

Fundamentos del Aprendizaje Automático

Febrero 2023



red.es



"El FSE invierte en tu futuro"
Fondo Social Europeo

1. Conceptos fundamentales

- 2.1 Diagrama de bloques de un sistema de aprendizaje
- 2.2 Elementos básicos
- 2.3 Objetivo
- 2.4 Conjuntos de entrenamiento y test
- 2.5 Pasos

1. Ejemplo práctico

2. Tipos de modelos de ML

- 3.1 Aprendizaje supervisado / no supervisado
- 3.2 Aprendizaje por refuerzo
- 3.3 Aprendizaje incremental vs batch
- 3.4 Modelos paramétricos / no paramétricos

3. Regresión

4. Clasificación

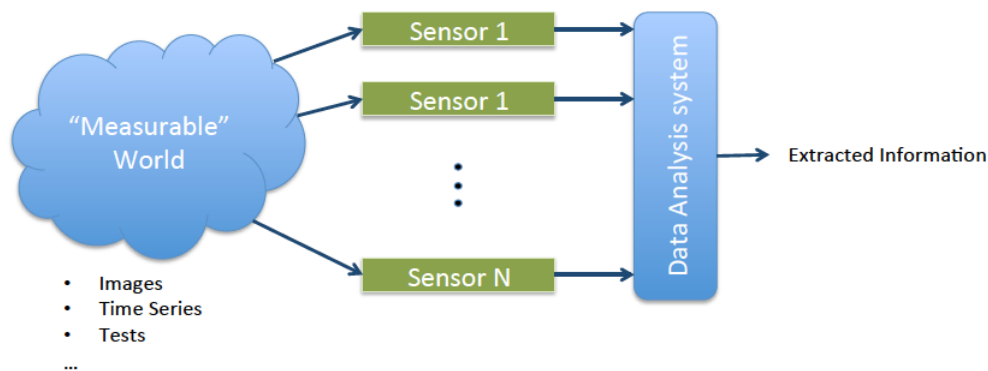
5. Modelos no supervisados

6. Redes neuronales

Conceptos fundamentales

Conceptos fundamentales del ML

Diagrama de bloques de un sistema de aprendizaje



- La información deseada no se puede acceder directamente, así que hay que usar una serie de variable estadísticas relacionadas.
- El análisis de datos explota esta información estadística para proporcionar resultados precisos...
- ...pero algunos errores son generalmente inevitables.
- Mucha inversión está justificada cuando el sistema se opera intensivamente en diferentes datos.

Conceptos fundamentales del ML

Elementos básicos

Construir modelos que se ajustan a una colección de datos

- **Modelo**: objeto de una clase (programa informático) o función matemática (modelo estadístico) con parámetros libres.
- **Conjunto de entrenamiento**: Conjunto de ejemplos sacados de la distribución de datos que se pretende modelar.
- **Optimizador**: Método que ajusta los valores de los parámetros libres del modelo (entrena) para que capturen patrones informativos contenidos en los datos.

Conceptos fundamentales del ML

Objetivo

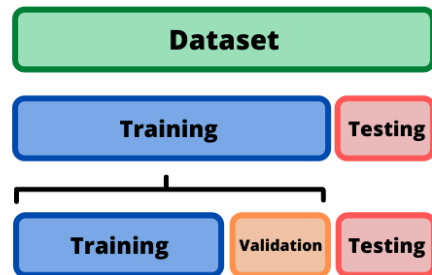
Conseguir que el modelo pueda hacer predicciones más o menos correctas cuando se le presenten datos que no se hayan usado durante el entrenamiento.

Generalización

Evitar el **sobreentrenamiento**, que provoca un error pequeño en entrenamiento pero elevado en test.

Conceptos fundamentales del ML

Conjuntos de...



Test

Conjunto de datos que nunca se usa durante el entrenamiento ni durante la optimización de los parámetros, sólo para calcular la puntuación de test tras la finalización de dicho proceso.

Entrenamiento

Conjunto de datos usado durante el entrenamiento del modelo.

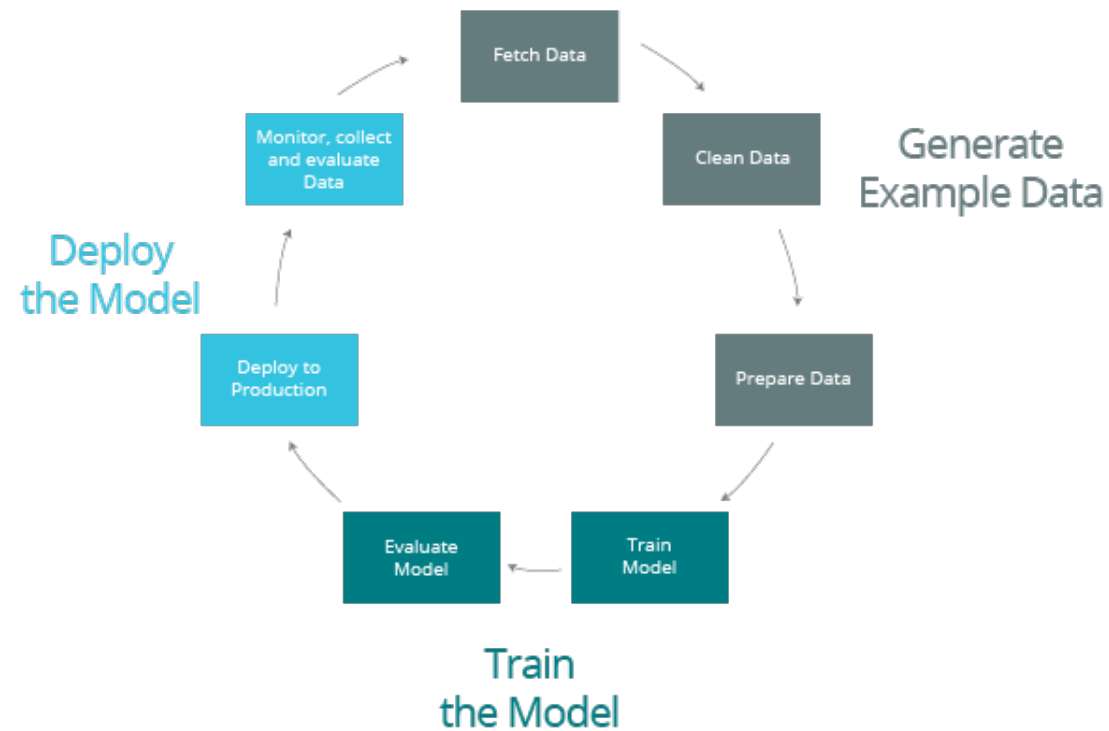
Validación

Conjunto de datos usado para evaluar el modelo durante el entrenamiento y elegir el mejor valor para los parámetros, los que obtengan mejor puntuación en dicho conjunto.

Conceptos fundamentales del ML

Pasos

Machine Learning Pipeline



Ejemplo práctico

Ejemplo práctico

Ejemplo: Reconocimiento de dígitos manuscritos

Entrada del modelo

Imagen de 28x28 píxeles = vector x de 784 números reales

Salida del modelo

Identidad del dígito 0,...,9

Conjunto de entrenamiento: N dígitos:

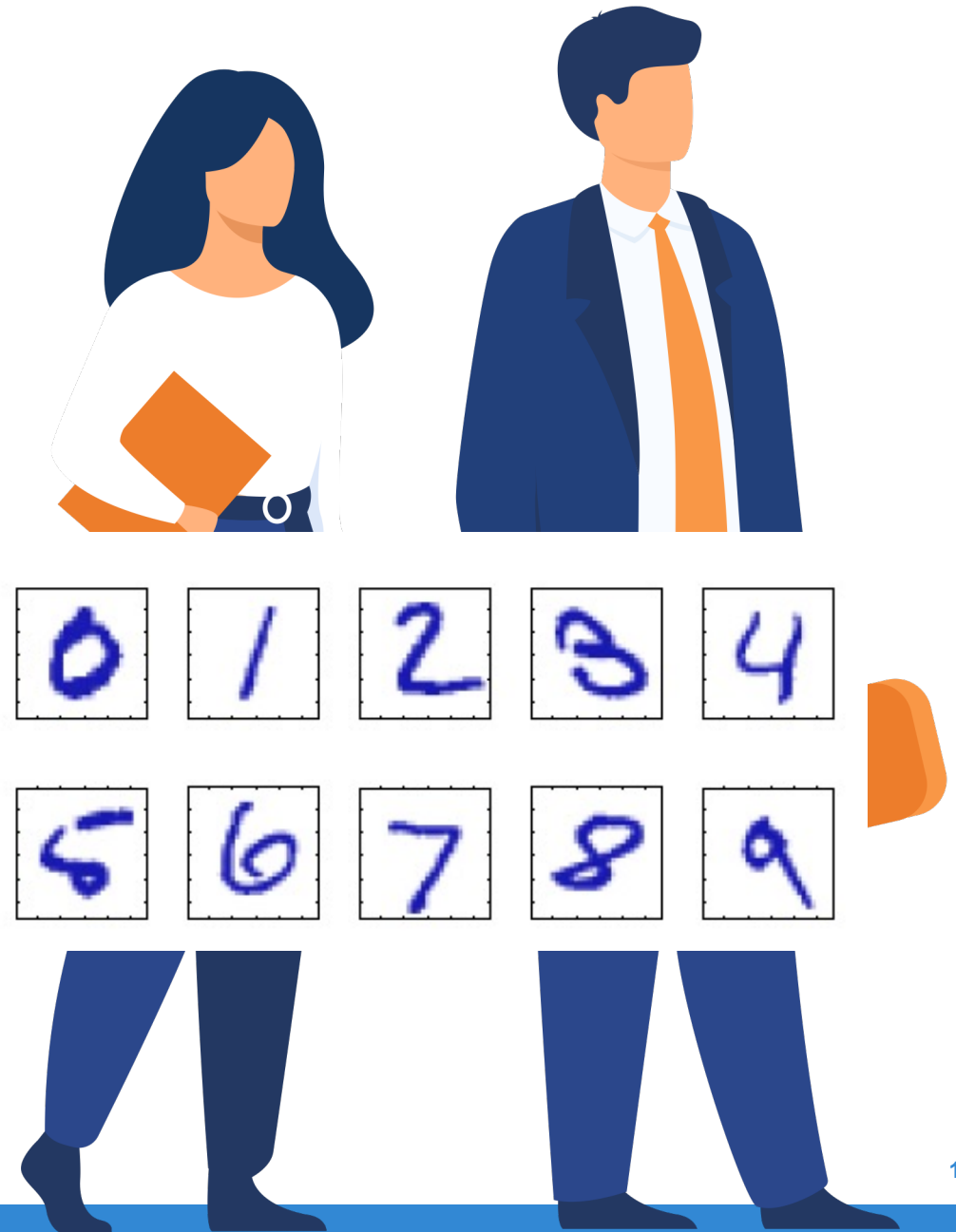
$$\{x_1, \dots, x_N\}$$

Vector objetivo: Identidad de los dígitos

$$\{t_1, \dots, t_N\}$$

Resultado del algoritmo: expresado como...

$$y(x)$$



Ejemplo práctico

Ejemplo: Reconocimiento de dígitos manuscritos

Conjunto de test: El modelo debe ser capaz de determinar la identidad de nuevas imágenes de dígitos

Generalización: Capacidad del modelo para clasificar nuevos ejemplos diferentes de los usados en entrenamiento.

