



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Árboles de clasificación

Alfons Juan
Jorge Civera
Albert Sanchis

DSIC

Departamento de Sistemas
Informáticos y Computación

Objetivos formativos

- Definir un árbol de clasificación
- Aplicar un algoritmo de aprendizaje convencional

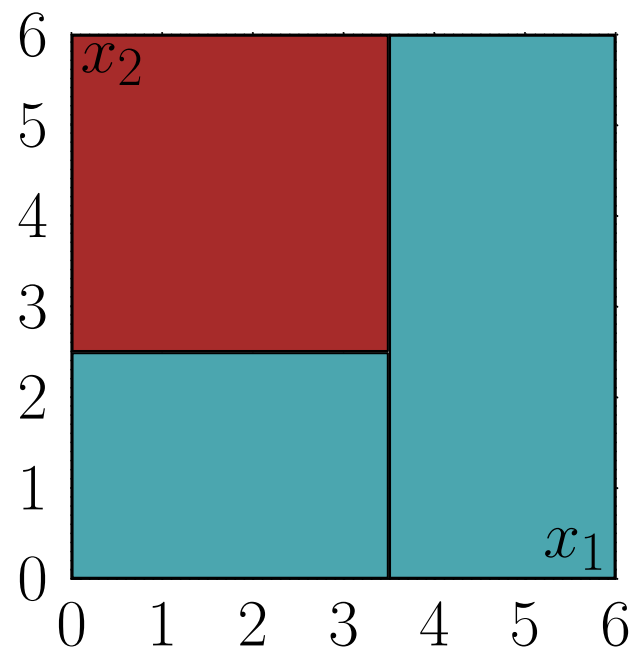
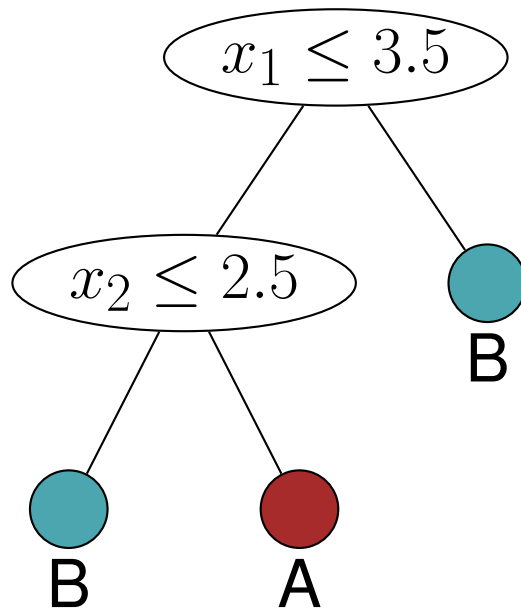
Índice

1	Árboles de clasificación	3
2	Algoritmo de aprendizaje	4

1. Árboles de clasificación

Un *árbol de clasificación* es una estructura jerárquica para la clasificación de objetos.

Ejemplo: objetos del tipo $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$



2. Algoritmo de aprendizaje

Árbol($S = \{(x_n, c_n)\}$) // S es un conjunto de aprendizaje
 ($\mathcal{C}, L, R, \Delta\mathcal{I}$) = **Dicotomiza**(S)
 si $\Delta\mathcal{I} < \epsilon$ **devuelve** **Nodo**(**Moda**($\{c_n\}$), $-$, $-$, $-$)
 si no **devuelve** **Nodo**($-$, \mathcal{C} , **Árbol**(L), **Árbol**(R))

donde:

