



## Aprendizaje automático (Parte 2)

### Clasificación no supervisada y supervisada mediante el enfoque geométrico

#### Actividad 1.

1. Descargar y revisar los datos contenidos en el archivo wineData. Leer la descripción de los datos para entender a qué problema corresponden.
2. Escribir un script que me permita cargar los datos e imprimirlos.
3. Combinar diagrama de puntos (*scatter*) que utilizando pares de características, me permitan visualizar los datos en el plano cartesiano. Probar al menos 3 combinaciones y determinar las que permiten discriminar mejor las clases
4. Aplique el algoritmo **kmeans** y determine que tan preciso es el agrupamiento que realiza dicho algoritmo frente a la clasificación verdadera, calculando el error en la predicción.

#### Actividad 2.

1. Utilizar las herramientas suministradas por **Pandas** para hacer un análisis descriptivo de los datos incluyendo datos estadísticos y gráficas que permitan la descripción de los mismos.
2. Aplicar las utilidades suministradas con la librería **ScikitLearn** para determinar la eficacia del clasificador del vecino más cercano (**knn**) implementado en el problema de clasificación de vinos tratado en el caso anterior. Aplicar validación cruzada con división en 5 conjuntos.

Escriba un breve informe en Word, incluya una descripción de los datos utilizados junto a la explicación de cada uno de sus experimentos. Realice además un análisis de resultados, explique los resultados obtenidos en la matriz de confusión. Adjunte el código escrito junto al archivo de Word.