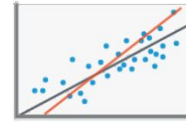


- **Aprendizaje por refuerzo:** El modelo se entrena mediante recompensas o castigos en función de sus acciones en un entorno y de los resultados obtenidos.
 - Ejemplo: aprender a jugar al ajedrez recompensando si gana y castigando si pierde. Es un tipo de aprendizaje muy común en robótica, donde el robot aprende un determinado comportamiento a través de su interacción con el entorno.

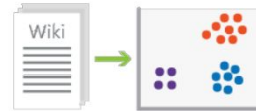
Tipos de aprendizaje - ejemplos



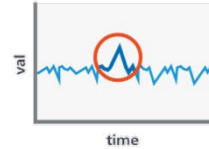
Classification
(supervised – predictive)



Regression
(supervised – predictive)



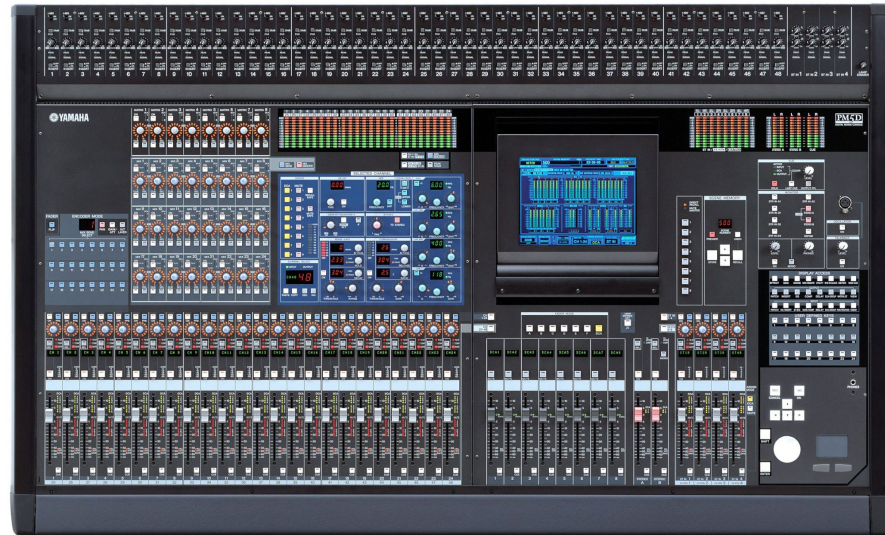
Clustering
(unsupervised – descriptive)



Anomaly Detection
(unsupervised – descriptive)

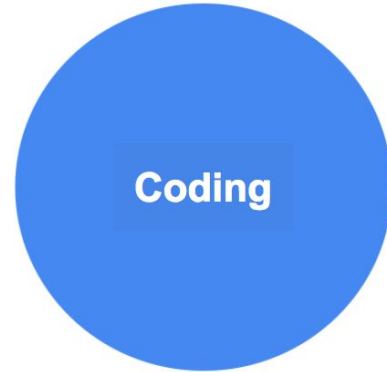
¿Cómo aprende?

- Concepto de “entrenamiento” y función objetivo.



Datos

Importancia de los datos



What people think AI is about

Tipos de datos



- Estructurados:
 - Datos tabulares con medidas y características.
 - Datos que tienen un modelo definido o provienen de un campo determinado en un registro.



- No estructurados:
 - Imágenes
 - Audio
 - Video
 - Otros formatos de archivo en los que no podemos definir una estructura



Datos estructurados

- Ejemplo: Iris dataset
- Base de datos con características de 3 tipos de plantas:
 - Setosa
 - Versicolor
 - Virginica
- Características:
 - Longitud del sépalo
 - Ancho del sépalo
 - Longitud del pétalo
 - Ancho del pétalo

Datos estructurados

- Ejemplo: Iris dataset



Muestras	#	Longitud del sépalo	Ancho del sépalo	Longitud del pétalo	Ancho del pétalo	Clase
	1	5,1	3,5	1,4	0,2	Setosa
	2	4,9	3,0	1,4	0,2	Setosa
	...					
	50	5,5	3,5	4,5	1,2	Versicolor
	...					
	150	5,9	3,0	5,0	1,8	Virginica
<div> <div>Características (atributos, medidas)</div> <div>Etiqueta (objetivo)</div> </div>						