Preprocesado: Las imágenes de los dígitos de entrenamiento se trasladan y se escalan a una caja de tamaño fijo

Extracción de características / *Feature extraction*

Aplicar también a los datos de test



Tipos de modelos de ML

Tipos de modelos de ML

Aprendizaje supervisado

Objetivo: Aprender un mapeo de entradas \mathbf{x} a salidas y , dado un conjunto etiquetado de pares de entrada-salida. Aprende una relación 1 a 1 entre observaciones y targets.

$$D = \{\mathbf{x}_i, y_i\}_{i=1}^N$$



D-dimensional vector: características / *features*

Aprendizaje no supervisado

Sólo se dan datos de entrada.

Objetivo: Encontrar 'patrones interesantes' en los datos. No se dispone de un target/salida.

$$D = \{\mathbf{x}_i\}_{i=1}^N$$

*Existen otros paradigmas como el **aprendizaje semi-supervisado** o el **aprendizaje por refuerzo**.

Tipos de modelos de ML

Aprendizaje supervisado

- Scores para recomendaciones de netflix, *spotify*, etc
- Sistemas de concesiones de créditos, seguros, etc
- Reconocimiento de caras, huellas, etc
- Filtrado de correos electrónicos, noticias, etc
- Sistemas de ayuda a la diagnosis en aplicaciones clínicas

Aprendizaje no supervisado

- Agrupar colecciones de datos
- Limpieza de *outliers*
- Segmentar vídeo o audio
- Segmentación de clientes
- Organización de colecciones de documentos
- Aprendizaje de funciones de densidad de probabilidad

Tipos de modelos de ML

Aprendizaje supervisado

Objetivo: Aprender un mapeo de entradas \mathbf{x} a salidas y , dado un conjunto etiquetado de pares de entrada-salida. Aprende una relación 1 a 1 entre observaciones y targets.

$$D = \{\mathbf{x}_i, y_i\}_{i=1}^N$$

Regresión

Aprender una función que para cada observación prediga un número real o *score*

y_i *Es un número real*

Clasificación

Aprender una función que para cada observación prediga a qué clase, dentro de un conjunto finito y establecido a priori, pertenece esta observación

y_i *Es una categórica*

Tipos de modelos de ML

Agrupamiento o clustering

Dividir un conjunto de datos (observaciones) en K partes que contengan datos parecidos entre sí y diferentes de los datos contenidos en los otros grupos.

Detección de novedad

Aprender el soporte de la distribución de datos para ayudar a decidir si un dato de test pertenece a esa distribución o no.

Densidad de probabilidad

Aprendizaje de **densidades de probabilidad** a partir de una muestra finita de datos

Reducción de dimensionalidad

Aprender transformaciones de las variables de entrada para conseguir un conjunto con menos variables

Aprendizaje no supervisado

Sólo se dan datos de entrada.

Objetivo: Encontrar 'patrones interesantes' en los datos.
No se dispone de un target/salida.

$$D = \{\mathbf{x}_i\}_{i=1}^N$$

Tipos de modelos de ML

Aprendizaje supervisado

Objetivo: Aprender un mapeo de entradas \mathbf{x} a salidas y , dado un conjunto etiquetado de pares de entrada-salida. Aprende una relación 1 a 1 entre observaciones y targets.

$$D = \{\mathbf{x}_i, y_i\}_{i=1}^N$$



D -dimensional vector: características / *features*

Aprendizaje no supervisado

Sólo se dan datos de entrada.

Objetivo: Encontrar 'patrones interesantes' en los datos. No se dispone de un target/salida.

$$D = \{\mathbf{x}_i\}_{i=1}^N$$

Tipos de modelos de ML

Aprendizaje por refuerzo

- Los datos están formados por **secuencias** de observaciones/decisiones que desembocan en una **recompensa**
- Se aprende una **estrategia** para encadenar una secuencia de decisiones que maximicen la recompensa global
- Aplicaciones:
 - Conducción autónoma
 - Robótica
 - Videojuegos y juegos de mesa
 - diseño de estrategias de *trading*

Tipos de modelos de ML

Aprendizaje incremental vs batch

Condicionado por el problema de optimización que haya que resolver

- **Batch:** se dispone de una vez del conjunto completo de datos de entrenamiento. Generalmente es el caso cuando buscamos una optimización que nos dé un modelo globalmente óptimo
- **Incremental** no se dispone del conjunto de datos de entrenamiento completo. Estrategia a emplear cuando el procesador en el que se ejecuta el algoritmo de entrenamiento no puede con todos los datos a la vez

Tipos de modelos de ML

Modelos paramétricos

Número fijo de parámetros

- ✓ Más rápidos
- Suposiciones más fuertes sobre la naturaleza de las distribuciones de los datos

Se suele dar a la función estimación una forma paramétrica a priori, y el propósito de diseño es encontrar los valores de los parámetros de acuerdo a cierto objetivo.

Ejemplos: regresores lineales, regresión logística, agrupamiento k medias,...

Modelos no paramétricos

El número de parámetros crece con la cantidad de muestras de entrenamiento.

- ✓ Más flexibles
- Intratables computacionalmente para datasets más grandes.

La forma analítica del modelo no se asume a priori.

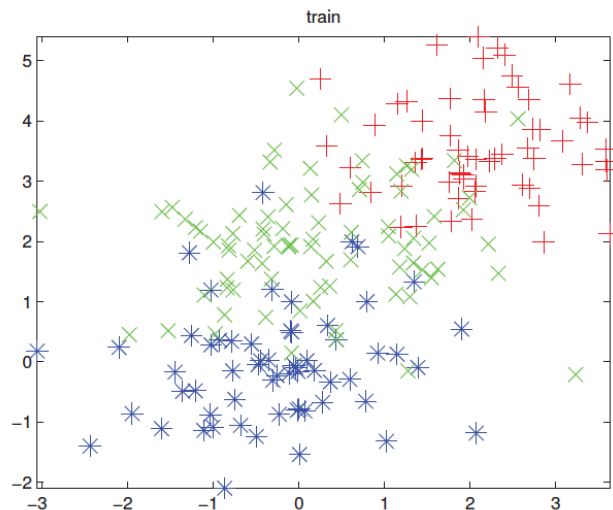
Ejemplos: KNN, SVM,...

Tipos de modelos de ML

K nearest neighbor KNN

Mirar los K vecinos del conjunto de entrenamiento más cercanos a la muestra de test, contar los miembros de cada clase y devolver la fracción empírica como la estimación.

Distancia: ejemplo la euclídea.

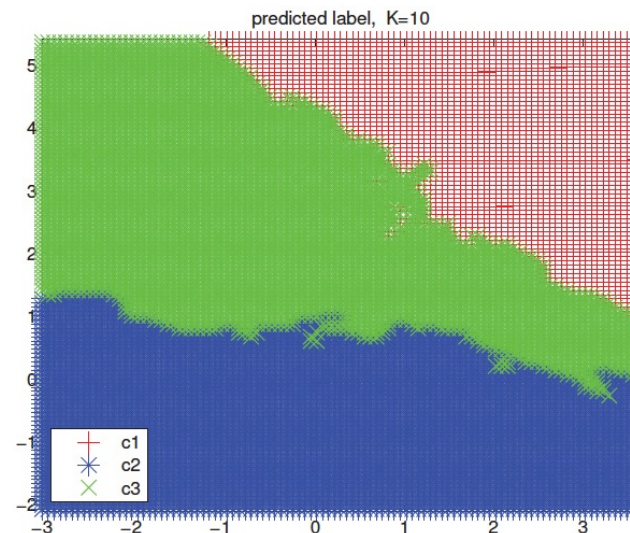


Modelos no paramétricos

El número de parámetros crece con la cantidad de muestras de entrenamiento.

- ✓ Más flexibles
- Intratables computacionalmente para datasets más grandes.

La forma analítica del modelo no se



Tipos de modelos de ML

K nearest neighbor KNN

Mirar los K vecinos del conjunto de entrenamiento más cercanos a la muestra de test, contar los miembros de cada clase y devolver la fracción empírica como la estimación.

Distancia: ejemplo la euclídea.

Solución: modelos paramétricos, que hacen suposiciones sobre la naturaleza de las distribuciones de los datos.

