





Introducción a la extracción de conocimiento a partir de Big Data

17 de febrero 2021







Preguntas a responder hoy

- ¿Qué es el Big Data?
- ¿Qué es un científico de datos?
- ¿Cómo se extrae conocimiento a partir de los datos?
- ¿Cómo aprende un ordenador?
- ¿Qué es el aprendizaje automático?
- ¿Cuáles son los componentes de este tipo de sistemas?
- ¿Qué aplicaciones tienen en la vida real?
- Y muchas más...



Contenidos

- Los datos son el nuevo petróleo
- Extracción de conocimiento
- Análisis de datos exploratorio
 - Estadísticas descriptivas
 - Visualización
- Aprendizaje automático
 - Datos
 - Características
 - Algoritmos

Contenidos

- Los datos son el nuevo petróleo
- Extracción de conocimiento
- Análisis de datos exploratorio
 - Estadísticas descriptivas
 - Visualización
- Aprendizaje automático
 - Datos
 - Características
 - Algoritmos

Vivimos en un mundo inundado de datos

- Páginas webs que monitorizan cada clic de sus usuarios
- Teléfonos móviles acumulando registros de ubicación
- Vehículos inteligentes recopilando hábitos de conducción
- Hogares inteligentes recopilando hábitos de vida
- Tiendas online recopilando hábitos de compra
- Todo tipo de estadísticas gubernamentales
- Internet es una enorme enciclopedia
- ► El Internet de las cosas (Internet of Things IoT)
- ▶ El yo cuantificado (Quantified self)
- ...

Enterradas en estos datos están las respuestas a incontables preguntas que nadie ha pensado ni siquiera hacer

Big Data

- Datos masivos
- Conjunto de datos cuya gestión resulta problemática con tecnologías y herramientas convencionales
 - Almacenamiento
 - Integración
 - Recuperación
 - Procesamiento
 - Análisis

Big Data

- Las tres Vs del Big Data
 - Volumen (tamaño)
 - Velocidad (crecimiento, rapidez de consumo...)
 - ▶ Variedad (estructurados, no estructurados ...)
- Las dos Vs extra del Big Data
 - Veracidad (confianza, autenticidad,...)
 - Valor (qué aporta a tu "negocio")



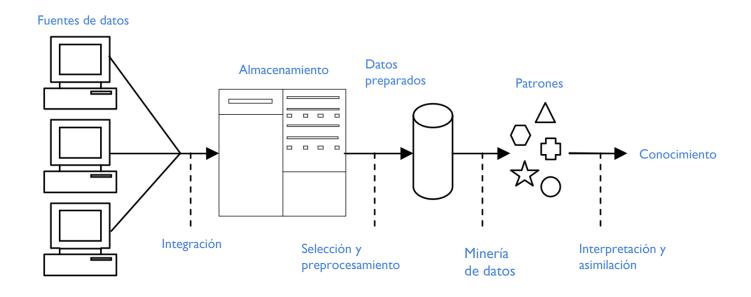
Big Data



Contenidos

- Los datos son el nuevo petróleo
- Extracción de conocimiento
- Análisis de datos exploratorio
 - Estadísticas descriptivas
 - Visualización
- Aprendizaje automático
 - Datos
 - Características
 - Algoritmos

- Extracción de conocimiento a partir de los datos
 - Knowledge Discovery in Databases (KDD)
 - "Extracción no trivial de información implícita, previamente desconocida y potencialmente útil a partir de datos"

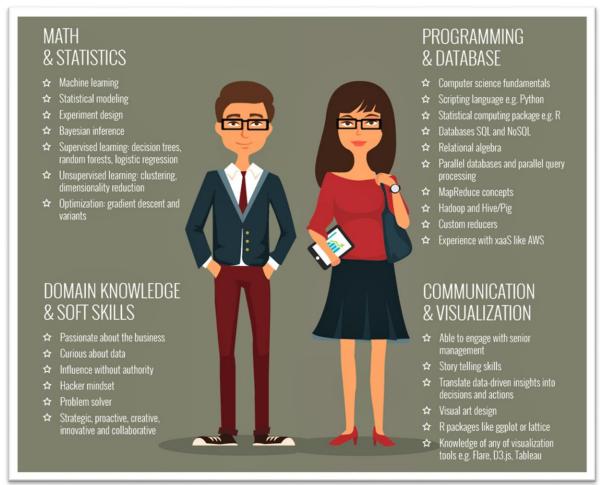


- ¿Quién se encarga de hacer este trabajo?
 - Un científico de datos es alguien que extrae conocimiento de los datos
 - Sabe más estadística que un informático y más informática que un estadístico