Datos no estructurados

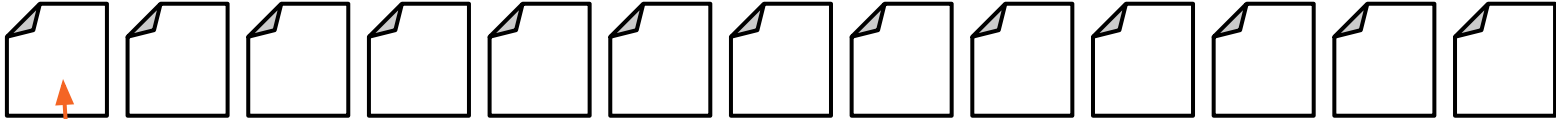


Preprocesado

- Preprocesado de los ficheros/datos
- Extracción de características
- Normalización
- Forma de los datos

Particiones

- Disponemos de un conjunto de datos:

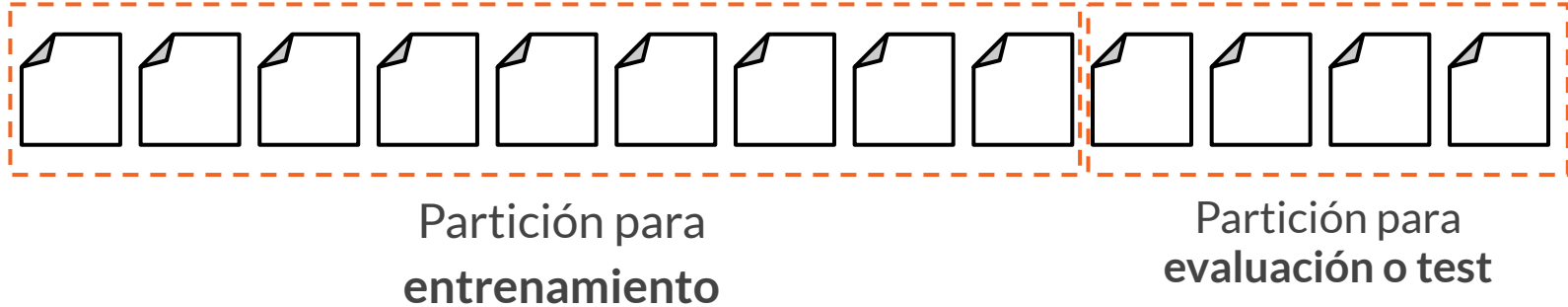


Muestra

(puede contener o no su etiqueta asociada)

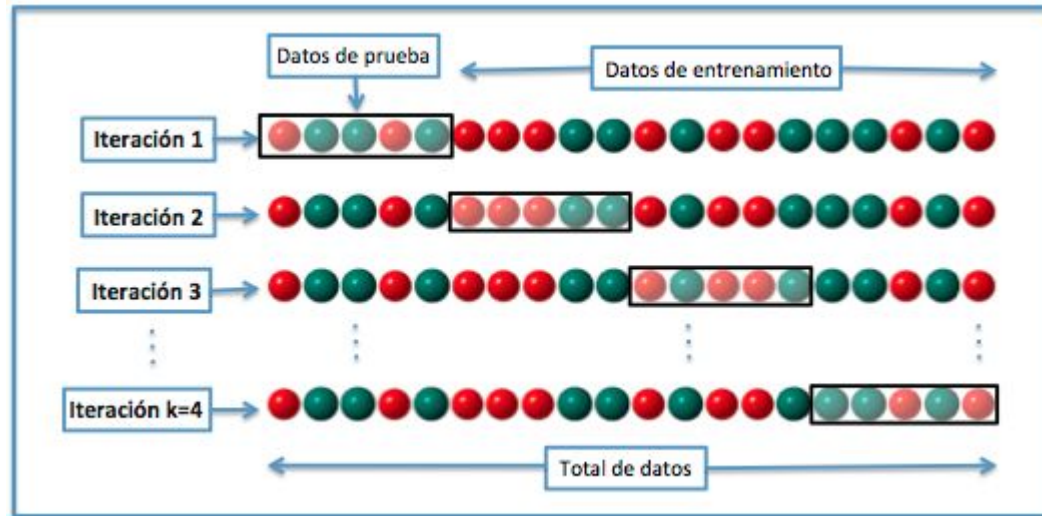
Particiones

- Creamos las particiones para entrenar y evaluar:



