

MÓDULO 8: Modelización predictiva

Ejercicio 1: modelo supervisado por k vecinos

Nota preliminar: realizar este trabajo **por parejas**.

- 1) Partiendo de los ejemplos de código Python empleados en clase, elige por favor uno de los siguientes conjuntos de datos y describe qué representan sus campos:

A) “from sklearn.datasets import load_breast_cancer” (básico)
B) “from sklearn.datasets import load_digits” (avanzado)
- 2) Describe el dataset en dimensiones como en número de características, número de categorías y número de samples por categoría **utilizando Python**.
- 3) Repite cada uno de los experimentos vistos en clase, razonando cada uno de los pasos y comentando los resultados parciales obtenidos con el dataset seleccionado.
- 4) Consulta la documentación de la librería Scikit-learn y configura el script de validación automatizada (#3) para que la función `GridSearchCV()` utilice validación leave-one-out en lugar de k-fold. Describe qué conclusiones pueden extraerse a partir de los scores medios de cross-validación y el de test utilizando un modelo por k-vecinos con k optimizada.
- 5) Razona si es necesario estratificar el proceso de cross-validación analizando la distribución de muestras por clase.
- 6) Introduce en el proceso de cross-validación el ajuste de los pesos de la métrica de distancia entre muestras de acuerdo al parámetro “weights” del modelo en scikit-learn:

[kNN en Scikit Learn](#)

Calcula y computa la ganancia/pérdida en desempeño del modelo cuando los pesos de la métrica de distancia es afinada dentro de la validación cruzada respecto al caso visto en clase (ajuste único de K)

- 7) Siguiendo la misma aproximación del último apartado, introduce el tipo de métrica de distancia (parámetro “metric”) dentro del proceso de validación cruzada. Evalúa los resultados y las ganancias/pérdidas de capacidad de generalización del modelo.
- 8) PARA NOTA: repite los apartados 1-7 con un dataset a tu elección del repositorio UCI:

[UCI repository](#)

(elige un dataset de **clasificación** con un número “contenido” de muestras)

Fecha de entrega: a fijar según progreso en la asignatura