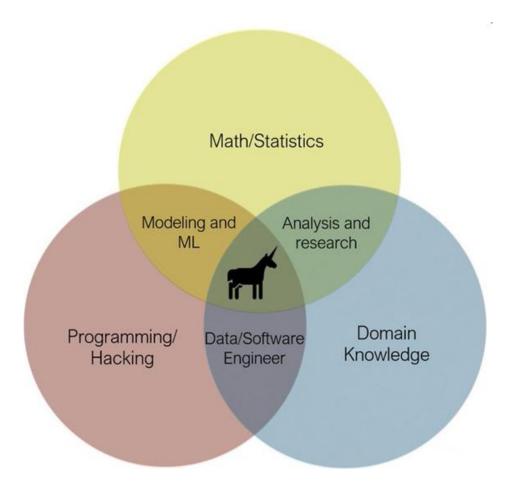
Data Scientist: The Sexiest Job of the 21st Century

by: Thomas H. Davenport and D.J. Patil

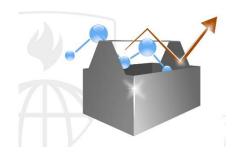
¿Cuáles son las cualidades de un Científico de Datos?



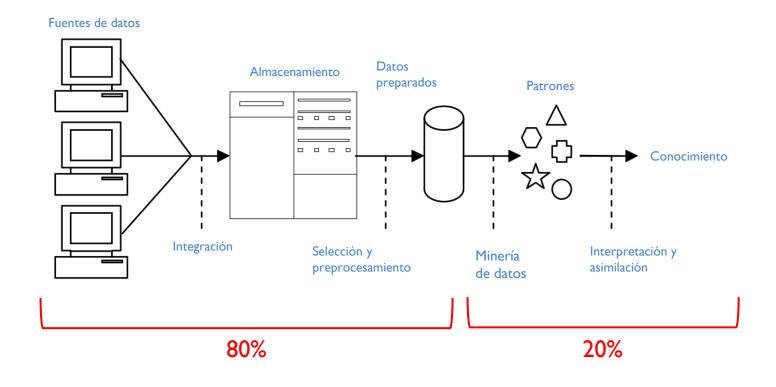
Les cualidades de un Científico de Datos?



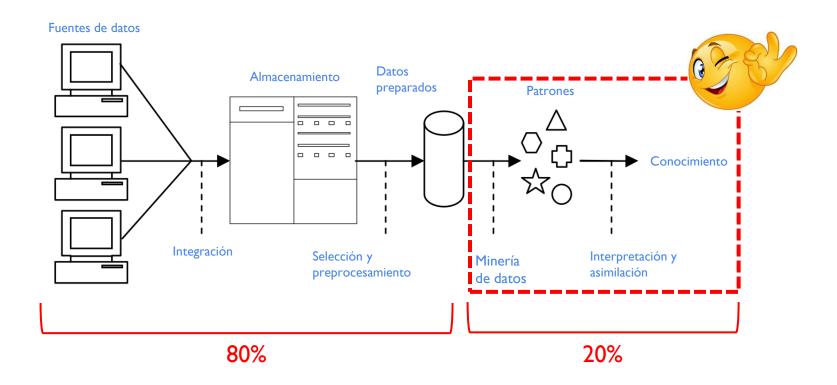
- La caja de herramientas del científico de datos
 - Python (tu lenguaje favorito de programación)
 - Google Colab (tu plataforma favorita de desarrollo)
 - Pandas (tu librería favorita para manipular tablas)
 - NumPy (tu librería favorita para operar con vectores)
 - Scikit-learn (tu librería favorita para machine learning)
 - Seaborn (tu librería favorita para visualización)
 - Kaggle (tu web favorita para juntarte con la gente guapa)



¿Cuánto tiempo se invierte en cada fase del proceso?



¿Cuánto tiempo se invierte en cada fase del proceso?



Minería de datos

- Data mining
- Proceso de descubrir patrones en grandes volúmenes de datos
- Analogía con un proceso de minería real
 - ▶ Partimos de un mineral (datos)
 - Llegamos al producto final refinado (conocimiento)
- Emplea métodos de aprendizaje automático, estadística y bases de datos



Minería de datos

- La extracción de conocimiento es un reto debido a la gran disparidad de problemas y tipos de datos existentes
- Un problema de recomendación de un producto comercial es muy diferente de una aplicación de detección de intrusos
 - Formato de los datos de entrada
 - Definición del problema
- Incluso dentro de tipos de problemas similares las diferencias son bastante significativas
 - Recomendación de un producto en una base de datos
 - Recomendación de contactos en una red social

Aplicaciones

- Analizar imágenes de satélite
- Análisis de compuestos orgánicos
- Detección de fraude en tarjetas de crédito
- Predicción de consumo eléctrico
- Diagnósticos médicos
- Valoración de inmuebles
- Marketing dirigido
- Pronóstico del tiempo
- Predecir audiencia de televisión
- ...

