XGBoost.

Ensamble → Combinación de varios modelos
xgboost → Árboles.

↓

Extreme Gradient Boosting ↔ Supervisada → Conocida la salida

NO supervisada → No conocemos salida.

Utilizamos los datos de train con múltiples funciones (x_i) para predecir salida (y_i).

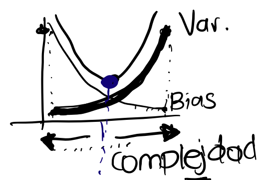
FUNCIÓN OBJETIVO: Pérdida (Training loss) + Regularización.

$$\underbrace{\text{Pérdida (Training loss)}}_{L} + \underbrace{\text{Regularización}}_{\Omega}$$

$$obj(\theta) = L(\theta) + \Omega(\theta).$$

L: Mide qué tan predictivo es el modelo respecto al train.

$$L(\theta) = \sum_i (y_i - \hat{y}_i)^2$$

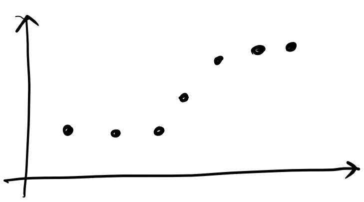


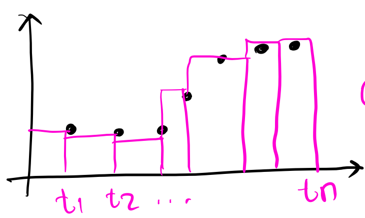
Pérdida de la regresión logística.

$$L(\theta) = \sum_i \left[y_i \ln(1+e^{-\hat{y}_i}) + (1-y_i)\ln(1+e^{\hat{y}_i}) \right].$$

Ω → optimizar.

Ω: Controla la complejidad del Modelo (evitar sobreajuste).

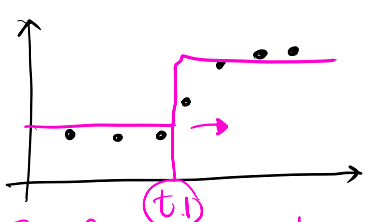




Over fiting.

Alta complejidad.

