

# Machine Learning

## Árboles de decisión

Christian Oliva Moya  
Luis Fernando Lago Fernández

# Introducción - Árboles de decisión

---

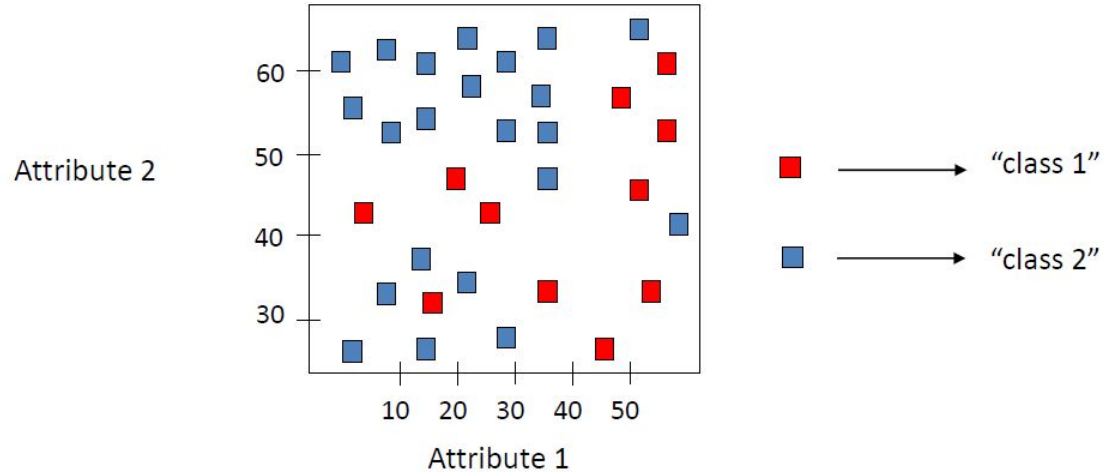
Un árbol de decisión es un algoritmo de ML:

- Supervisado
- De clasificación
- No paramétrico

Divide el espacio de forma recursiva utilizando reglas de decisión

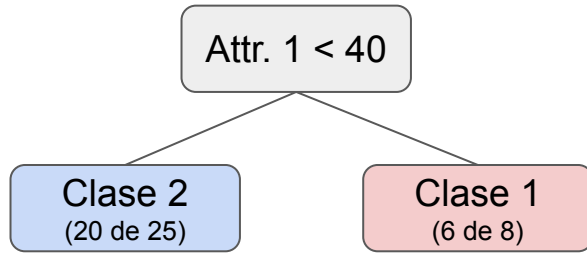
# Árboles de decisión - Ejemplo

Pongamos un ejemplo:

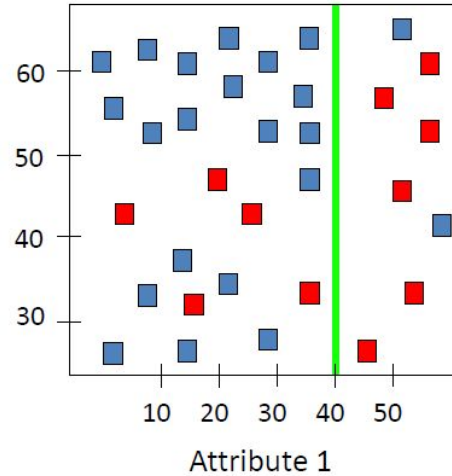


# Árboles de decisión - Ejemplo

Pongamos un ejemplo:



Attribute 2

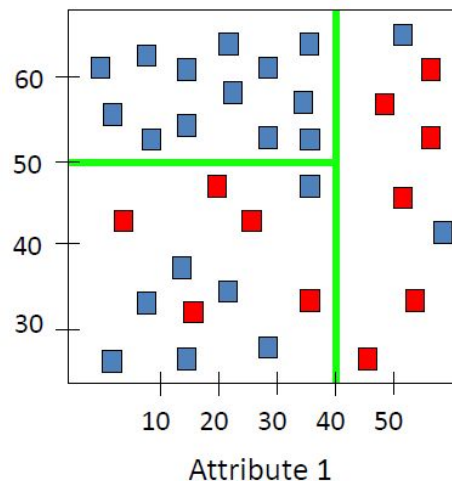
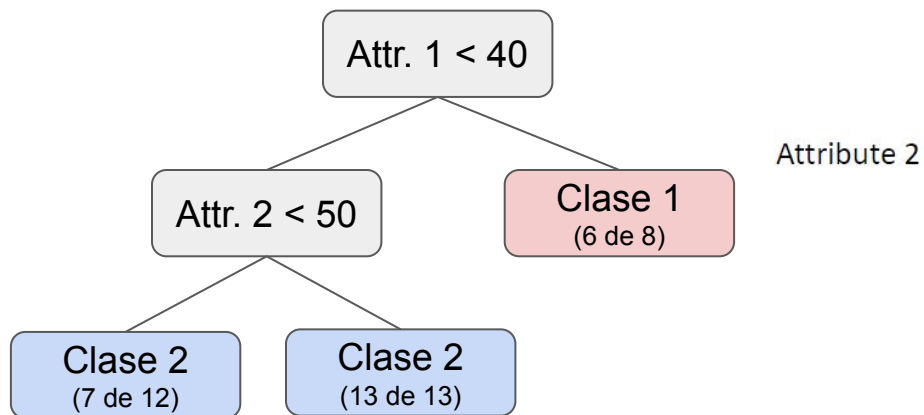