Modelos no paramétricos

La maldición de la dimensionalidad

- KNN funciona bien con datos de baja dimensionalidad, pero no alta.
- El método deja de ser local a altas dimensiones.
- El problema de mirar vecinos que están tan lejos es que puede que no sean buenos predictores

Tipos de modelos de ML

Modelos paramétricos

Número fijo de parámetros

- ✓ Más rápidos
- Suposiciones más fuertes sobre la naturaleza de las distribuciones de los datos

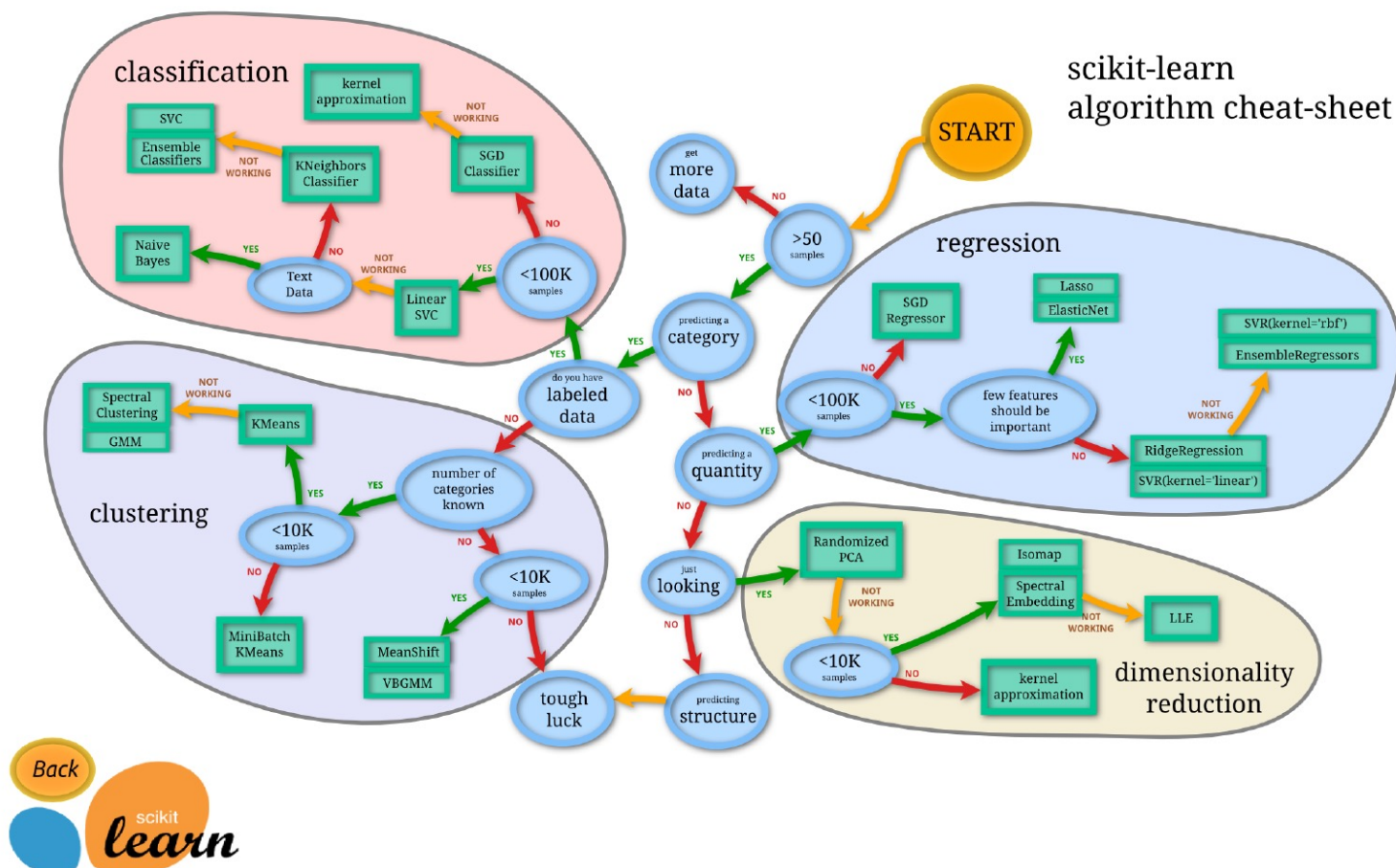
Se suele dar a la función estimación una forma paramétrica a priori, y el propósito de diseño es encontrar los valores de los parámetros de acuerdo a cierto objetivo.

Regresión lineal

Regresión logística

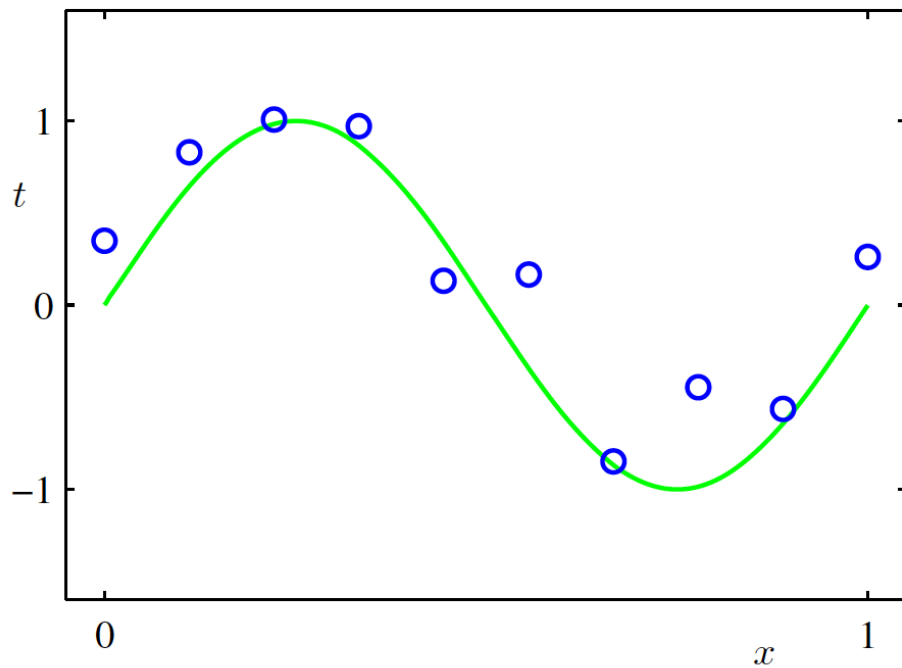
Probability theory is nothing but common sense reduced to calculation. — Pierre Laplace, 1812

Tipos de modelos de ML



Regresión

Ejemplos prácticos

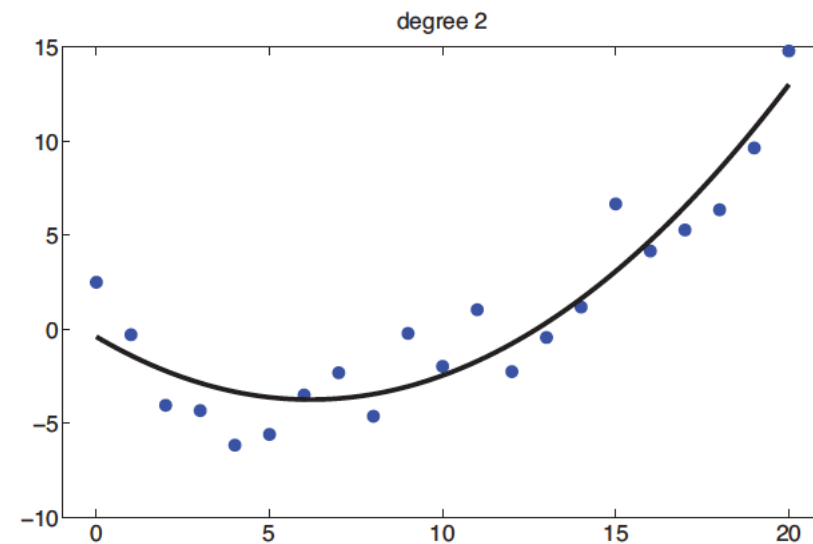
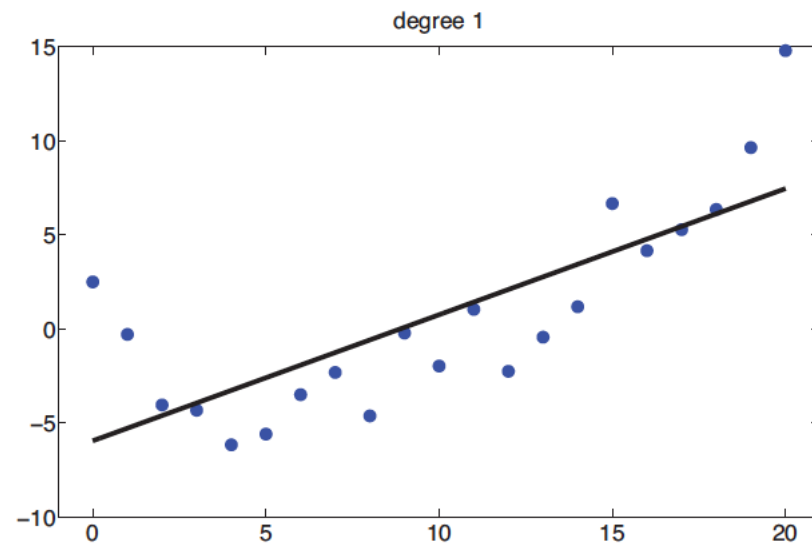


Regresión

- La variable de respuesta es continua

$$y_i \in \mathbb{R}$$

Ejemplos: linear vs polinómica



Regresión

Aplicaciones del mundo real

- Predecir el precio del mercado de valores de mañana dadas las condiciones del mercado actual y otra posible información lateral.
- Predecir la edad de un espectador viendo un video de YouTube dado.
- Predecir la localización en 3D de un brazo robot dadas señales de control enviadas a sus motores.
- Predecir la cantidad de antígeno específico de la próstata en el cuerpo como una función de un número de diferentes medidas clínicas.
- Predecir la temperature en cualquier localización dentro de un edificio utilizando datos meteorológicos, tiempo, sensores,...



