

## Trabajo Práctico 2: Selección de Variables

### Introducción

Pondremos en práctica los métodos de selección de variables que discutimos, utilizándolos para seleccionar variables en algunos datasets, y analizando los resultados obtenidos para comprender si la selección fue exitosa o no. Se espera que el alumno intente dar una explicación, sin demostrar, del por qué de los resultados obtenidos.

Se pide entregar un notebook con código funcionando y con comentarios y análisis correspondientes a cada resultado obtenido.

### Ejercicio 1

Implemente en Python 3 los siguientes métodos:

- a) Wrapper greedy backward
- b) Filter con test no-paramétrico (Kruskal-Wallis)
- c) Eliminación recursiva de características (RFE, por sus siglas en inglés)

Se puede usar como base el código de ejemplo del wrapper greedy forward proporcionado. Las implementaciones tienen que hacer uso de todas las funciones útiles necesarias de sklearn, pero no las versiones de estos tres métodos que están en la librería. Los asistentes de programación (chatGPT, copilot, etc. ) y repositorios públicos son válidos para consultar sobre el lenguaje de programación, no sobre los métodos que tienen que implementar.

### Ejercicio 2

En el código proporcionado se crean dos datasets artificiales (datosA y datosB). Analice cada dataset usando proyecciones que den información de las variables (por ejemplo un biplot de una PCA).

Aplique los 4 métodos de selección de variable (los 3 desarrollados más el wrapper greedy forward ya provisto) a los dos datasets para crear un ranking de todas las variables. Analice los resultados.

### Ejercicio 3

Prepare un dataset del problema diagonal del práctico 1 con 10 variables, 50 puntos por clase y  $C=2$ . Agregue 90 variables de ruido gaussiano con la misma dispersión que las relevantes. Aplique los 4 métodos de selección presentados anteriormente (los 3 desarrollados más el wrapper greedy forward ya provisto). Repita el experimento 30 veces (repetir todo: crear el dataset y elegir las variables).

## Trabajo Práctico 2: Selección de Variables

Calcule para cada método el porcentaje de aciertos en el ranking (cuántas de las 10 variables relevantes están en los primeros 10 lugares del ranking). Analizar los resultados.

### Ejercicio 4

Busque el dataset sonar<sup>1</sup> de la University of California, Irvine. Realice una selección de variables con RFE y el método de clasificación que considere adecuado, evaluando el error de clasificación resultante para cada cantidad de variables en el modelo (curva de error vs features). Repita el procedimiento al menos 5 veces, utilizando distintos conjuntos para medir el error.

### Ejercicio 5 (Opcional, 2 puntos)

Busque un dataset que considere interesante, muestre un biplot (PCA más variables originales), seleccione variables sobre el mismo con alguno de los métodos estudiados. Mida el error en función de la cantidad de variables en el modelo. Analice y explique los resultados obtenidos.

---

<sup>1</sup> <https://archive.ics.uci.edu/dataset/151/connectionist+bench+sonar+mines+vs+rocks>