■ → "class 1"

■ → "class 2"

# Árboles de decisión - Ejemplo

Pongamos un ejemplo:

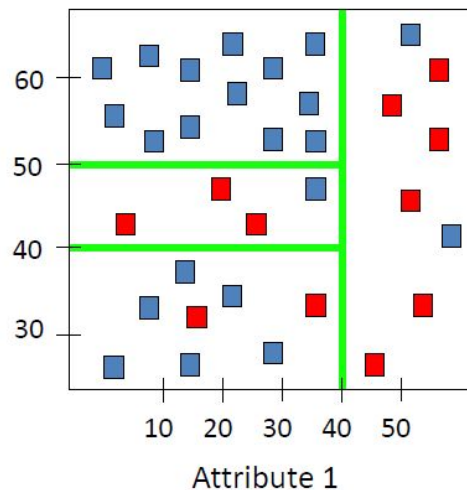
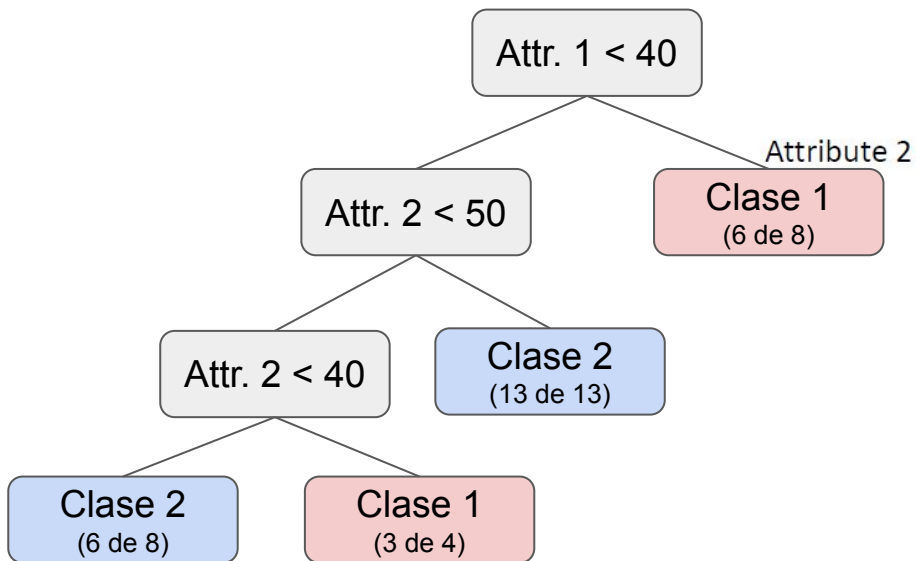


■ → "class 1"

■ → "class 2"

# Árboles de decisión - Ejemplo

Pongamos un ejemplo:



■ → "class 1"

■ → "class 2"

# Árboles de decisión - Conceptos

---

- Diagrama en forma de árbol que representa condiciones sucesivas sobre los atributos para clasificar una instancia
- Tipos de nodos:
  - **Nodos internos:**
    - Definen preguntas condicionales sobre los atributos
  - **Nodos hoja:**
    - Ejemplos que cumplen la condición y dan una predicción
- **Objetivo:** Construir el árbol más sencillo que mejor separe los ejemplos por clase

# Árboles de decisión

---

- ¿Cómo funciona?

Es un algoritmo iterativo. Mientras exista un corte que mejore el criterio de separación:

1. Se selecciona **la mejor separación** de acuerdo al criterio
2. Se añade la condición al árbol incluyendo las posibles respuestas
3. Se calculan los nodos hoja de la nueva condición añadida
4. **Criterio de parada**. Si no se cumple se vuelve al paso 1

- ¿Cómo saber la mejor separación?
- ¿Qué criterio de parada puede haber?