Regresión

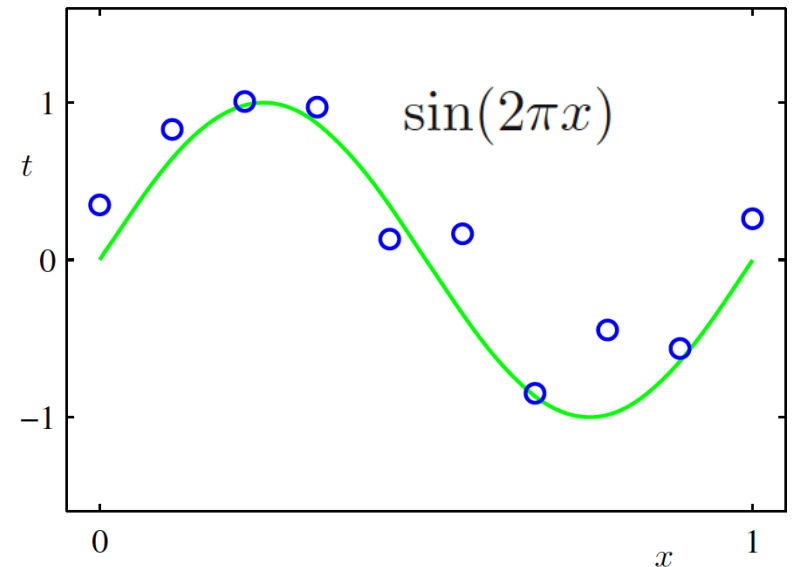
Ejemplo: ajustando una curva polinómica

- Datos de entrada:

$$\mathbf{x} \equiv (x_1, \dots, x_N)^T$$

$$\mathbf{t} \equiv (t_1, \dots, t_N)^T$$

- Generados a partir de una función seno con un pequeño nivel de ruido Gaussiano añadido.
- Objetivo:** Explotar el conjunto de entrenamiento para hacer predicciones del valor t de la variable objetivo para nuevos valores de la variable x de entrada.
- Implícitamente se trata de descubrir la función generadora de los datos.



$$y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j$$

Regresión

Ejemplo: ajustando una curva polinómica

- Modelo **lineal**:

Función lineal en los coeficientes, \mathbf{w} .

- \mathbf{w} ?

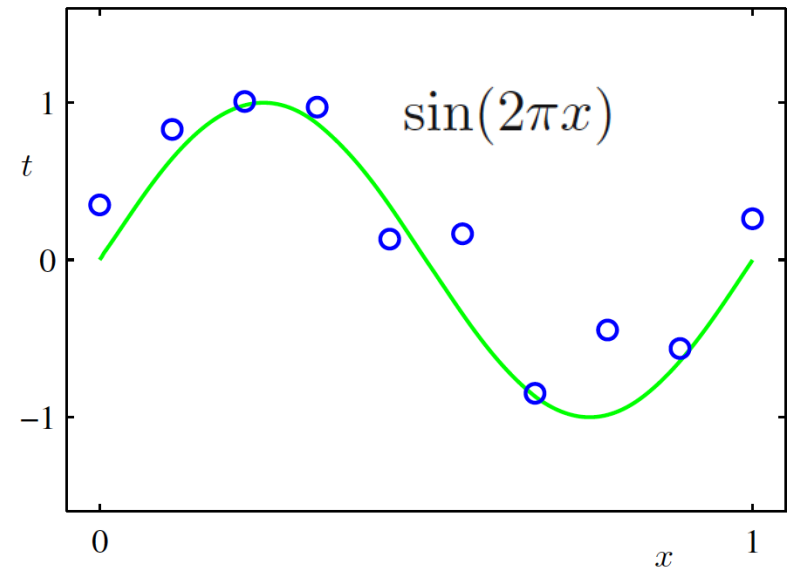
Ajustar el polinomio a los datos de entrenamiento

Minimizar una **función error**, que mide la diferencia entre la función $y(x, \mathbf{w})$, para unos valores de \mathbf{w} dados, y los datos del conjunto de entrenamiento.

- Función error ejemplo: suma de los errores cuadráticos.

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2$$

$$y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j$$



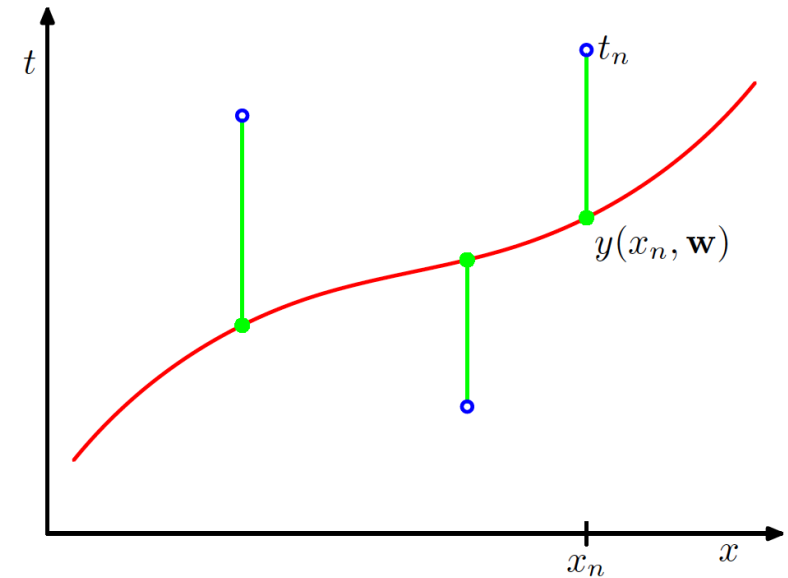
Regresión

Ejemplo: ajustando una curva polinómica

- Función no-negativa
- Minimizar

Calcular derivadas con respecto a los parámetros e igualar a cero

- Solución: lineal, solución única $y(x, \mathbf{w}^*)$
- Grado M? **Selección del modelo / model selection**