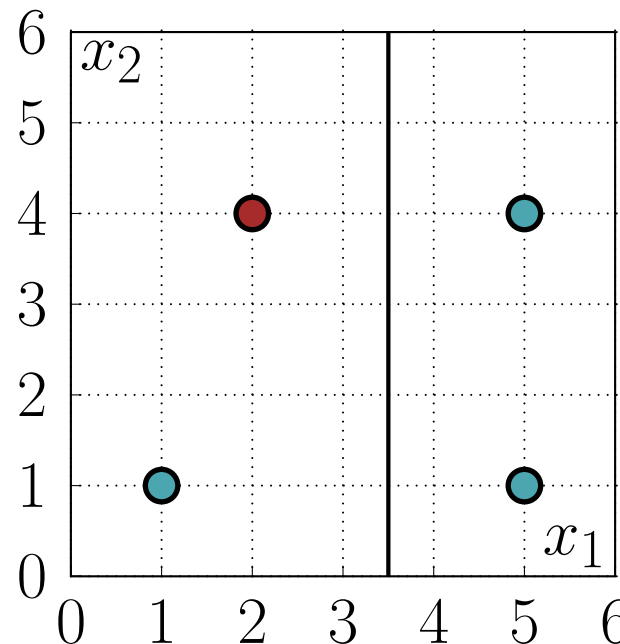
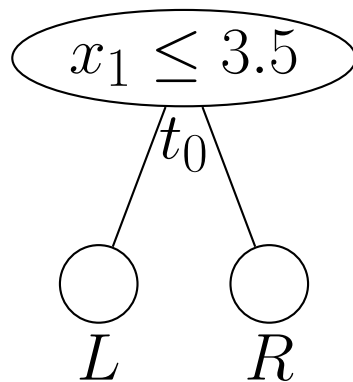
- **Dicotomiza** divide el conjunto S en dos subconjuntos, L y R , aplicando un ***criterio de partición***, \mathcal{C} , con el cual consigue un ***decremento de impureza*** $\Delta\mathcal{I}$
- ϵ es un ***umbral de impureza*** dado

Criterio de partición

Un criterio de partición usual consiste en elegir un par variable-umbral, (d, r) , y dividir los datos $\{(\mathbf{x}_n, c_n)\}$ de la siguiente manera:

$$L = \{(\mathbf{x}_n, c_n) : x_{nd} \leq r\} \quad \text{y} \quad R = \{(\mathbf{x}_n, c_n) : x_{nd} > r\}$$

Ejemplo: partición de los datos de la figura con $(d, r) = (1, 3.5)$



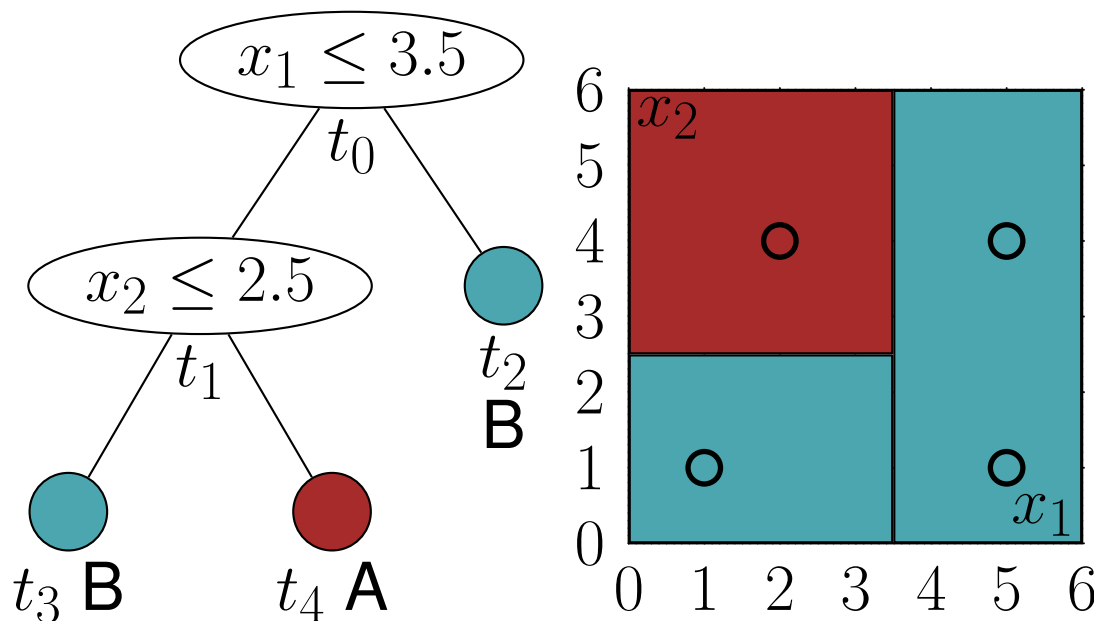
$$L = \left\{ \begin{pmatrix} 1 \\ 1 \\ \bullet \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ \bullet \end{pmatrix} \right\}$$
$$R = \left\{ \begin{pmatrix} 5 \\ 1 \\ \bullet \end{pmatrix}, \begin{pmatrix} 5 \\ 4 \\ \bullet \end{pmatrix} \right\}$$