Ejemplos

- Facebook
 - Pregunta a los usuarios su ciudad natal y ubicación actual
 - Aparentemente el objetivo es facilitar que tus amigos te encuentren y se conecten contigo
 - También analiza estas ubicaciones para identificar los patrones de migración global y dónde viven los fans de los diferentes equipos de fútbol



Ejemplos

- Campaña de Obama 2012
 - Empleó a docenas de científicos de datos para identificar votantes que necesitaban atención extra, cuyo voto era más probable que fuera útil
 - Identificaron los programas óptimos de recaudación de fondos específicos para los donantes



Ejemplos

- OkCupid
 - Hace miles de preguntas a sus miembros (desde el cambio climático al cilantro) para encontrar las parejas más apropiadas
 - Analiza estos resultados para identificar preguntas "inocuas" que hacer para saber qué probabilidad hay de que alguien se acueste contigo en la primera cita



Un inciso



Google Colaboratory

- Entorno gratuito de Jupyter Notebook que se ejecuta en la nube de Google (acceso mediante cuenta de Gmail)
 - https://colab.research.google.com
- Jupyter es un entorno web interactivo que permite editar y ejecutar código Python
- Podemos ejecutar nuestro código en CPU y GPU en la nube
- Los cuadernos de Jupyter se guardan en Google Drive y se pueden compartir como cualquier otro documento
- Tiene algunas limitaciones:
 - Máquina inicial de 12GB de RAM y 100GB de disco duro
 - Tiempo máximo de ejecución: 12 horas
 - Si estamos más de 90 minutos sin usar el cuaderno se desconecta

Google Colaboratory

Accede a

Hazte tu propia copia en Drive

```
CO La Text Mining ☆
File Edit View Insert Runtime Tools Help

□ Table of contents 

Table of contents 

Text Mining ☆
File Edit View Insert Runtime Tools Help

Copy to Drive
```

Seguimos



Contenidos

- Los datos son el nuevo petróleo
- Extracción de conocimiento
- Análisis de datos exploratorio
 - Estadísticas descriptivas
 - Visualización
- Aprendizaje automático
 - Datos
 - Características
 - Algoritmos

Análisis de datos exploratorio

Definición

- Exploratory Data Analysis
- Técnicas para analizar datos mediante tratamiento estadístico
 - ▶ Importar, limpiar y validar
 - Visualizar distribuciones
 - Explorar relaciones entre variables
 - Selección de características
 - Identificación de valores extremos
 - **...**
- Se utiliza en la fase inicial de todo proyecto de ciencia de datos
- El objetivo es tener un conocimiento sólido de los datos

Análisis de datos exploratorio

Tipos de análisis

- Estadísticas descriptivas
 - Media
 - Mediana
 - ▶ Moda
 - Varianza
 - **...**
- Visualización
 - Histograma
 - Diagrama de dispersión
 - Diagrama de caja
 - Nubes de palabras
 - **...**

Contenidos

- Los datos son el nuevo petróleo
- Extracción de conocimiento
- Análisis de datos exploratorio
 - **Estadísticas descriptivas**
 - Visualización
- Aprendizaje automático
 - Datos
 - Características
 - Algoritmos

Definición

- Técnicas matemáticas para resumir o describir conjuntos de datos de manera cuantitativa
- ldentificar propiedades de los datos, ruido y valores extremos
- Medidas habituales usadas para describir los datos
 - Tendencia central
 - Dispersión

Tendencia central

- Las columnas pueden de tener miles de valores distintos
- Un paso básico al explorar los datos es obtener un valor típico para cada columna
- Tendencia central: estimación de dónde está localizada la mayoría de los datos
- Medidas habituales
 - Media
 - Mediana
 - ▶ Moda

Dispersión

- La tendencia central es una forma de resumir una variable
- Otra forma de hacerlo es mediante la dispersión (variabilidad),
 midiendo si los valores están agrupados o dispersos
- Es útil para identificar valores extremos (outliers)
- Medidas habituales
 - Rango
 - Cuantil
 - Desviación media
 - Varianza
 - Desviación estándar

¡Practiquemos!

https://bit.ly/37gxDa6

Contenidos

- Los datos son el nuevo petróleo
- Extracción de conocimiento
- Análisis de datos exploratorio
 - Estadísticas descriptivas
 - Visualización
- Aprendizaje automático
 - Datos
 - Características
 - Algoritmos

Definición

- Visualizar los datos permite resaltar sus principales características
- Formas de presentar los datos
 - Textual
 - ▶ Tabular
 - Gráfica
- La representación gráfica es atractiva y fácil de entender
 - Explorar los datos
 - Comunicar los datos (no solo a expertos)

- Las técnicas de visualización pueden proporcionar una respuesta rápida a muchas preguntas importantes
 - ¿Qué rango cubren las observaciones?
 - ¿Cuál es la tendencia central?
 - La distribución es simétrica o hay asimetría en alguna dirección?
 - Hay evidencia de bimodalidad?
 - ¿Hay valores extremos significativos?
 - ...