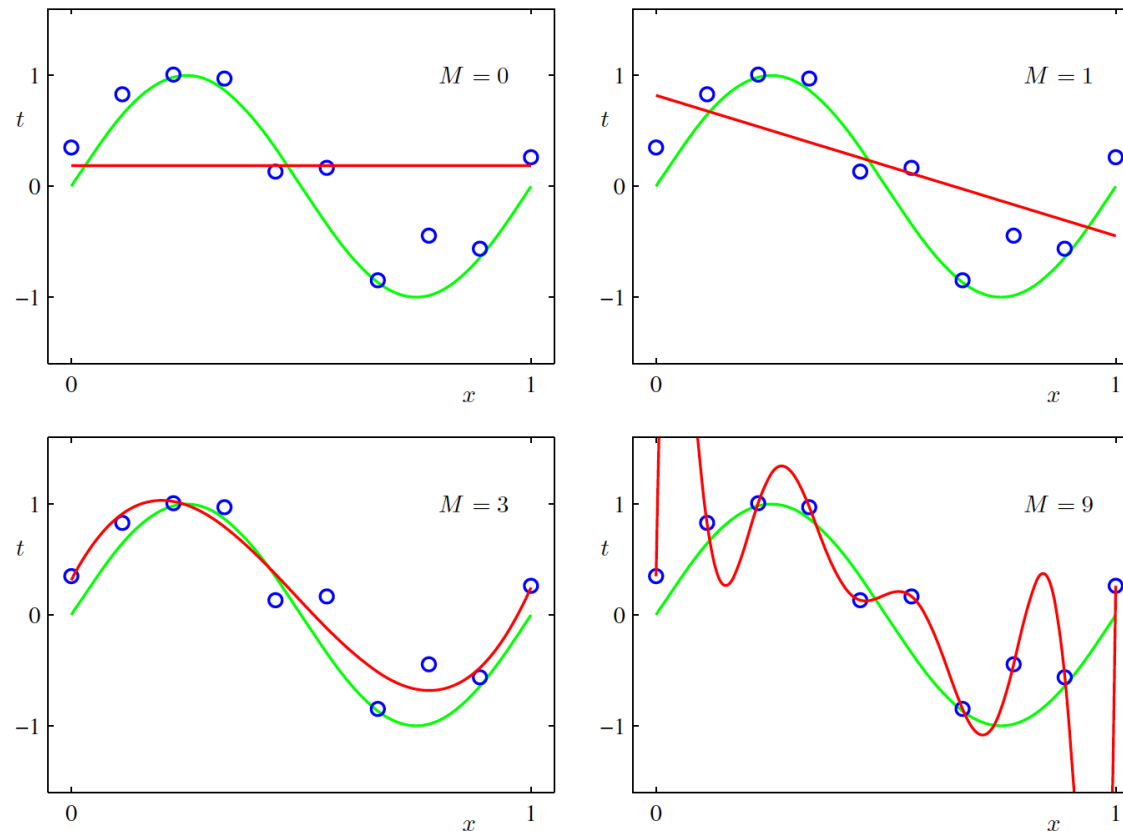


$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2$$

$$y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j$$

Regresión

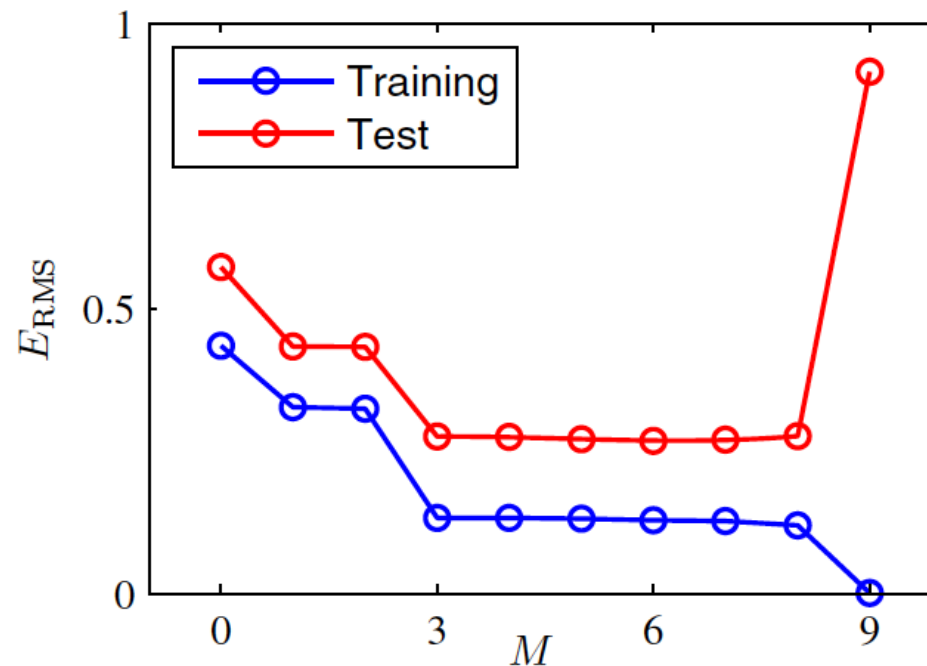
Ejemplo: ajustando una curva polinómica



Sobre-entrenamiento
Overfitting

Regresión

Ejemplo: ajustando una curva polinómica



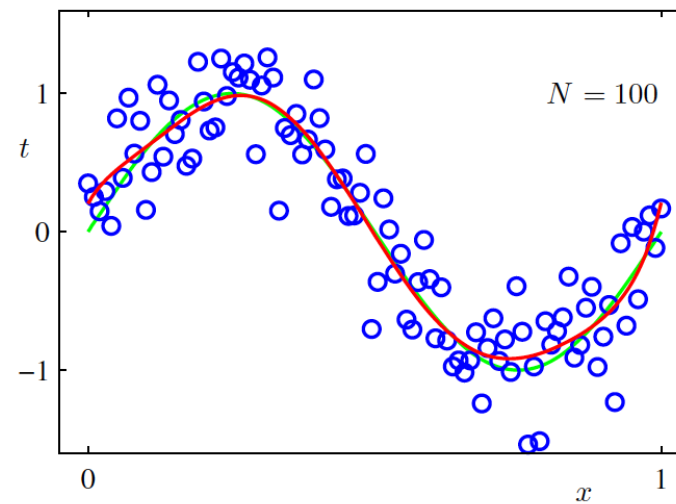
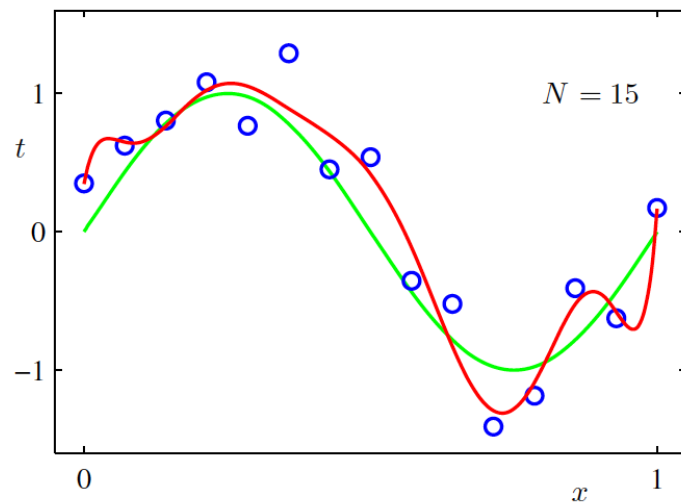
	$M = 0$	$M = 1$	$M = 6$	$M = 9$
w_0^*	0.19	0.82	0.31	0.35
w_1^*		-1.27	7.99	232.37
w_2^*			-25.43	-5321.83
w_3^*			17.37	48568.31
w_4^*				-231639.30
w_5^*				640042.26
w_6^*				-1061800.52
w_7^*				1042400.18
w_8^*				-557682.99
w_9^*				125201.43

Sobre-entrenamiento
Overfitting

Regresión

Ejemplo: ajustando una curva polinómica

¿Cómo evitarlo?



Sobre-entrenamiento
Overfitting

Ajustar número de parámetros al número de muestras

Regresión

Ejemplo: ajustando una curva polinómica

Regularización

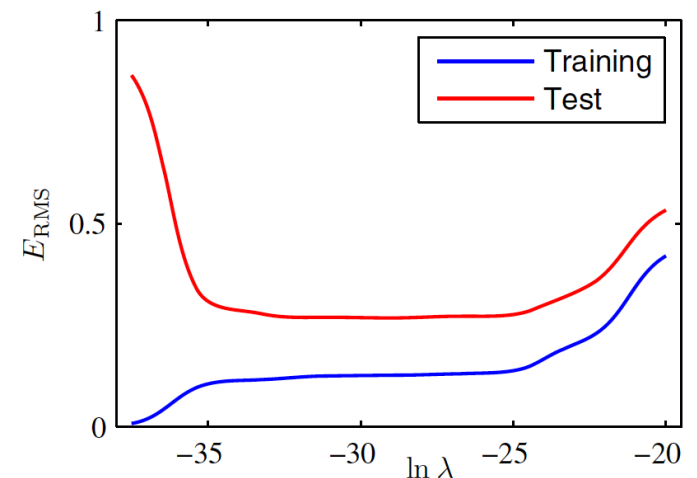
Regresión Rigde

$$\tilde{E}(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2$$

$$\|\mathbf{w}\|^2 \equiv \mathbf{w}^T \mathbf{w} = w_0^2 + w_1^2 + \dots + w_M^2$$

Añadir un término de penalización para que los parámetros no alcancen valores grandes.

¿Cómo evitarlo?



Sobre-entrenamiento
Overfitting

Regresión

Ejemplo: ajustando una curva polinómica

Regularización

- Conjunto de entrenamiento: usado para calcular los coeficientes, \mathbf{w} .
- Conjunto de validación: usado para seleccionar el modelo, M o λ

¿Cómo evitarlo?

Sobre-entrenamiento
Overfitting

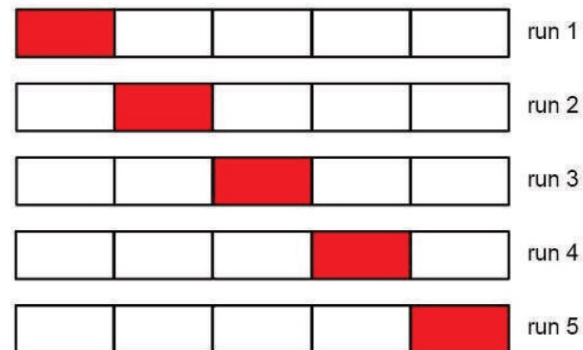
Regresión

Ejemplo: ajustando una curva polinómica

¿Cómo evitarlo?

Validación cruzada / *cross validation* (CV)

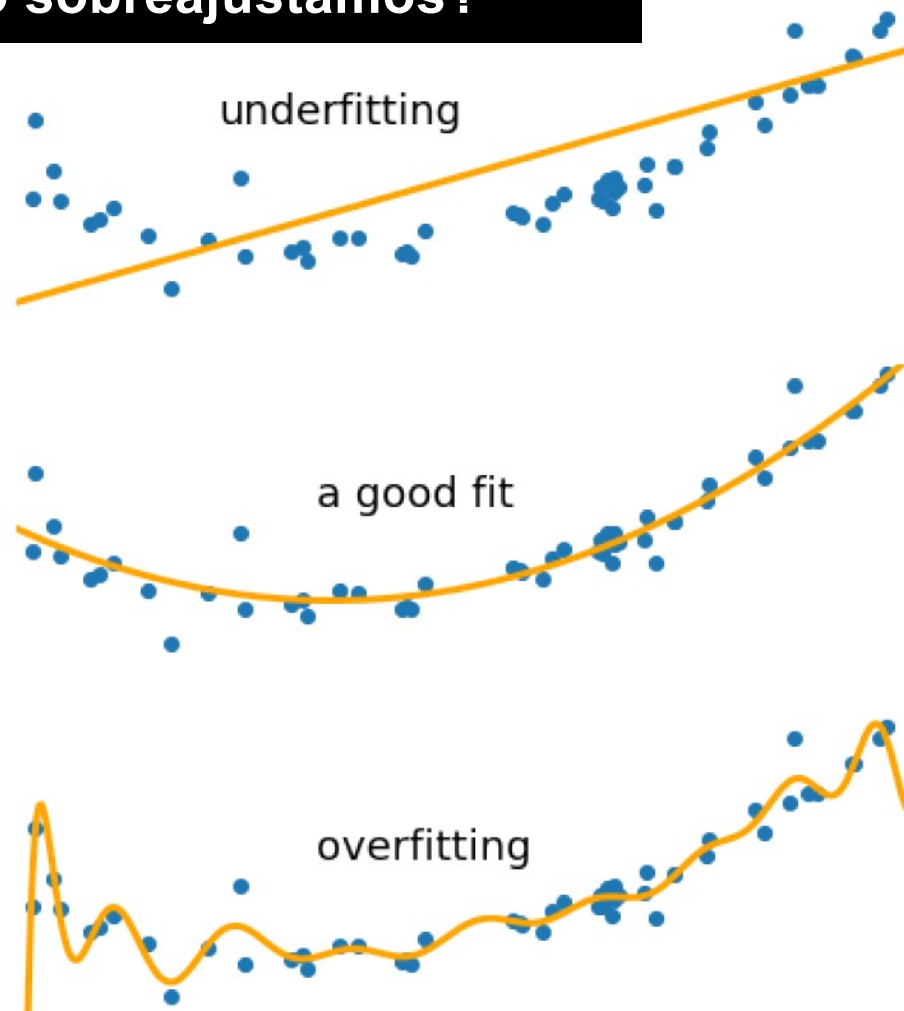
- Dividir el conjunto de entrenamiento en K particiones.
- Cada cada partición, entrenar en todas menos en una, donde testeamos.
- Error de test: Calculamos el error promedio en todas las particiones.
- Útil cuando hay pocos datos.
- *Leave-one out cross validation*.



