

# Ajuste de Hiperparámetros en Modelos de Aprendizaje de Máquina con Algoritmos Bioinspirados

Módulo 4 : Técnicas computacionales avanzadas para modelar fenómenos sociales  
Concentración en Economía Aplicada y Ciencia de Datos  
ITESM

27 de abril de 2023



# Ajuste de Hiperparámetros en Modelos de Aprendizaje de Máquina

Algunos modelos de aprendizaje de máquina necesitan definir algunos **parámetros** antes de llevar a cabo el aprendizaje. A estos parámetros se les conoce como **hiperparámetros** e influyen en la forma como se realiza el aprendizaje.

Por ejemplo, para el caso de Árboles de Decisión, algunos hiperparámetros son:

- La profundidad máxima.
- El método para medir la calidad de la división (índice de gini, entropía, log loss).
- La estrategia utilizada para elegir la división en cada nodo (best o random).

# Ajuste de Hiperparámetros en Modelos de Aprendizaje de Máquina

- Dado que la elección de estos hiperparámetros tiene un impacto considerable en el desempeño de los modelos de aprendizaje de máquina, los científicos de datos dedican una cantidad considerable de tiempo en elegir la mejor combinación de hiperparámetros.
- Este proceso se le conoce como **ajuste de hiperparámetros**.

# Ajuste de hiperparámetros por búsqueda en Grid

- El método consiste en elegir un subconjunto de los valores posibles de cada hiperparámetro.
- En el caso de Árboles de Decisión, podemos elegir un subconjunto de 2, 5 y 10 para el parámetro de profundidad máxima, y usar todas las opciones disponibles para los otros dos hiperparámetros.
- En total, tendríamos  $3 \times 3 \times 2 = 18$  combinaciones posibles.

# Ajuste de hiperparámetros por búsqueda en Grid

- Para cada combinación posible, entrenamos el clasificador y lo evaluamos contra algún criterio de desempeño.
- Se elige la combinación de hiperparámetros que tiene el mejor desempeño.
- El principal inconveniente de este método es que la búsqueda exhaustiva sobre todas las combinaciones posibles que puede resultar muy tardado.

# Gradient Boosting Regression Trees

- *Boosting* es un método de ensamble (*ensemble method*) de aprendizaje que se basa en entrenar secuencialmente predictores con el fin de intentar corregir el desempeño de su predecesor ([Géron, 2017](#), cap 7).
- Los dos algoritmos de *boosting* más conocidos son el *AdaBoost* y el *Gradient Boosting*.
- Para el caso de *árboles de decisión*, el algoritmo es conocido como *Gradient Boosted Regression Trees* (GBRT).

# Gradient Boosting Regression Trees

- El algoritmo *Gradient Boosting* intenta ajustar el nuevo predictor al residuo del error producto del predictor previo.
- El problema central es encontrar la estructura óptima del árbol.

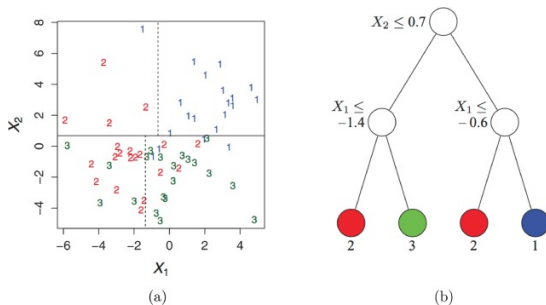


Figura: a) Partición y b) estructura del árbol de decisión (tomado de [Lamperti et al. \(2018\)](#))

El algoritmo se describe a continuación:

- Para  $b = 1, 2, \dots, B$

- ▶ Ajusta un árbol  $\hat{f}^b$  con  $d$  particiones ( $d + 1$ ) nodos terminales a los datos de entrenamiento  $(X, r)$ .
- ▶ Actualiza  $\hat{f}$  agregando una versión reducida del árbol nuevo:

$$\hat{f} \leftarrow \hat{f} + \lambda \hat{f}^b$$

- ▶ Actualiza los residuales

$$r_i \leftarrow r_i - \hat{f}^b$$

- El output del algoritmo es:

$$\hat{f}(x) = \sum_{i=1}^B \hat{f}^b(X)$$



- Un parámetro importante del GBRT es la *learning rate* ( $\lambda$ ), el cual es un hiperparámetro que escala la contribución de cada árbol.
- Hay una relación inversa entre la *learning rate* y el número de particiones( $B$ ).
- A una *learning rate* igual a 0.1, se necesitarán una cantidad grande de árboles para que el ensamble generalice mejor.
- Un número muy grande de  $B$  puede provocar sobreajuste. Una propuesta es usar *validación cruzada* para escoger  $B$ .

# References I

- Géron, A. (2017). Hands-on machine learning with scikit-learn and tensorflow: Concepts. *Tools, and Techniques to build intelligent systems*.
- Lamperti, F., Roventini, A., and Sani, A. (2018). Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control*, 90:366–389.