- Hay múltiples tipos de diagramas para distintos objetivos:
 - Diagrama de barras (ranking)
 - Histograma (distribución)
 - Diagrama de densidad (distribución)
 - Diagrama de líneas (evolución)
 - Diagrama de dispersión (correlación)
 - Mapa de calor (correlación)
 - Diagrama de caja (distribución y ranking)
 - Diagrama de enjambre (distribución)
 - Diagrama de violín (distribución)
 - Diagrama de árbol (todo-parte)
 - Nube de palabras (ranking)
 - ...

¡Practiquemos!

https://bit.ly/3priqJI

Contenidos

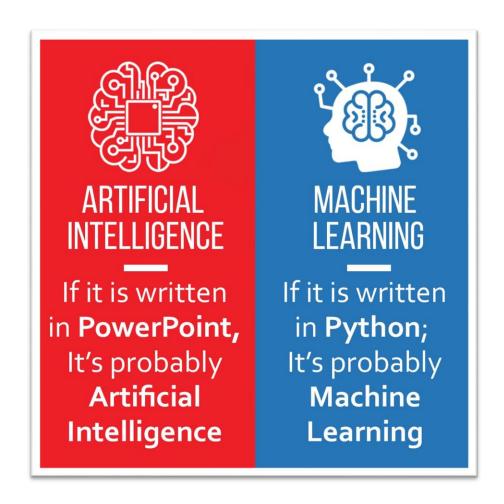
- Los datos son el nuevo petróleo
- Extracción de conocimiento
- Análisis de datos exploratorio
 - Estadísticas descriptivas
 - Visualización
- Aprendizaje automático
 - Datos
 - Características
 - Algoritmos

Aprendizaje automático

- ▶ También llamado machine learning
- Rama de la Inteligencia Artificial
- Objetivos
 - Desarrollar técnicas que permitan a los ordenadores aprender a partir de datos
 - Generalizar a partir de la experiencia (inducción) y construir un modelo

"Se dice que un programa de ordenador aprende a partir de una experiencia E con respecto a una tarea T y una medida de rendimiento P, si su rendimiento en la tarea T, medida usando P, mejora con la experiencia E", Tom M. Mitchell

Aprendizaje automático



Aprendizaje automático

- Componentes de un sistema de aprendizaje automático
 - Datos
 - ▶ Conjunto de muestras que se utilizan para entrenar/evaluar el sistema
 - Características
 - Atributos que representan a cada una de las muestras del conjunto de datos
 - Algoritmos
 - Deraciones que permiten aprender a partir de las características obtenidas de los datos de entrenamiento para generar un modelo

Contenidos

- Los datos son el nuevo petróleo
- Extracción de conocimiento
- Análisis de datos exploratorio
 - Estadísticas descriptivas
 - Visualización
- Aprendizaje automático
 - Datos
 - Características
 - Algoritmos

Datos

- Los datos son colecciones de objetos
 - Pacientes de un hospital
 - Clientes de una operadora telefónica
 - Viajes en tren de Barcelona a Madrid
 - Accesos a un servidor Web
 - Animales de un zoo
 - Pisos vendidos en una zona
 - ...
- Son el combustible de los sistemas de aprendizaje automático
- Al conjunto de datos disponible en una aplicación se le llama dataset

Datos

- Hay que invertir mucho esfuerzo en garantizar que los datos son de calidad
 - Obtener un conjunto amplio
 - Que sea representativo
 - Eliminar falsas observaciones
 - Limpiar
 - Dar formato
 - ...
- No importa lo sofisticados que sean los algoritmos si los datos no son adecuados

Datos

We don't have better algorithms.
We just have more data.



Peter Norving (Google Inc.)

Contenidos

- Los datos son el nuevo petróleo
- Extracción de conocimiento
- Análisis de datos exploratorio
 - Estadísticas descriptivas
 - Visualización
- Aprendizaje automático
 - Datos
 - Características
 - Algoritmos

Características

- Cada dato (objeto) es descrito por un número de características/atributos (features) que representan sus propiedades
 - Ej: para una persona: color de ojos, altura, peso, edad, ...
- Dos tipos fundamentales de características
 - Discretas
 - Contienen etiquetas que representan categorías
 - Ej: color de un objeto, código postal, aprobado/suspendido, ...
 - Continuas
 - Toman valores numéricos
 - Ej: número de hijos, altura, edad, peso, ...