Sobre-entrenamiento
Overfitting

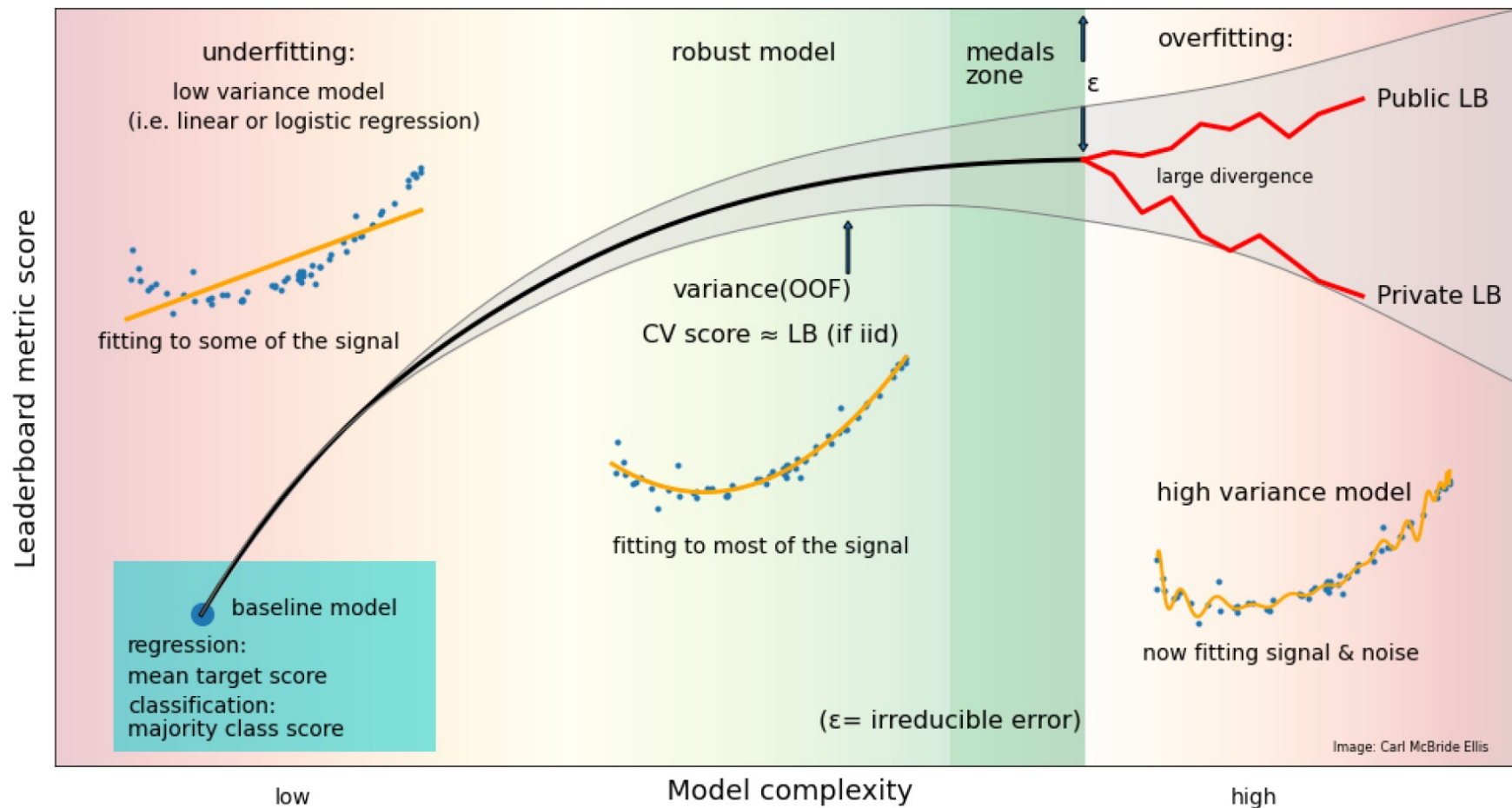
Regresión

¿Qué pasa cuando sobreajustamos?



Regresión

¿Qué pasa cuando sobreajustamos?



Ejercicios

Hora de Notebooks!!

Todos los de regresión 😊

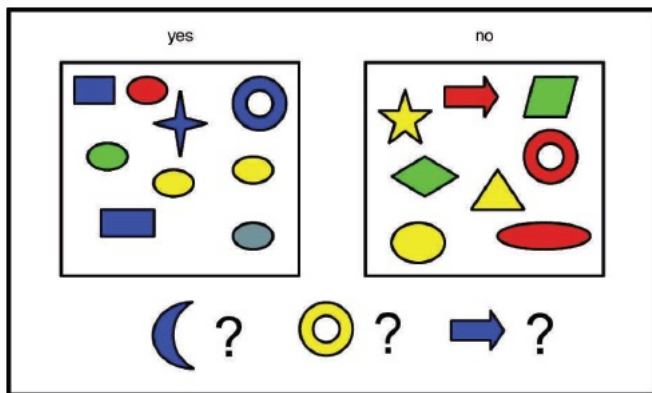
Duración estimada: Todo el día!



Clasificación

Ejemplos prácticos

		D features (attributes)			
N cases	Color	Shape	Size (cm)	Label	
	Blue	Square	10	1	
	Red	Ellipse	2.4	1	
	Red	Ellipse	20.7	0	



Clasificación

- Aprender un mapeo de entradas a salidas, donde $y \in \{1, \dots, C\}$

Tipos

- Si $C=2$, **clasificación binaria**
- Si $C>2$, **clasificación multiclase**
- Si las etiquetas no son mutuamente exclusivas, **clasificación multi-etiqueta**

Clasificación

Necesidad de predicciones probabilísticas

- Para manejar casos ambiguos, es deseable retornar una **probabilidad**.
- Decisión:

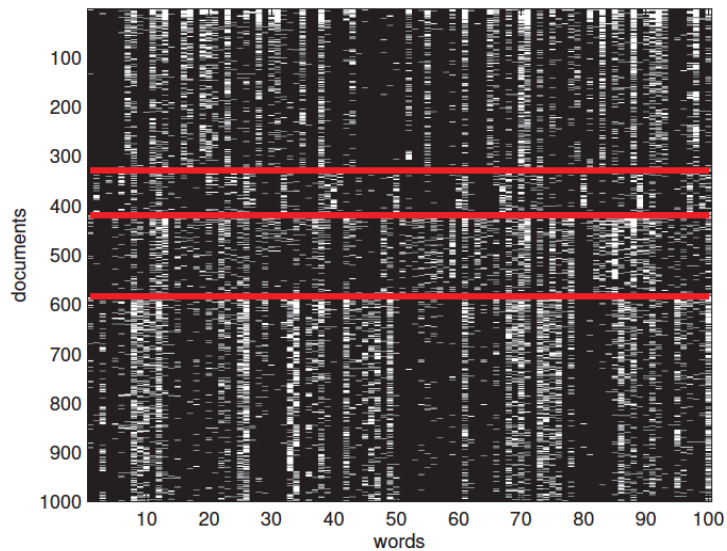
$$\hat{y} = \hat{f}(\mathbf{x}) = \operatorname{argmax}_{c=1}^C p(y = c | \mathbf{x}, \mathcal{D})$$

- Corresponde a la moda de la distribución: **MAP (Maximum a posteriori)**
- Evaluación del riesgo: importante especialmente en medicina y finanzas.

Clasificación

Aplicaciones del mundo real

Clasificación de documentos y filtrado de spam



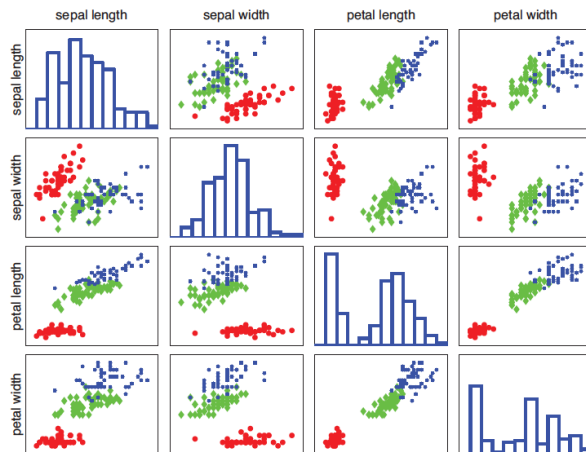
Bolsas de palabras / *bag of words*



Clasificación

Aplicaciones del mundo real

Clasificación de flores



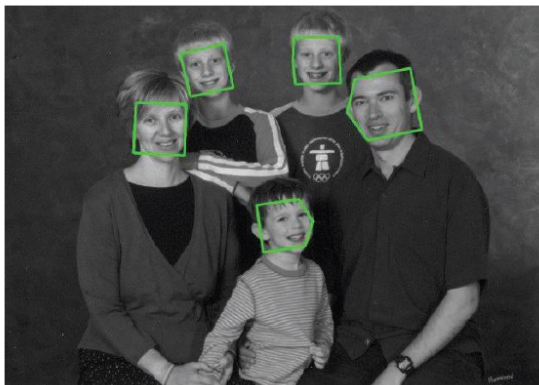
Roja: setosa
Verde: versicolor
Azul: virginica



Clasificación

Aplicaciones del mundo real

Detección y reconocimiento facial



Clasificación

Algunas métricas

Machine learning \ Manual counting	True	False
True	True Positive (TP)	False Positive (FP)
False	False Negative (FN)	True Negative (TN)

Equations:

$$\text{False positive rate (FPR)} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

$$\text{False negative rate (FNR)} = \frac{\text{FN}}{\text{FN} + \text{TP}}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{Youden index} = \text{Sensitivity} + \text{Specificity} - 1$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Ejercicios

Hora de Notebooks!!

Todos los de clasificación 😊

Duración estimada: Todo el día!



