

# Introducción al Business Analytics

David Masip Rodó



# Business Analytics

## Definición

- “La constante e iterativa exploración y estudio del rendimiento pasado del negocio, con el objetivo de ganar conocimiento útil para la consecución de los objetivos estratégicos de la organización.”
- Aplicación de los campos:
  - Estadística
  - Inteligencia artificial

# Aprendizaje supervisado vs aprendizaje no supervisado

## Modelos descriptivos

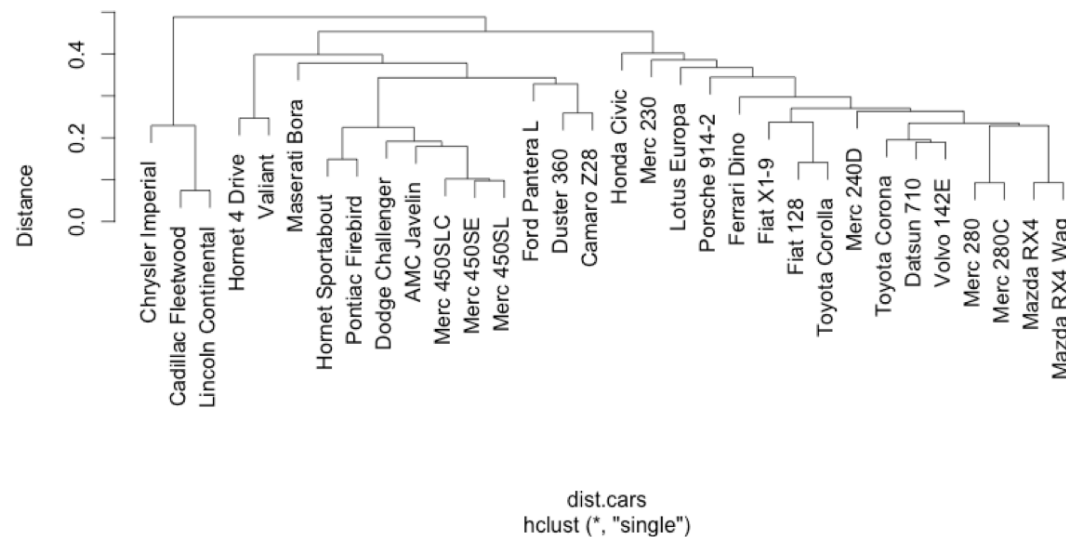
- Dado un conjunto de variables sobre unos datos de una empresa, pretendemos buscar en ellos características relevantes o segmentar los datos en grupos (o clustering).

## Modelos predictivos

- Dado un conjunto de variables sobre unos datos, pretendemos aquí predecir una variable especial (etiqueta) que depende de alguna forma del conjunto de datos iniciales. El objetivo es “**aprender**” un modelo que nos permita adivinar la etiqueta a partir de las relaciones encontradas en el pasado

# Clustering jerárquico

Se agrupan sucesivamente los datos uniendo progresivamente los ejemplos en clusters, que a su vez se unen entre si hasta definir dos subconjuntos



# Clustering no jerárquico

## El algoritmo k-means

- Proceso iterativo que permite encontrar un determinado número de subconjuntos de datos que minimizan la suma de distancias a sus centroides

### Centroides

Como su nombre indica, el algoritmo k-means utiliza como representante  $\mu_i$  de cada grupo  $S_i$  su centro geométrico (como puntos de un espacio  $d$ -dimensional):

$$\mu_i = \frac{1}{|S_i|} \left( \sum_{\mathbf{x}_j \in S_i} \mathbf{x}_j \right)$$

pero podría utilizarse cualquier otro elemento como representante, siempre que proporcione una solución estable al problema de particionar el conjunto original.

# Análisis de Componentes Principales

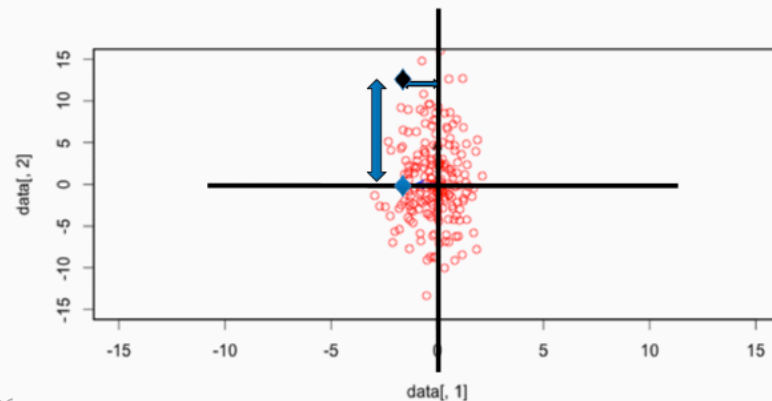
## Extracción de características

- Obtener un menor número de atributos que representan globalmente la información presente en el conjunto de datos original

### Análisis de componentes principales

#### Minimización del error

- Proyección lineal en la que se minimiza el error cuadrático de los datos reconstruidos
- Maximización de la varianza en el espacio proyectado



EIMT.UOC.EDU

# Resumen

- En este módulo deberéis
  - Mirar los videos docentes.
  - Preguntar las dudas que pudieran surgir en el foro.
  - Practicar (opcionalmente) los algoritmos explicados mediante el lenguaje R.
  - Realizar los ejercicios de evaluación.
- Otros recursos
  - Mas información sobre datamining en R
    - <http://data-mining.business-intelligence.uoc.edu/>