Características

- Un conjunto de datos se representa habitualmente como una tabla o una serie de vectores de características
 - Cada columna es una característica
 - Cada fila es una instancia (objeto)

Características

Id Reembolso **Estado Civil** Salario Fraude Sí Soltero 125.000€ No Instancias No Casado 100.000€ No No Soltero 70.000€ No No 95.000€ Sí Divorciado

Características

- Datos etiquetados (labelled)
 - Hay una característica especial para cada instancia llamada clase clase

ld	Reembolso	Estado Civil	Salario	Fraude
I	Sí	Soltero	125.000€	No
2	No	Casado	100.000€	No
3	No	Soltero	70.000€	No
• • •	•••	•••		

- Datos no etiquetados (unlabelled)
 - No hay una clase definida

Contenidos

- Los datos son el nuevo petróleo
- Extracción de conocimiento
- Análisis de datos exploratorio
 - Estadísticas descriptivas
 - Visualización
- Aprendizaje automático
 - Datos
 - Características
 - Algoritmos

- Aprendizaje supervisado (métodos predictivos)
 - Utilizan datos etiquetados
 - Usar algunas variables para predecir valores futuros o no conocidos de otras variables
 - Clasificación
 - Regresión
- Aprendizaje no supervisado (métodos descriptivos)
 - Utilizan datos no etiquetados
 - Encontrar patrones interpretables por un humano que describan los datos
 - Agrupamiento
 - Reglas de asociación

Clasificación

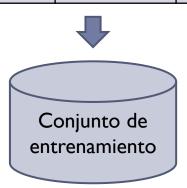
- Trabaja con datos etiquetados
- El atributo clase es de tipo discreto
 - Contiene un conjunto de etiquetas como posibles valores
 - ▶ **Ej**:{positivo, negativo, neutro},{hombre, mujer}, ...
- El objetivo es predecir el valor del atributo *clase* para instancias no vistas con anterioridad
- Debemos encontrar un modelo para el atributo *clase* en función de los valores de los otros atributos

Clasificación

- Conjunto de entrenamiento (training set)
 - Colección de instancias inicial, con sus clases ya asignadas
 - Se utiliza para construir el modelo
- Conjunto de validación (validation set)
 - Colección de instancias, con sus clases ya asignadas
 - Se utiliza para ajustar los parámetros del modelo y seleccionar la mejor configuración
 - No siempre es necesario
- Conjunto de evaluación (test set)
 - Colección de instancias, con sus clases ya asignadas
 - Se utiliza para validar el modelo, comparando las clases preasignadas con las que el modelo produce

Re.	Estado	Salario	Fraude
Sí	Soltero	125.000€	No
No	Casado	100.000€	No
No	Soltero	70.000€	No
Sí	Casado	120.000€	No
No	Divorciado	95.000€	Sí
No	Casado	60.000€	No
•••	•••	•••	•••

Re.	Estado	Salario	Fraude
No	Soltero	75.000€	?
Sí	Casado	50.000€	?
No	Casado	150.000€	?
Sí	Divorciado	90.000€	?
•••	•••	•••	•••



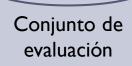


Re.	Estado	Salario	Fraude
Sí	Soltero	125.000€	No
No	Casado	100.000€	No
No	Soltero	70.000€	No
Sí	Casado	120.000€	No
No	Divorciado	95.000€	Sí
No	Casado	60.000€	No
•••	•••	•••	



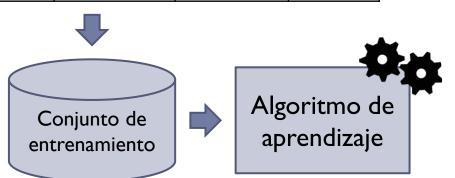


Re.	Estado	Salario	Fraude
No	Soltero	75.000€	?
Sí	Casado	50.000€	?
No	Casado	150.000€	?
Sí	Divorciado	90.000€	?
•••	•••		•••



Re.	Estado	Salario	Fraude
Sí	Soltero	125.000€	No
No	Casado	100.000€	No
No	Soltero	70.000€	No
Sí	Casado	120.000€	No
No	Divorciado	95.000€	Sí
No	Casado	60.000€	No
•••	•••	•••	•••

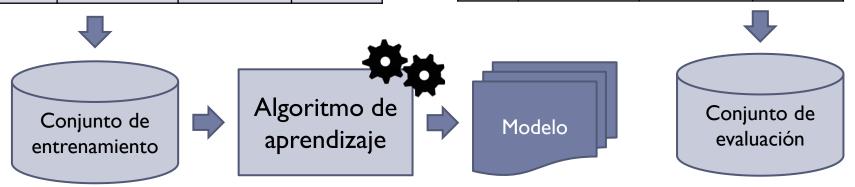
Re.	Estado	Salario	Fraude
No	Soltero	75.000€	?
Sí	Casado	50.000€	?
No	Casado	150.000€	?
Sí	Divorciado	90.000€	?
• • •	•••	•••	•••