Modelos no supervisados

Modelos no supervisados

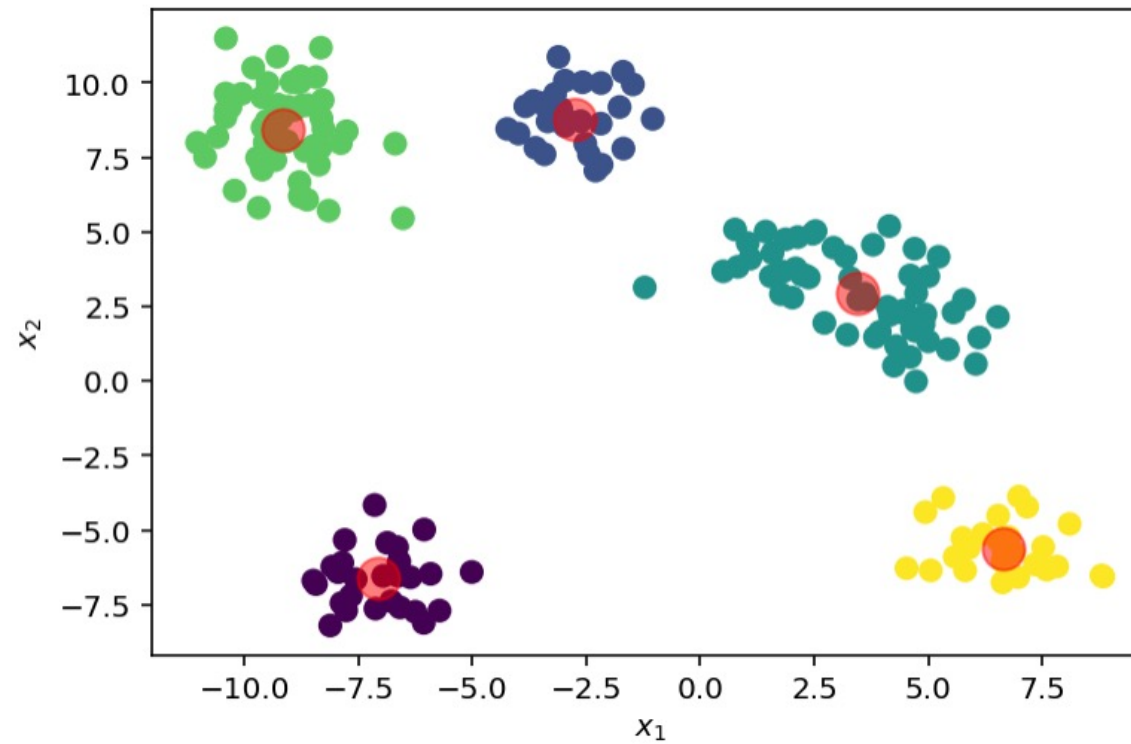
Agrupamiento o *clustering*

El agrupamiento es quizá el problema más inmediato que se puede resolver con aprendizaje automático.

El algoritmo de agrupamiento más referenciado posiblemente sea el ***Kmedias***. Divide los datos en K grupos (K es un parámetro que se debe fijar a priori) encontrando los K **centroides** o **prototipos** que mejor representan estos grupos.

Modelos no supervisados

Agrupamiento o *clustering*



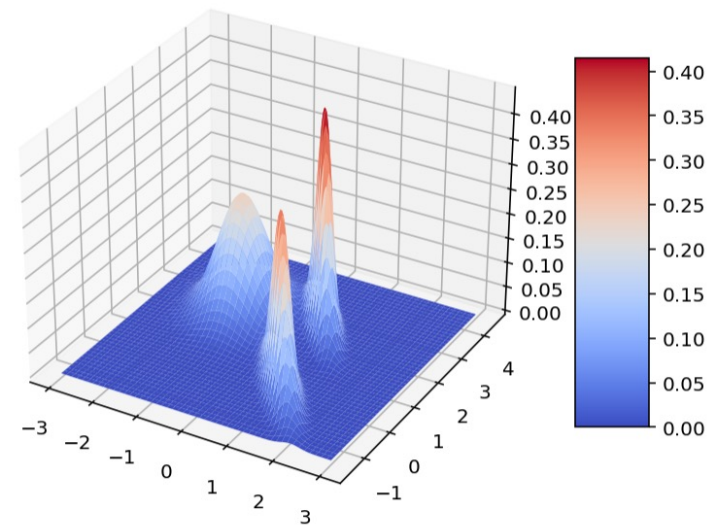
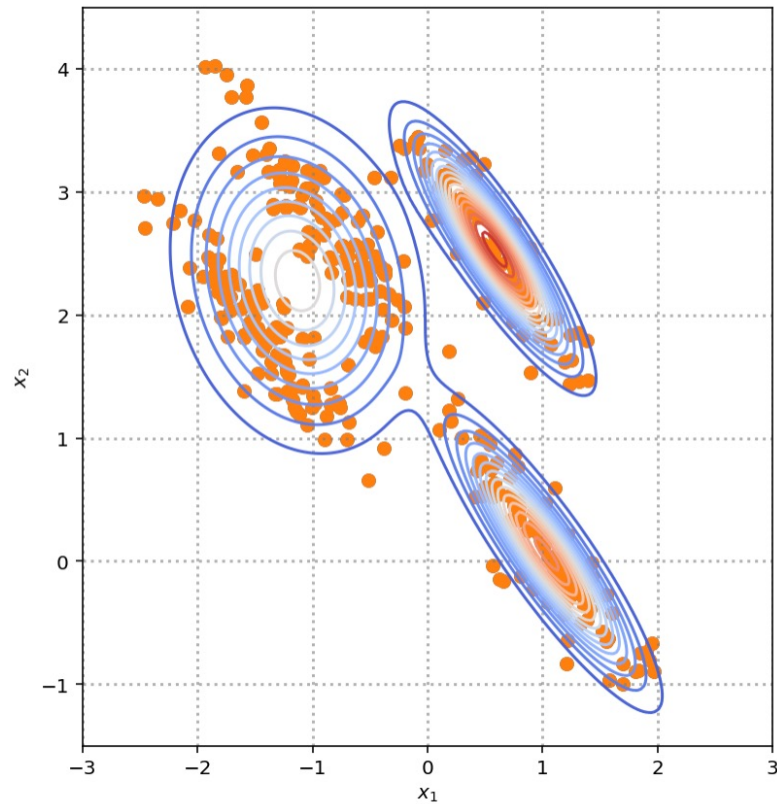
Modelos no supervisados

Aprendizaje de densidad de probabilidad

- Caracterización de los datos. La caracterización más completa desde un punto de vista matemático es mediante **la función de densidad de probabilidad** que modela el **proceso generativo** de los datos.
- Estimar densidades de probabilidad es complejo, especialmente si no se dispone de información a priori sobre el proceso real responsable de la generación de los datos.
- Frecuente: uso de **densidades aproximadas** en la construcción de los modelos mejora considerablemente las prestaciones de los mismos.
- Quizá el método más habitual para aprender densidades de probabilidad que se ajusten a una colección de datos definidos en términos de variables numéricas continuas sea la **mezcla de Gaussianas**.

Modelos no supervisados

Aprendizaje de densidad de probabilidad



Modelos no supervisados

Reducción de dimensionalidad

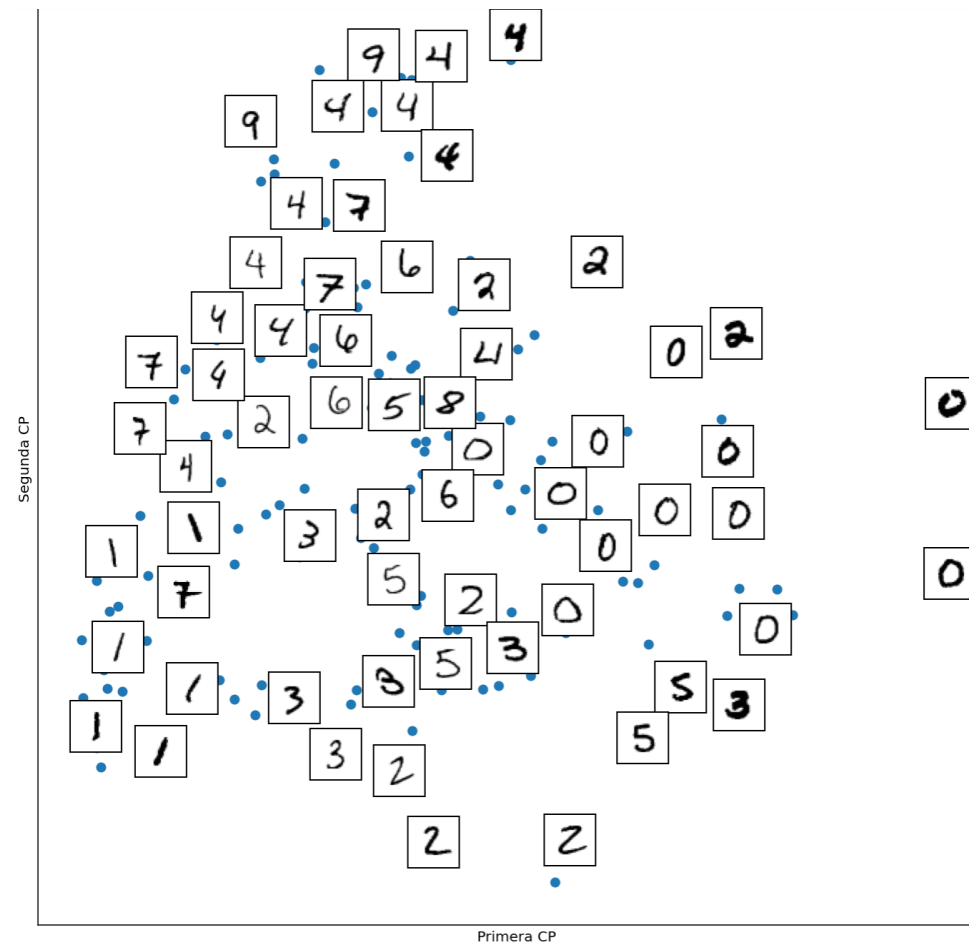
Objetivos:

1. Simplificar la carga computacional de los optimizadores al emplear datos con menos dimensiones
2. Eliminar componentes ruidosas o que no estén alineadas con el patrón que queremos capturar

El algoritmo más comúnmente empleado para este tipo de problemas es el **Análisis en Componentes Principales o PCA**.

