



# Aprendizaje Automático I

## Ejercicio: Componentes Principales

DSL3

noviembre, 2023

En este ejercicio vamos a trabajar con el conjunto de datos (Boston Housing Price) de la librería `mlbench`. El conjunto de datos consta de 506 observaciones de 14 atributos. El valor medio del precio de la vivienda en miles de dólares, denominado MEDV, es la variable de interés. A continuación se ofrece una breve descripción de cada característica:

- CRIM - tasa de delincuencia per cápita por ciudad
- ZN: proporción de suelo residencial con parcelas de más de 25.000 pies cuadrados.
- INDUS - proporción de acres comerciales no minoristas por ciudad
- CHAS - Variable ficticia del río Charles (1 si la zona linda con el río; 0 en caso contrario)
- NOX - concentración de óxidos nítricos (partes por 10 millones)
- RM - número medio de habitaciones por vivienda
- AGE - proporción de unidades ocupadas por sus propietarios construidas antes de 1940
- DIS - distancias ponderadas a cinco centros de empleo de Boston
- RAD - índice de accesibilidad a autopistas radiales
- TAX - tipo del impuesto sobre bienes inmuebles por cada 10.000 dólares
- PTRATIO - ratio alumnos-profesor por ciudad
- B -  $1000(B_k - 0,63)^2$  donde  $B_k$  es la proporción de negros por ciudad
- LSTAT - % de estatus inferior de la población
- MEDV - Valor medio de las viviendas ocupadas por sus propietarios en miles de \$.

1. Carga los datos y explora su contenido.
2. Realiza un Análisis de Componentes Principales.
3. Estudia el efecto de escalar, o no, las variables.

4. ¿Cuántas PCs deberías de conservar?
5. Realiza e interpreta un biplot. ¿Puedes averiguar dónde se sitúan las casas más caras?
6. Busca una interpretación a las dos primeras componentes principales.