

Machine Learning para Inteligencia Artificial

Ensembles: Boosting

Universidad ORT Uruguay

7 de Mayo, 2025

Índice

Introducción a boosting

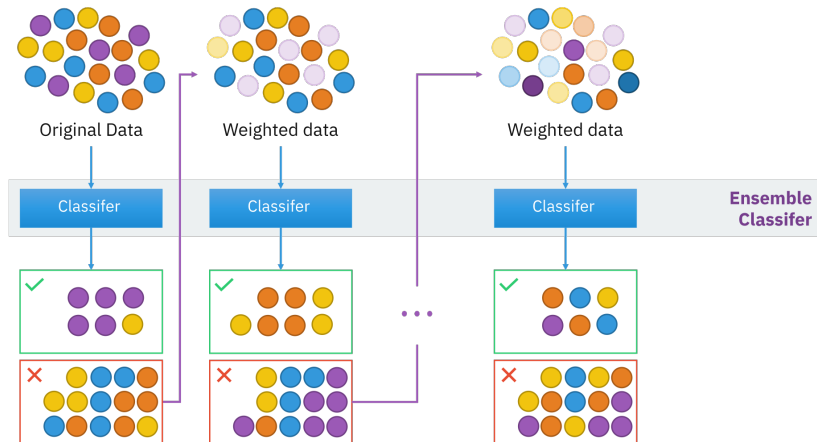
AdaBoost

Gradient Boosting

Ensemble: Boosting

- Apunta a **reducir el sesgo** en modelos con alto sesgo
- Boosting genera **secuencialmente** predictores
- **Corrigiendo errores** del modelo anterior en cada iteración
- La predicción final proviene de un **promedio** ponderado o **voto** ponderado.

Ensemble: Boosting



[https://en.wikipedia.org/wiki/Boosting_\(machine_learning\)](https://en.wikipedia.org/wiki/Boosting_(machine_learning))

Boosting a alto nivel

- El conjunto de entrenamiento T_1 es T
- El conjunto T_{k+1} se obtiene **modificando** T_k de acuerdo al error de h_k
- El predictor final se obtiene a partir de la suma **pesada** de los predictores

$$h(\mathbf{x}) = \text{FUNCIÓN} \left(\sum_{k=1}^K \alpha_k h_k(\mathbf{x}) \right)$$

- **Adaboost**: El coeficiente α_k representa el voto de h_k en el ensemble
- **Gradient boost**: $\alpha_k = \lambda$ son constantes (shrinkage/learning rate)

AdaBoost: Adaptive Boosting (etiquetas en $\{-1, 1\}$)

- Primera implementación exitosa de la idea de boosting (1995).
- Construye una secuencia de K clasificadores binarios (débiles)

$$A_1(\mathbf{x}), A_2(\mathbf{x}), \dots, A_K(\mathbf{x})$$

- Sólo usa la predicción final de los modelos base, y no sus probabilidades.
- Las predicciones individuales no son tratadas por igual:

$$A_{\text{boost}}(\mathbf{x}) = \text{Signo} \left\{ \sum_{k=1}^K \alpha_k A_k(\mathbf{x}) \right\}.$$

- El coeficiente α_k puede considerarse como un grado de confianza en las predicciones realizadas por el miembro k del ensemble.

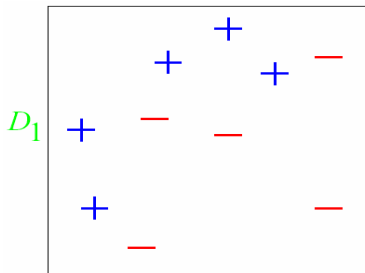
AdaBoost: Adaptive Boosting (binario $\{-1, 1\}$)