Validación cruzada

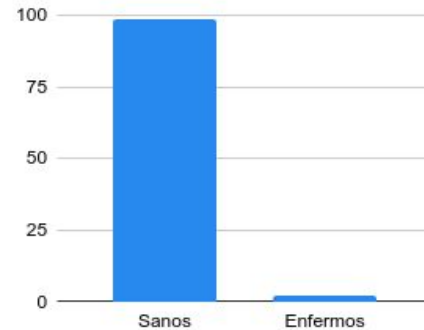
- Existen más esquemas de entrenamiento / evaluación:



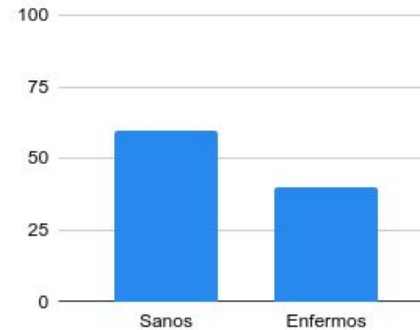
Problemas con las particiones

Hay que llevar cuidado al crear las particiones para no crear distintas distribuciones y que todas las clases o tipos de datos estén igualmente representados.

Conjunto de entrenamiento



Conjunto de test



Etapas

Etapas

1. Preparar los datos
2. Entrenar el modelo de ML
3. Validar el modelo generado
4. Evaluación final

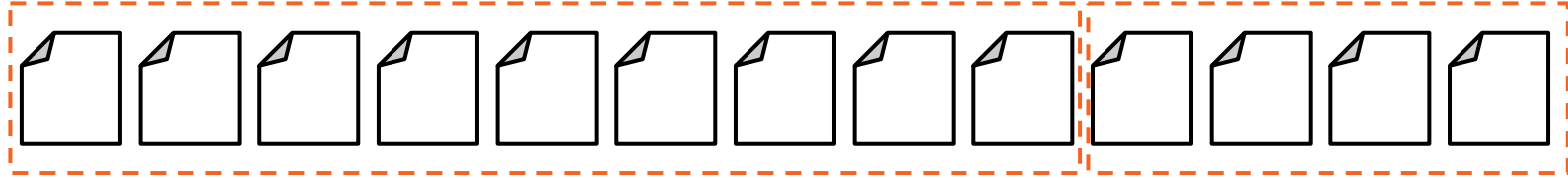
Etapas



Ajuste de hiperparámetros

- Problema:
 - Siguiendo el esquema anterior corremos el riesgo de **sobreajustar** el modelo a los datos de test.
 - A veces no hay test para realizar este proceso.
 - Incluso si no usamos el test estaríamos sobreajustando la configuración a los datos de entrenamiento.
- Solución:
 - Sacar una tercera partición de los datos de entrenamiento (que llamaremos **validación**) para el ajuste de los hiperparámetros.
- **Nota:** Esto solo es necesario si hay que ajustar hiperparámetros.