# Árboles de decisión - Impureza: Entropía

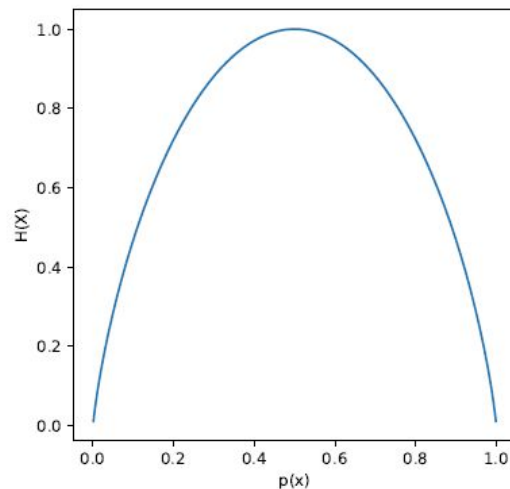
- ¿Cómo saber la mejor separación?

La **entropía** mide la incertidumbre en la fuente de información, es decir...

**Cómo de desordenados están los ejemplos**

$$E(X) = - \sum_{i=1}^c P(C_i) \log_2(P(C_i))$$

Donde  $P(C_i)$  es la proporción de ejemplos que son clasificados como clase  $C_i$



# Árboles de decisión - Impureza: Gini

---

- ¿Cómo saber la mejor separación?

Una métrica alternativa es la **impureza de Gini**

Mide la probabilidad de clasificar incorrectamente un dato aleatorio si seguimos la distribución:

$$Gini(X) = 1 - \sum_{i=1}^c P(C_i)^2$$

Donde  $P(C_i)$  es la proporción de ejemplos que son clasificados como clase  $C_i$

# Árboles de decisión - Impureza

- ¿Cómo saber la mejor separación?

Otras funciones de impureza:

Criterion	Impurity function $I(q_1, q_2, \dots, q_C)$	Comments
Expected error	$1 - \max(q_1, q_2, \dots, q_C)$	Causes many ties
GINI (CART)	$1 - \sum q_k^2$	If 2 classes: $2 q_1 q_2$
Entropy (ID3, C5)	$-\sum q_k \log_2 q_k$	$0 \log_2 0 \equiv 0$
DKM	$2 \sqrt{q_1 \cdot q_2}$	Only 2 classes. Robust for unbalanced classes
CHAID	$\chi^2$	"Chi-square Automatic Interaction Detector"

# Árboles de decisión - Impureza

---

- ¿Cómo saber la mejor separación?

En cualquier caso, buscamos la separación que minimice la impureza global:

$$\text{global } I(s) = \sum_{j=1}^n P_j \cdot I(q_1(j), q_2(j), \dots, q_c(j))$$

# Árboles de decisión - Ejemplo

Ejemplo de ejecución con impureza de Gini y dos clases:

Gini with **two classes**:  $I = 2 q_1 q_2$

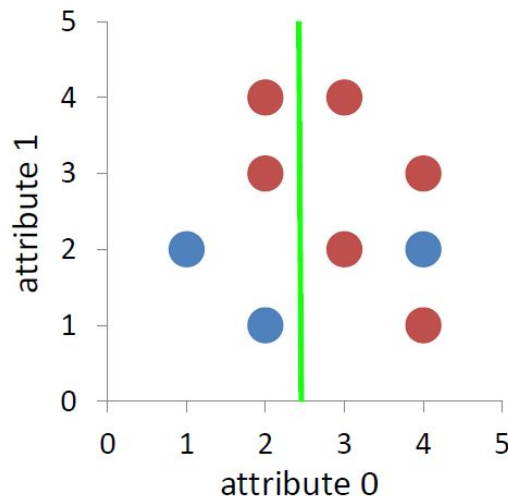
Total impurity of “**attribute 0 > 2.5**” ?

response = Yes:

$$4 \text{ (red)} \quad 1 \text{ (blue)} \quad I = 2 \cdot \frac{4}{5} \cdot \frac{1}{5} = \frac{8}{25}$$

response = No:

$$2 \text{ (red)} \quad 2 \text{ (blue)} \quad I = 2 \cdot \frac{2}{4} \cdot \frac{2}{4} = \frac{1}{2}$$

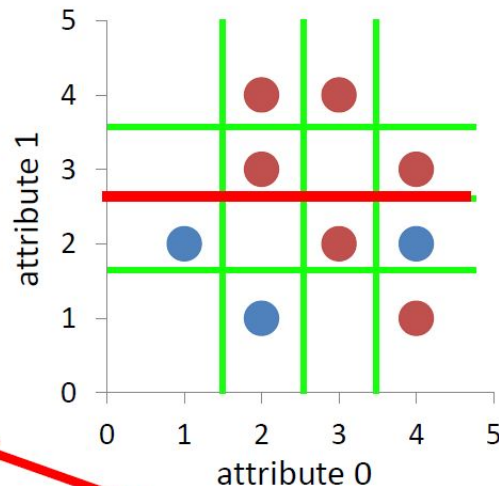


$$\text{Total impurity} = P(\text{Yes}) \cdot \frac{8}{25} + P(\text{No}) \cdot \frac{1}{2} = \frac{5}{9} \cdot \frac{8}{25} + \frac{4}{9} \cdot \frac{1}{2} = \frac{2}{5}$$

