

## Práctica 2: Exploración de datos

Realice los siguientes ejercicios usando el módulo Pandas y cualquier otro módulo adicional que considere:

1. 1.- Obtenga tres ejemplos de ficheros de datos en formato CSV, ARFF u otro cualquiera de:
  - [Weka datasets](#)
  - [UCI MLR](#)
2. Usando Pandas cargue los ficheros y evalúe qué información puede obtener del histograma de atributos.
3. Estudie el efecto de la normalización (reescalar en el intervalo  $[0, 1]$ ) y la estandarización ( $\mu = 0$ ,  $\sigma = 1$ ) sobre el histograma.
4. (\*) Estudie el efecto del análisis en componentes principales (`sklearn.decomposition.PCA`) sobre el histograma.
5. Usando la visualización del diagrama de dispersión (scatter plot) estudie qué información puede obtener de dicha representación gráfica.
6. Estudie el efecto de la normalización y la estandarización sobre el diagrama de dispersión.
7. (\*) Estudie el efecto del análisis en componentes principales sobre el diagrama de dispersión.
8. Estudie el diagrama de correlaciones de los tres conjuntos e indique qué información relativa a las diferentes clases puede obtener.
9. Estudie la representación en coordenadas paralelas de los tres conjuntos e indique qué información relativa a las diferentes clases puede obtener.

NOTAS:

- Los ejercicios marcados con un \* son opcionales.