

# Ensemble Learning (BOOSTING)

→ "Wisdom of the crowd" ~~✗~~

Ensemble learning

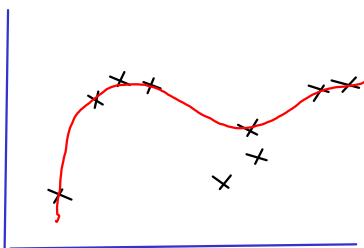
→ Entrenar un grupo de algoritmos y obtener una predicción a partir de ellos.

Por ejemplo: VOTAR  
(soft voting / hard voting)

- Bagging
- Random forest
- **BOOSTING**

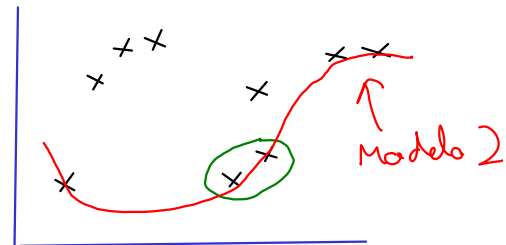
## Boosting

→ ADABOOST (Adaptative Boosting)

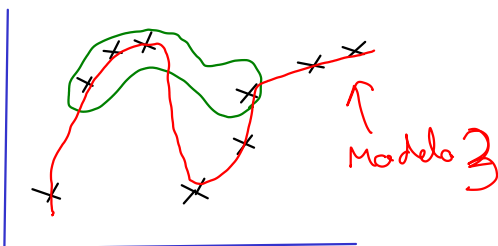


Modelo 1

→ Damos mayor peso relativo a los mal clasificados



Modelo 2



Modelo 3

→ Finalmente los pongo a votar

(Algoritmo)

**ADABOOST**

Base de datos:  $X = \begin{pmatrix} x^{(1)} \\ \vdots \\ x^{(m)} \end{pmatrix}$

$\vec{y} = \begin{pmatrix} y^{(1)} \\ \vdots \\ y^{(m)} \end{pmatrix}$  ← TARGET

1 (Originalmente todos los puntos valen lo mismo)

Fijo mas pesos:  $w^{(i)} = \frac{1}{m}$

2 Modelo 1  $\rightarrow$   $r_1 = \sum_{i=1}^m w^{(i)}$   
 Predicción del modelo 1  $\rightarrow \hat{y}_1$     Error del modelo 1  $\hat{y}_1 \neq y^{(i)}$

$\sum_{i=1}^m w^{(i)} = 1$

3 Peso del modelo:  $\alpha_1 = \eta \log \frac{1-r_1}{r_1}$  ← Lo que el valor del modelo

↑  
Hiperparámetro (LEARNING RATE)

4 Actualiza los pesos:

$$w^{(i)}_{(NUEVO)} = \begin{cases} w^{(i)} & \text{si } \hat{y}_1^{(i)} = y^{(i)} \\ w^{(i)} \exp(-\alpha_1) & \text{si } \hat{y}_1^{(i)} \neq y^{(i)} \end{cases}$$

Se dividen por  $\sum_i w^{(i)}_{NUEVOS}$   
(NORMALIZACIÓN)  
(Para que sumen 1)

DOY MAS PESO A LOS MAL CLASIFICADOS

MODELO [2]

Sigo iterando:

$W_j$  → Modelo  $j$  →

Predicción del modelo  $j$  →  $\hat{y}_j$

Error del modelo  $j$   $r_j = \sum_{i=1}^m w_j^{(i)}$  /  $\hat{y}_j^{(i)} \neq y^{(i)}$  /  $\sum_{i=1}^m w_j^{(i)} = 1$

Peso del modelo:  $\alpha_j = \eta \log \frac{1-r_j}{r_j}$  ← Lo que elvota del modelo

↑  
Hiperparámetro (LEARNING RATE)

Actualiza los pesos:

$$w_{j+1}^{(i)} = \begin{cases} w_j^{(i)} & \text{si } \hat{y}_j^{(i)} = y^{(i)} \\ w_j^{(i)} \exp(-\alpha_j) & \text{si } \hat{y}_j^{(i)} \neq y^{(i)} \end{cases}$$

SI GO ITERANDO

$j = 1, \dots, N$  ↑  
NÚMERO DE MODELOS

5 Predicción final

$$\hat{y}^{(i)} = \underset{K}{\operatorname{argmax}} \sum_{j=1}^N q_j$$

↑  
(CLASES)  $\hat{y}_j^{(i)} = K$  ← EL MÁS VOTADO

↑  
 $C_1, \dots, C_K$   
los posibles valores de  $\hat{y}$