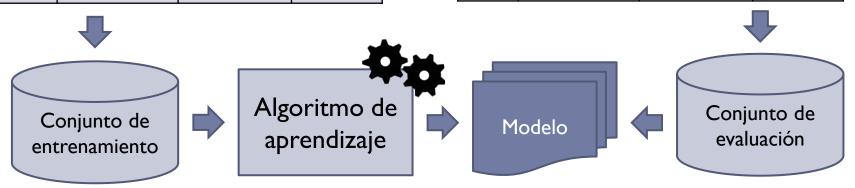
Re.	Estado	Salario	Fraude
Sí	Soltero	125.000€	No
No	Casado	100.000€	No
No	Soltero	70.000€	No
Sí	Casado	120.000€	No
No	Divorciado	95.000€	Sí
No	Casado	60.000€	No
		•••	•••

Re.	Estado	Salario	Fraude
No	Soltero	75.000€	?
Sí	Casado	50.000€	?
No	Casado	150.000€	?
Sí	Divorciado	90.000€	?
•••	•••	•••	•••



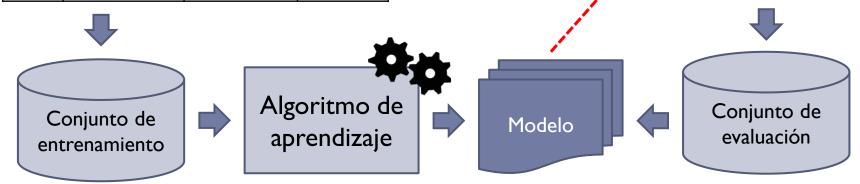
Re.	Estado	Salario	Fraude
Sí	Soltero	125.000€	No
No	Casado	100.000€	No
No	Soltero	70.000€	No
Sí	Casado	120.000€	No
No	Divorciado	95.000€	Sí
No	Casado	60.000€	No
•••	•••	•••	•••

Re.	Estado	Salario	Fraude
No	Soltero	75.000€	?
Sí	Casado	50.000€	?
No	Casado	150.000€	?
Sí	Divorciado	90.000€	?
•••	•••	•••	•••



Re.	Estado	Salario	Fraude
Sí	Soltero	125.000€	No
No	Casado	100.000€	No
No	Soltero	70.000€	No
Sí	Casado	120.000€	No
No	Divorciado	95.000€	Sí
No	Casado	60.000€	No
•••	•••	•••	•••

Re.	Estado	Salario	Fraude
No	Soltero	75.000€	No
Sí	Casado	50.000€	Sí
No	Casado	150.000€	No
Sí	Divorciado	90.000€	No
•••	, , , , ,	•••	•••



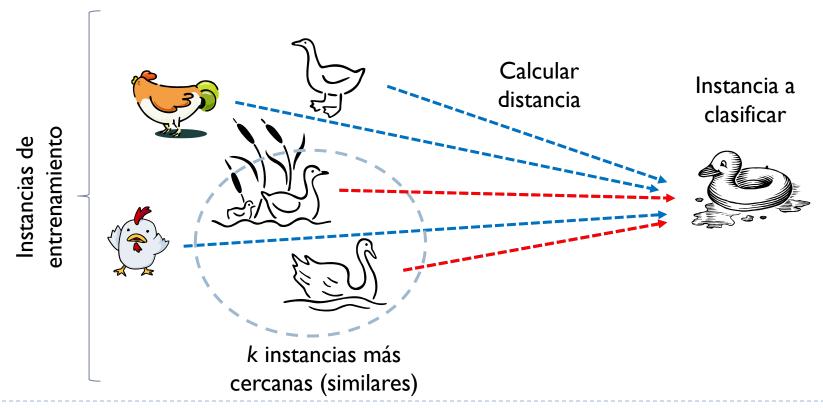
- Clasificación
 - Ejemplo de aplicación
 - Marketing directo
 - □ Objetivo
 - □ Reducir el coste de envío de correo seleccionando un conjunto de clientes que sean candidatos a comprar un nuevo modelo de teléfono móvil
 - □ Aproximación
 - ☐ Usar datos de un producto similar existente anteriormente
 - □ Sabemos qué cliente lo compró y quién no
 - □ La decisión {comprará, no_comprará} constituye el atributo clase que queremos predecir
 - □ Recolectar información demográfica, de estilo de vida, tipo de negocio, salario, etc. para cada cliente potencial
 - ☐ Usar esa información como *características* de entrada para entrenar el clasificador

