# Unidad Temática 2 - Práctica Dirigida 1

## Ejercicio 2

# Estudio del dataset Wine de UCI

## 1. Descargar el dataset Wine de UCI

El dataset Wine de UCI se puede descargar de la siguiente URL: https://archive.ics.uci.edu/dataset/109/wine. El mismo viene en un archivo comprimido en formato ZIP, conteniendo un archivo Index, un archivo de datos y un archivo de descripción del dataset.

## 2. Analizar y describir el dataset

El dataset contiene 178 instancias, cada una de las cuales representa un vino. Cada instancia tiene 14 atributos, que son los siguientes:

1. Clase: 1, 2 o 3
2. Alcohol
3. Ácido málico
4. Ceniza
5. Alcalinidad de la ceniza
6. Magnesio
7. Fenoles totales
8. Flavonoides
9. Fenoles no flavonoides
10. Proantocianinas
11. Intensidad del color
12. Matiz
13. OD280/OD315 de vinos diluidos
14. Prolina

Los atributos 2 a 14 son numéricos, y representan distintas características de los vinos.

Analizando los datos, se puede ver que no hay valores faltantes en el dataset. Además, todos los atributos son numéricos, por lo que no es necesario realizar ninguna transformación de datos. Por último, los atributos solo contienen un valor atípico en el "Color Intensity" que arroja un número extremadamente alto, por lo que se decide eliminarlo.

## Modelado del dataset

Para probar la normalización de los datos, se utiliza el Algoritmo Naive Bayes, que es un algoritmo de clasificación probabilístico. El mismo se utiliza para predecir la clase de una instancia, dada la probabilidad de

que pertenezca a cada una de las clases. Para esto se tomaron 2 caminos, uno con los datos sin normalizar y otro con los datos normalizados.

Se observo como resultado que el algoritmo Naive Bayes funciona mejor con los datos normalizados, ya que la precisión de la clasificación es mayor. Esto se debe a que los datos normalizados tienen una distribución más uniforme, por lo que el algoritmo puede clasificar mejor las instancias.