Algoritmo Adaboost

- 1: Sea K el número de rondas/iteraciones
- 2: Inicializar los pesos $w_i^{(1)} = 1/N$ para todo $i = 1, \dots, N$
- 3: Inicializar $A = 0$
- 4:
- 5: Para $k = 1$ a K repetir
 - 6: Para todo i : $w_i^{(k)} = w_i^{(k)} / \sum_j w_j^{(k)}$ Normalizar pesos
 - 7: $A_k = \text{EntrenarWeakLerner}(T, \mathbf{w}^{(k)})$
 - 8: Determinar voto de A_k en el ensemble:
 - 9: $\varepsilon_k = \sum_i w_i^{(k)} \mathbb{1}(A_k(\mathbf{x}_i) \neq y_i)$ Calcular error
 - 10: $\alpha_k = \frac{1}{2} \ln [(1 - \varepsilon_k) / \varepsilon_k]$ Calcular voto
 - 11: $w_i^{(k+1)} = w_i^{(k)} \times \begin{cases} e^{-\alpha_k} & \text{si } A_k(\mathbf{x}_i) = y_i \\ e^{\alpha_k} & \text{si } A_k(\mathbf{x}_i) \neq y_i \end{cases}$ Nuevos pesos
 - 12: Actualizar $A \leftarrow A + \alpha_k A_k$
 - 13: Devolver clasificador $\text{signo}(A)$

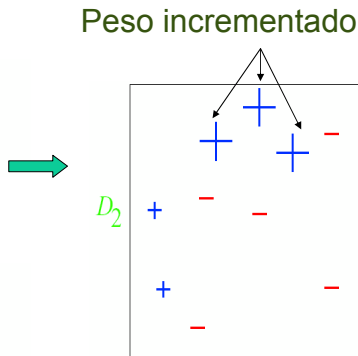
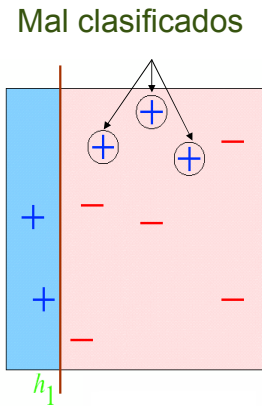
Ejemplo: Dataset inicial

Dado un dataset: $\{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbb{R}^2, y_i \in \{+1, -1\}, i \in [1, M]\}$