# Árboles de decisión - Ejemplo

Ejemplo de ejecución con impureza de Gini y dos clases:

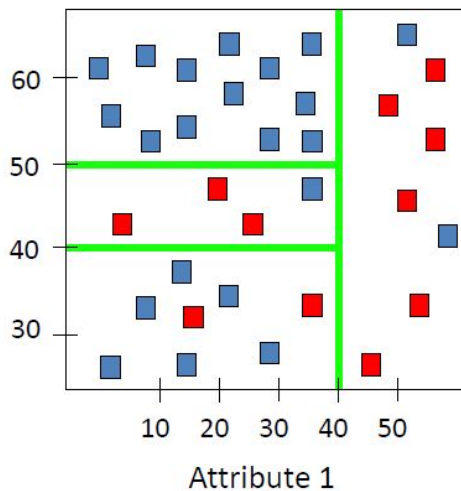
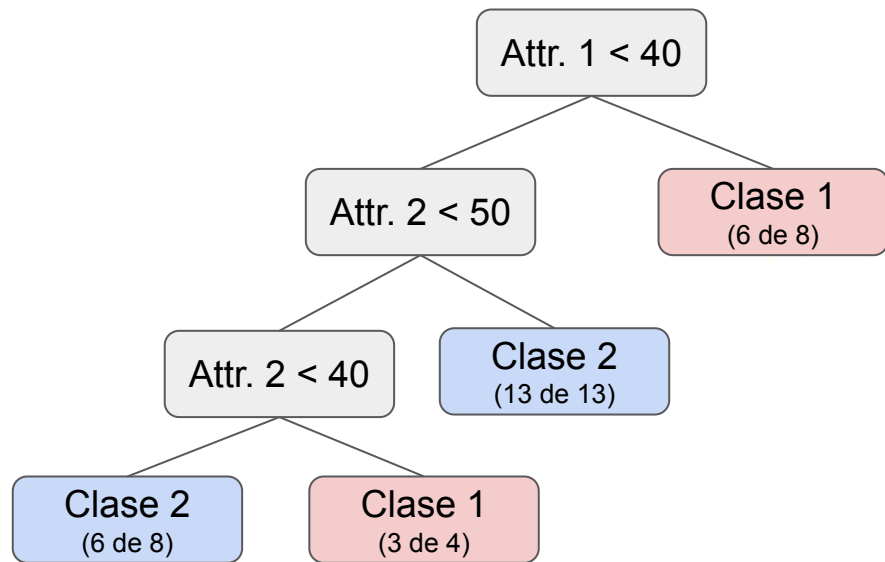
Query	Yes		No		I(s) total
	$n_+, n_-$	I	$n_+, n_-$	I	
$x_0 > 1.5$	6, 2	$\frac{3}{8}$	0, 1	0	$\frac{1}{3} = 0.333$
$x_0 > 2.5$	4, 1	$\frac{8}{25}$	2, 2	$\frac{1}{2}$	$\frac{2}{5} = 0.400$
$x_0 > 3.5$	2, 1	$\frac{4}{9}$	4, 2	$\frac{4}{9}$	$\frac{4}{9} = 0.444$
$x_1 > 1.5$	5, 2	$\frac{20}{49}$	1, 1	$\frac{1}{2}$	$\frac{3}{7} = 0.429$
$x_1 > 2.5$	4, 0	0	2, 3	$\frac{12}{25}$	$\frac{4}{15} = 0.267$
$x_1 > 3.5$	2, 0	0	4, 3	$\frac{24}{49}$	$\frac{8}{21} = 0.381$



**Best query: attribute  $_1 > 2.5$   
(lowest total impurity)**

# Árboles de decisión - Criterio de parada

- ¿Qué criterio de parada puede haber?
  - Profundidad máxima del árbol
  - Que la impureza sea menor que un umbral
  - Número mínimo de datos en un nodo hoja. En este árbol el mínimo es 10.



■ → "class 1"

■ → "class 2"

# Árboles de decisión - Ventajas e Inconvenientes

---

- Ventajas

Se pueden representar **visualmente**. Son fáciles de interpretar

Realizan una **selección de características** de forma implícita

Soportan datos categóricos y numéricos

Son muy rápidos

- Inconvenientes

Rápido **sobreajuste** (overfitting) a los datos de entrenamiento con árboles con mucha profundidad

Árboles sesgados para clases no balanceadas

Es un algoritmo Greedy