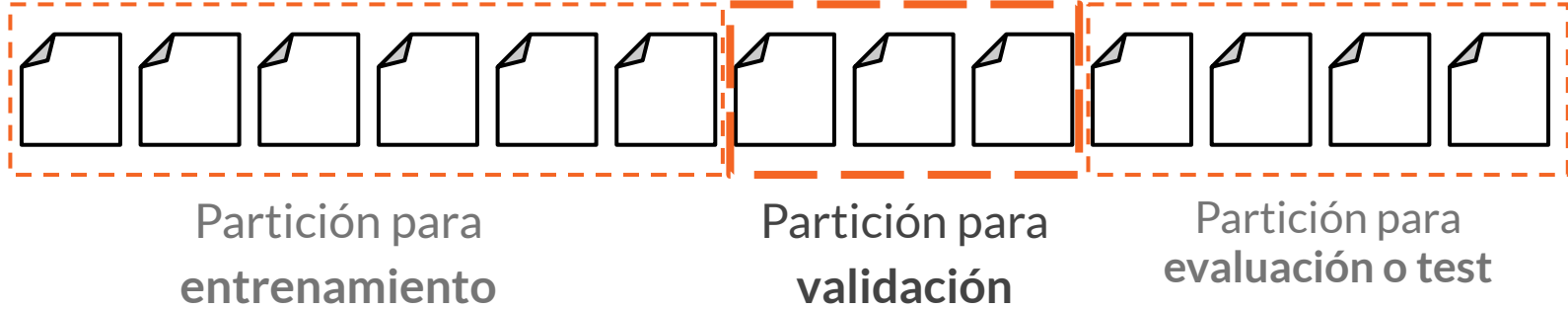
Particiones



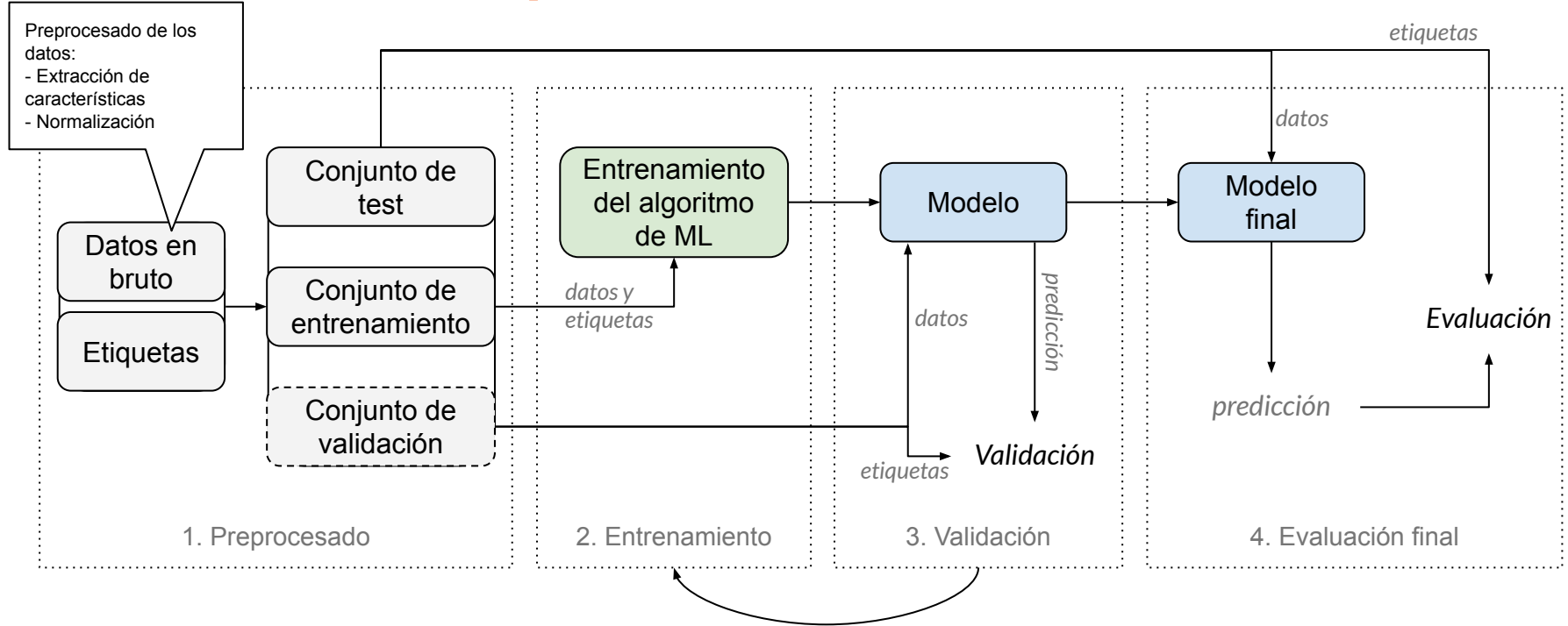
Partición para
entrenamiento

Partición para
evaluación o test

Partición de validación



Etapas



Librerías

Librerías

- Python
- Sklearn (<https://scikit-learn.org/stable/index.html>)
- Numpy
- Matplotlib
- Pandas



cursos d'estiu | cursos de verano
rafael
altamira

Introducción al *Machine Learning*

Conceptos básicos sobre el aprendizaje
automático

Profesor: Antonio Javier Gallego