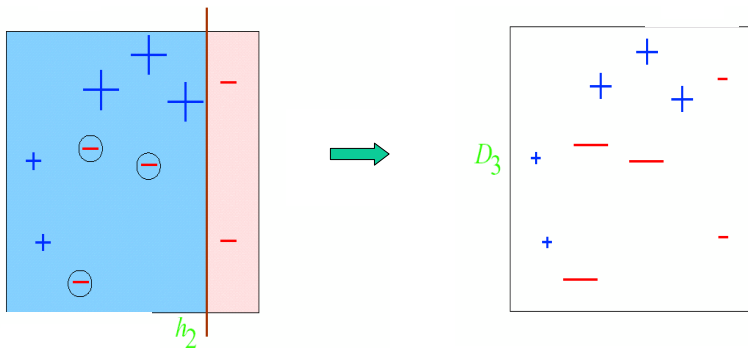


Todos tienen el mismo peso: $w_i = \frac{1}{N}, i \in [1, M]$

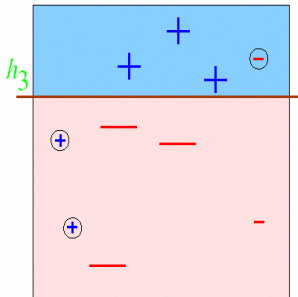
Ejemplo: Primera ronda



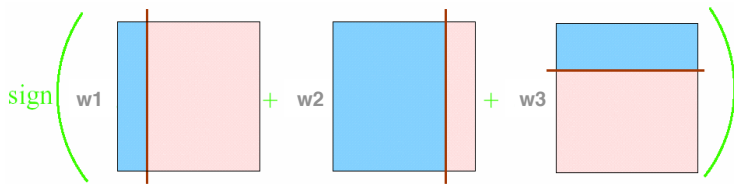
Ejemplo: Segunda ronda



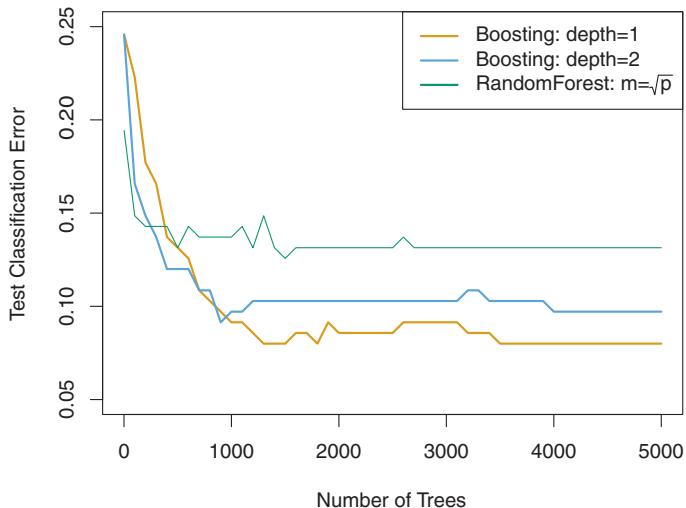
Ejemplo: Tercera ronda



Ejemplo: Modelo final



Boosting vs Random forest



Gradient Boosting

Algoritmo Gradient boosting (vanilla para regresión)

- 1: K número de rondas
 - 2: $(h(\mathbf{x}) - y)^2$ pérdida cuadrática $\ell(h(\mathbf{x}), y)$
 - 3: λ tasa de aprendizaje
-
- 4: $h(\mathbf{x}) = \bar{y}$ $\arg \min_c \sum_i \ell(c, y)$
 - 5: Repetir K veces
 - 6: Calcular residuos $r_i = y_i - h(\mathbf{x}_i)$ $-\frac{\partial \ell}{\partial h(\mathbf{x}_i)}(h(\mathbf{x}), y)$
 - 7: Entrenar árbol A con para predecir residuos $\{(\mathbf{x}_i, r_i)\}$
 - 8: Actualizar $h(\mathbf{x}) \leftarrow h(\mathbf{x}) + \lambda A(\mathbf{x})$
 - 9: Devolver regresor h

Ejemplo: Primera ronda

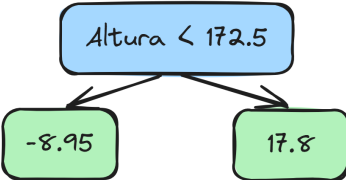
$h(x) =$

66.7

Sexo	Altura	Mano	Peso
M	179	26	75
M	163	21	59
F	163	20	53
M	175	24	94
F	170	19	57
F	154	19	62

Sexo	Altura	Mano	Residuo
M	179	26	8.3
M	163	21	-7.7
F	163	20	-13.7
M	175	24	27.3
F	170	19	-9.7
F	154	19	-4.7

Ejemplo: segunda ronda

$$h(x) = \boxed{66.7} + \underbrace{0.1}_{\lambda} \times$$


Sexo	Altura	Mano	Residuo
M	179	26	8.3
M	163	21	-7.7
F	163	20	-13.7
M	175	24	27.3
F	170	19	-9.7
F	154	19	-4.7

Sexo	Altura	Mano	Residuo
M	179	26	6.5
M	163	21	-6.8
F	163	20	-12.8
M	175	24	25.5
F	170	19	-8.8
F	154	19	-3.8

Bibliografía

- An introduction to statistical learning with applications in Python. Cap 8.
- Machine Learning - A First Course for Engineers and Scientists. Capítulo 7.
- Raschka, S. [Introduction to Machine Learning](#). Lecture 7. (2021)
- M. Stamp. Introduction to ML with Applications to Information Security, 7.4 y 7.5.
- Zhi-Hua Zhou. Ensemble Methods: Foundations and Algorithms.