Modelos no supervisados

Reducción de dimensionalidad



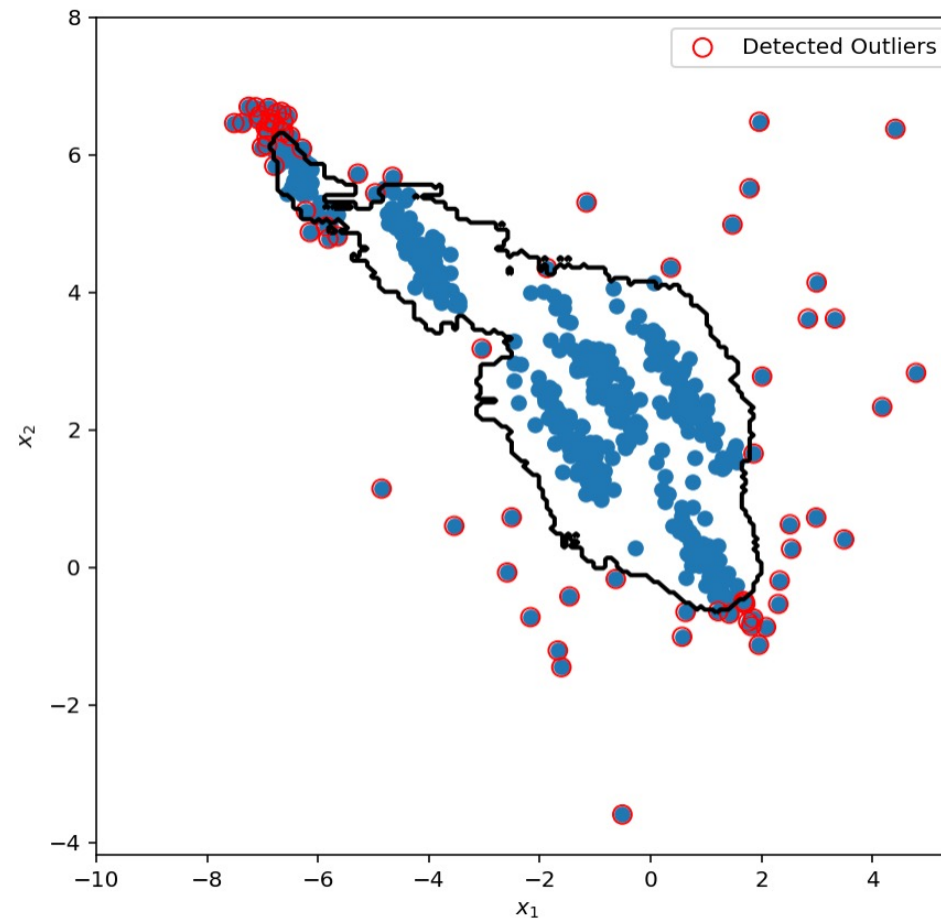
Modelos no supervisados

Detección de novedad

- Muy relacionada con el aprendizaje de densidades de probabilidad.
- Evaluar en qué medida cada una de las observaciones puede ser un **outlier**, es decir, es lo suficientemente diferente del resto de observaciones del conjunto de entrenamiento como para considerar que es una observación ruidosa o que no se ha generado con el modelo que suponemos ha generado la mayoría de los datos.
- No hace falta aprender una densidad de probabilidad que represente los datos muy fielmente, sólo nos basta con encontrar una métrica que nos permita decidir si un dato está suficientemente lejos de otras zonas más densamente pobladas de observaciones.

Modelos no supervisados

Detección de novedad



Ejercicios (EXTRA!)

Hora de Notebooks!!

Todos los de modelos no supervisados 😊



Redes neuronales

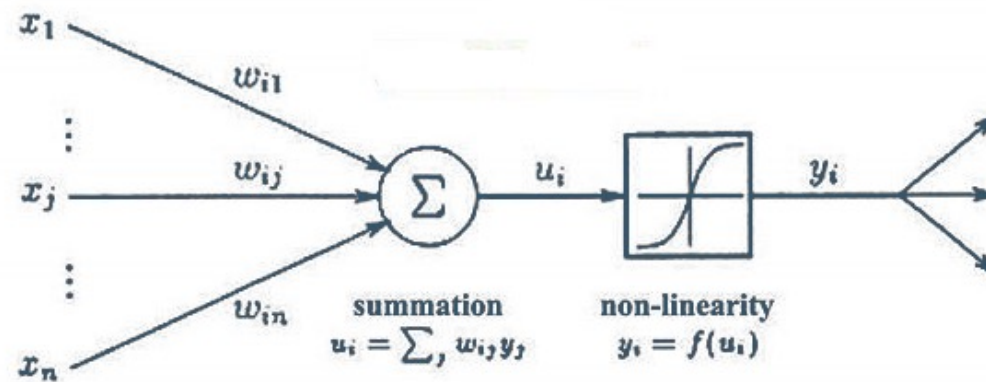
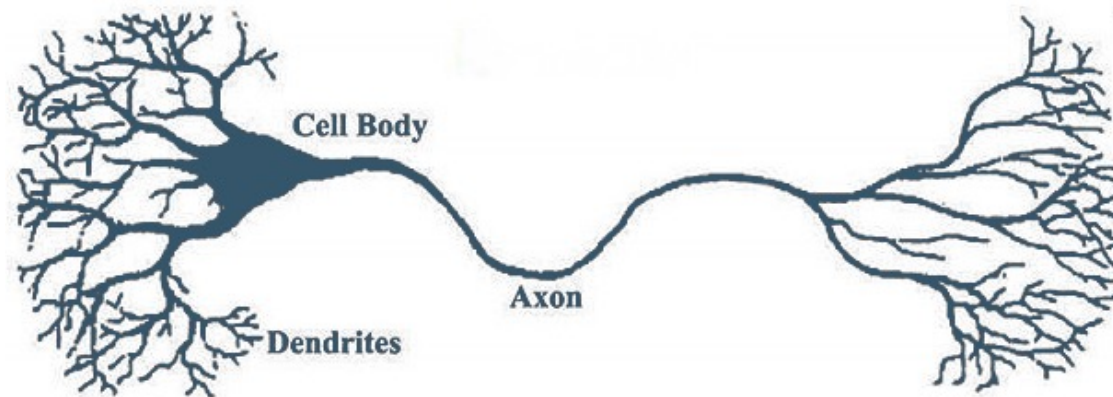
Redes neuronales

Motivación

- Posiblemente el método **más representativo** del aprendizaje automático. Siempre vuelven al foco.
- Son un **método general** que puede adaptarse a problemas de clasificación, regresión, detección de patrones, etc.
- Inspiración en las redes de neuronas naturales.
- Conexión con **neurociencia**.
- Una neurona es una unidad elemental de cálculo que recibe unas entradas, las combina linealmente y dispara con una función sigmoide.

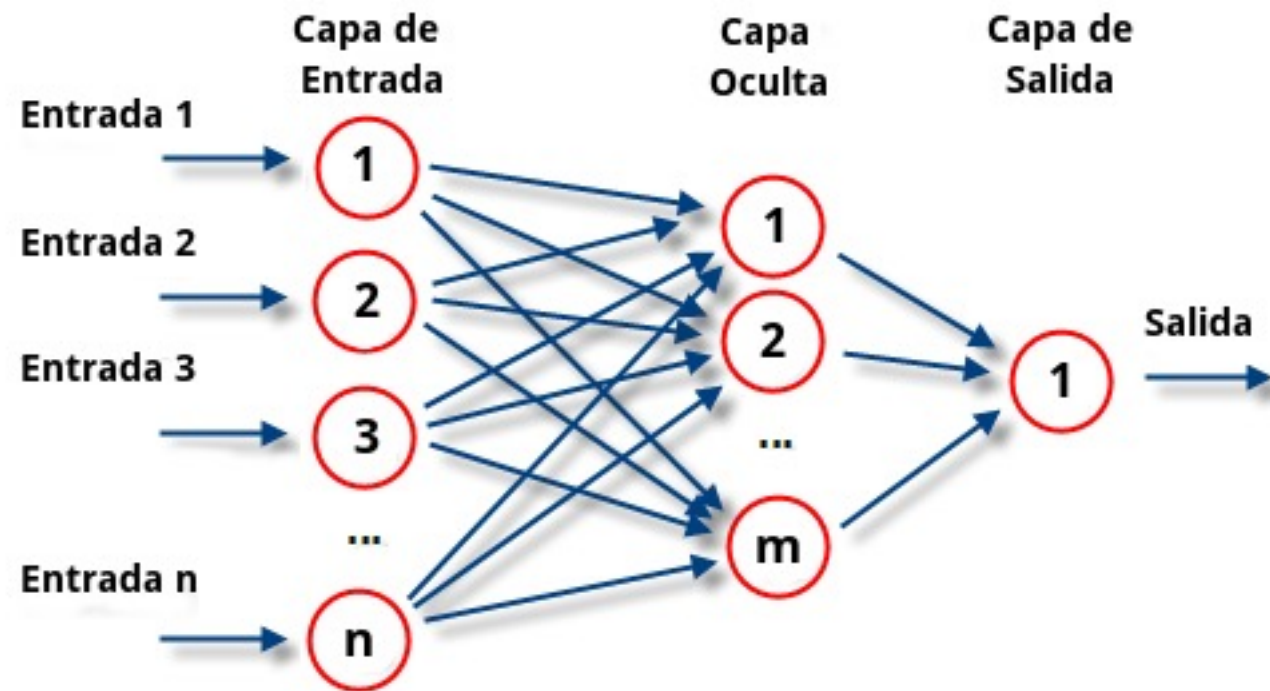
Redes neuronales

Motivación



Redes neuronales

Motivación



Ejercicio (EXTRA!)

Hora de Notebooks!!

De NLP y redes neuronales 😊



Contacto

Correo: a.cobo.aguilera@gmail.com

LinkedIn: [Aurora Cobo Aguilera](#)

GitHub: [AuroraCoboAguilera](#)

Google Scholar: [Aurora Cobo Aguilera](#)





red.es



"El FSE invierte en tu futuro"

Fondo Social Europeo