- Clasificación
 - Ejemplo de aplicación
 - Fidelización de clientes
 - □ Objetivo
 - □ Predecir cuándo una compañía puede perder un cliente
 - □ Aproximación
 - □ Usar instancias de transacciones de clientes pasados y presentes
 - Con qué frecuencia llama el cliente, dónde llama, a qué hora del día, situación económica, estado civil, etc.
 - □ Etiquetar a los clientes como {leal, desleal} (ésta será la clase)
 - ☐ Encontrar un modelo para predecir la lealtad de los clientes

Clasificación

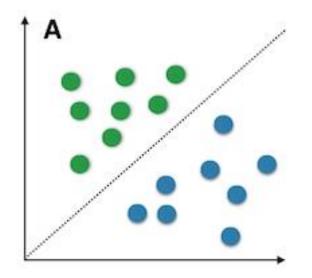
- Existen numerosos algoritmos de clasificación
- Algunos funcionan mejor para determinadas tareas
 - □ Dependiendo del número de instancias
 - □ Dependiendo del número de características
- Tipos
 - □ Árboles de decisión
 - □ Razonamiento basado en ejemplos
 - □ Redes neuronales
 - □ Bayesianos
 - □ Separadores lineales
 - □ ...

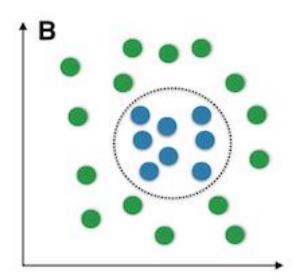
- k-Nearest Neighbors (k-NN)
 - Idea intuitiva: "si anda como una pato y grazna como un pato, probablemente sea un pato"



- Deep Neural Networks (Redes neuronales profundas)
 - > Separan las muestras en un espacio multidimensional

Linear vs. nonlinear problems





Regresión

- Trabaja con datos etiquetados (como la clasificación)
- El atributo *clase* es de tipo continuo
 - Contiene un conjunto de valores numéricos
 - Ej: precio estimado de una casa, de una acción,
- Debemos encontrar un modelo para el atributo *clase* en función de los valores de los otros atributos
- El objetivo es predecir el valor del atributo continuo *cla*se para instancias no vistas con anterioridad

Regresión

- Emplea los mismos conjuntos de datos que para la clasificación
 - Conjunto de entrenamiento
 - Conjunto de validación
 - Conjunto de evaluación
- Algoritmos
 - Perceptrón multicapa
 - ▶ k-NN
 - Máquinas de Vectores de Soporte (SVM)
 - Árboles de decisión (M5P)
 - **...**

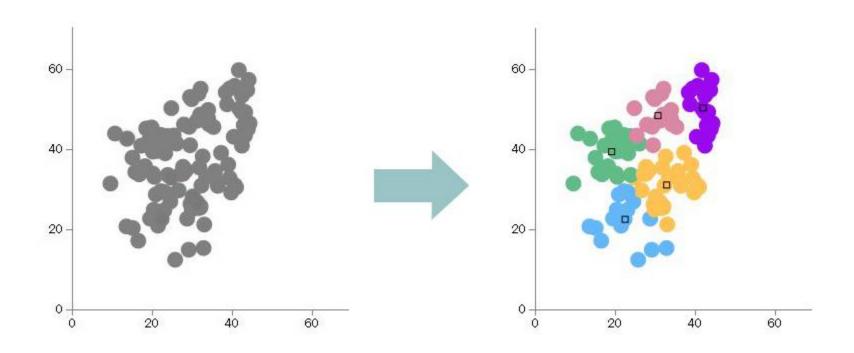
Regresión

- Ejemplos de aplicación
 - Predecir el número de ventas de un nuevo producto basado en los gastos de publicidad
 - Predecir la velocidad del viento como una función de la temperatura, humedad, presión del aire, etc.
 - Predicción de series temporales en índices del mercado de valores

Agrupamiento

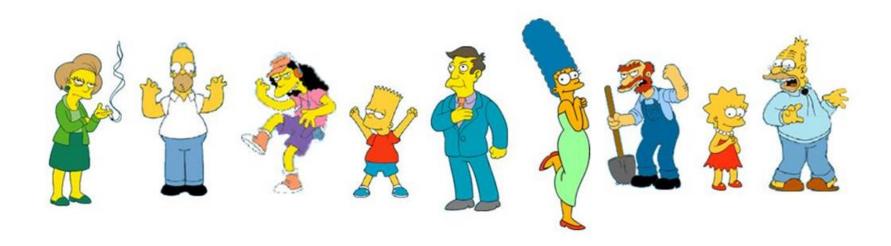
- También conocido como clustering
- Trabaja con datos no etiquetados
- El objetivo es encontrar agrupamientos (clusters) en los datos (instancias) de manera que
 - Las instancias de un cluster se parecen más entre ellas
 - Las instancias en distintos *clusters* se parecen menos entre ellas
- Es necesaria una medida de similitud entre instancias
 - Ej: la distancia euclídea si los atributos son numéricos