Muchos splits. $\Omega \uparrow$

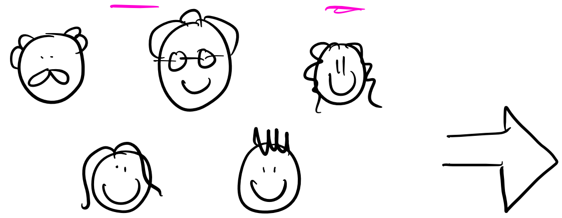


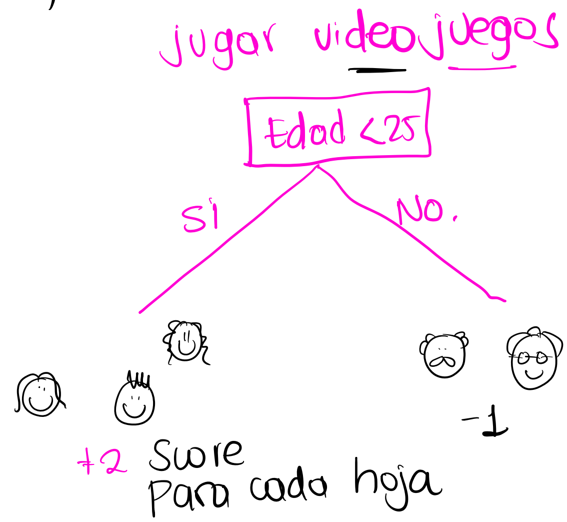
Pocos splits $\Omega \downarrow$

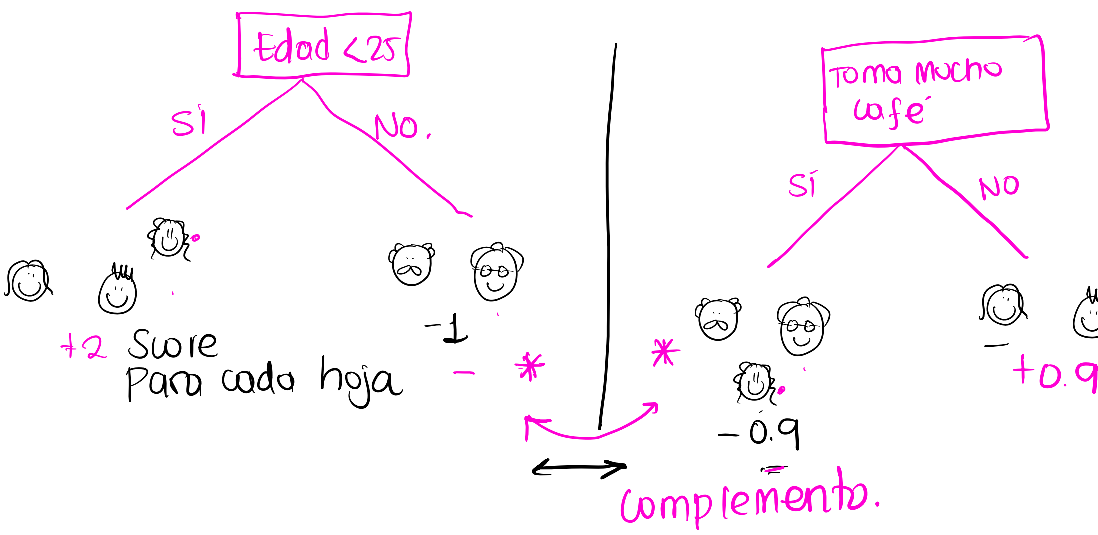
$L(\theta) \uparrow$

Xgboost → conjunto de arboles de desicion

input: Edad, Genero, Ocupación, etc.







$f(\ ) = -1 - 0.9 = -1.9$

$f(\ ) = +2 + 0.9 = \|2.9\|$

Cómo traducirlo al modelo:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F}$$

Posibles para tu modelo y tus datos.

K: Número de árboles
f_k: función en el espacio funcional $\mathcal{F}$
F: conjunto de todos los $\boxed{\text{CART}}$ Posibles.

FUNCIÓN OBJETIVO:

$$obj(\theta) = L(\theta) + \Omega(\theta).$$

$$= \underbrace{\sum_{i}^{n} l(y_i, \hat{y}_i)}_{\text{Loss}} + \underbrace{\sum_{k=1}^{K} w(f_k)}_{\text{Regularización.}}$$

$w(f_k) \rightarrow$ Complejidad del modelo. del árbol f_k.

Boosting Tree: (función objetivo de forma recurrente).

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \boxed{\sum_{i=1}^{t} w(f_i)}$$

$\rightarrow w(f_t) + \text{Const.}$

$\hat{y}_i$: predicción.

Predicción para cada paso t:

$$\hat{y}_i^{(0)} = 0$$

$$\hat{y}_i^{(1)} = f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i)$$

$$\hat{y}_i^{(2)} = f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i)$$

$$\vdots$$

$$\boxed{\hat{y}_i^{(t)}} = \sum_{k=1}^{t} f_k(x_i) = \boxed{\hat{y}_i^{(t-1)} + f_t(x_i)}.$$

La función objetivo:

$$obj^{(t)} = \sum_{i=1}^{n} \underset{\downarrow \text{MSE}}{l}(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + w(f_t) + \text{Cte}$$

Función obj. Para MSE (Error cuadrático Medio).

$$obj^{(t)} = \sum_{i=1}^{n} \boxed{\left(y_i - \left(\hat{y}_i^{(t-1)} + f_t(x_i) \right) \right)^2} + \omega(f_t) + conste.$$

$$= \sum_{i=1}^{n} \boxed{\left[2\left(\hat{y}_i^{(t-1)} - y_i \right) f_i(x_i) + f_i(x_i)^2 \right]} + \omega(f_t) + cost.$$