Agrupamiento

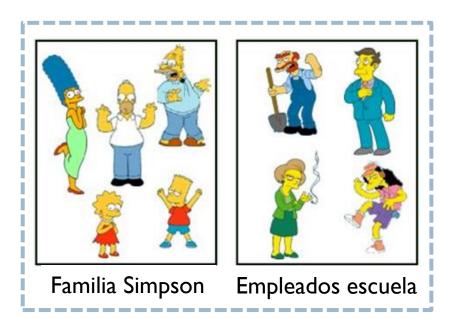


- Agrupamiento
 - Ejemplo de aplicación
 - Segmentación del mercado
 - □ Objetivo
 - □ Dividir el mercado en distintos subconjuntos de clientes
 - Cualquier subconjunto puede ser seleccionado como mercado objetivo a alcanzar con una aproximación de marketing distinta
 - □ Aproximación
 - □ Recolectar atributos de los clientes basados en su información geográfica y de estilo de vida
 - ☐ Encontrar clusters de clientes similares
 - Medir la calidad del cluster observando los patrones de compra de clientes en el mismo cluster con respecto a otros en distintos clusters

- Agrupamiento
 - ¿Cuál es el agrupamiento natural de estos objetos?



- Agrupamiento
 - ¿Cuál es el agrupamiento natural de estos objetos?
 - Agrupar es una tarea subjetiva
 - Las características y las medidas de similitud son importantes





- Agrupamiento
 - Algoritmos
 - ▶ K-means
 - Expectation Maximisation (EM)
 - Cobweb
 - **...**
 - En algunos algoritmos es necesario definir el número de clusters

Reglas de asociación

- Trabaja con datos no etiquetados (como el clustering)
- Dbjetivo: obtener unas reglas de dependencia para predecir la ocurrencia de un ítem basado en la ocurrencia de otros ítems
- Uso habitual para análisis de afinidad (market basket analysis)
 - Si conocemos las compras hechas por todos los clientes durante un periodo, podemos encontrar relaciones entre dichos productos

```
SI queso Y leche ENTONCES pan (probabilidad = 0,7)
```

Reglas de asociación

 Partimos de un conjunto de instancias, cada una con una serie de elementos de una colección

Ĺ	2
2.	ರ
٠,	╮
2	2
2	=
400	ত্
•	ζ.
7	"
	=
_	
C	ָ כ
τ	ׄ כ
	_
(2
+	ر
2	=
=	3
.=	_
2	=
()
ı`	`
	J

Id	Ítems
- 1	pan, refresco, leche
2	cerveza, pan
3	cerveza, refresco, pañales, leche
4	cerveza, pan, pañales, leche
5	refresco, pañales, leche





 ${leche}$ → ${refresco}$ ${pañales, leche}$ → ${cerveza}$

- Reglas de asociación
 - Ejemplos de aplicación
 - Gestión de un supermercado
 - □ Objetivo
 - ☐ Identificar los ítems que se suelen compran juntos por los clientes
 - □ Aproximación
 - □ Procesar los datos de punto de venta recolectados en las cajas para encontrar dependencias entre los ítems
 - □ Regla clásica
 - La parábola de los pañales y la cerveza
 - Si un cliente compra pañales y leche, es muy probable que compre cerveza



- Reglas de asociación
 - Ejemplos de aplicación
 - Manejo de inventario
 - □ Objetivo
 - □ Una compañía de reparaciones quiere anticipar la naturaleza de las reparaciones de sus productos
 - Mantener los vehículos de servicio equipados con los componentes adecuados para reducir el número de visitas a un domicilio
 - □ Aproximación
 - □ Procesar los datos sobre herramientas y componentes necesarios en reparaciones previas en diferentes localizaciones de consumidores
 - □ Descubrir la coocurrencia de patrones

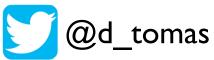
¡Practiquemos!

https://bit.ly/20GLal6

... y para saber más

- Webs
 - ▶ Towards Data Science
 - KDnuggets
 - Kaggle competitions
 - Coursera
- YouTube
 - Dot CSV
- Cursos (¡Ojo! Publicidad encubierta)
 - "Big Data: fundamentos tecnológicos y aplicaciones prácticas", curso de verano de la UA
 - "Máster Universitario en Ciencia de Datos", título de la UA





David Tomás Díaz