Impureza de un nodo

La **impureza de un nodo** t es la incertidumbre sobre la clase de los objetos en t ; suele calcularse como la **entropía** de la distribución empírica de la probabilidad posteriori de clases en t :

$$\mathcal{I}(t) = - \sum_{c=1}^C \hat{P}(c | t) \log_2 \hat{P}(c | t) \quad \text{donde} \quad \hat{P}(c | t) = \frac{N_c(t)}{N(t)}$$

Ejemplo:



$$\begin{aligned} \mathcal{I}(t_0) &= -\frac{1}{4} \log_2 \frac{1}{4} - \frac{3}{4} \log_2 \frac{3}{4} \\ &= .50 + .31 = .81 \end{aligned}$$

$$\begin{aligned} \mathcal{I}(t_1) &= -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} \\ &= \frac{1}{2} + \frac{1}{2} = 1 \end{aligned}$$

$$\mathcal{I}(t_2) = \mathcal{I}(t_3) = \mathcal{I}(t_4) = 0$$

Maximización del decremento de impureza

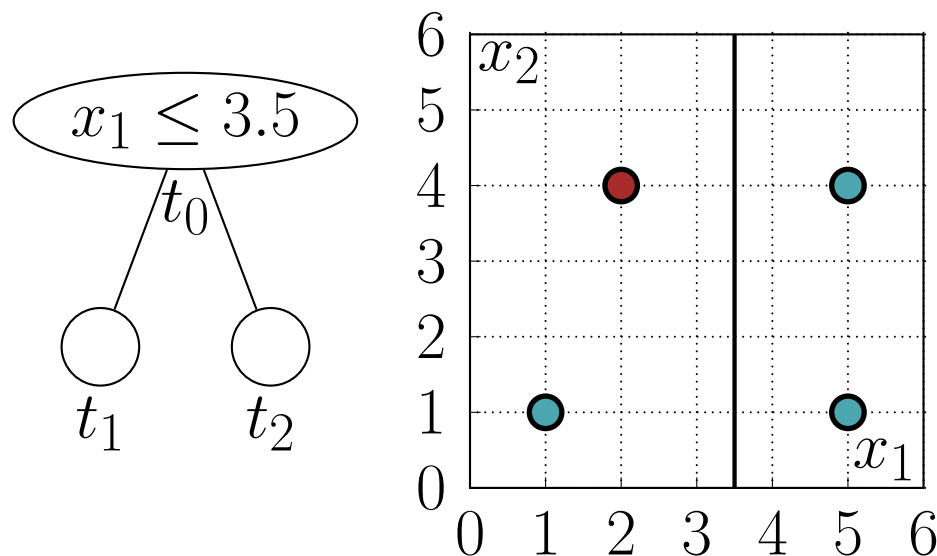
La partición de un nodo t se realiza buscando un par variable-umbral que consigue un máximo decremento de impureza:

$$(d^*, r^*) = \arg \max_{d, r} \Delta \mathcal{I}(d, r)$$

donde

$$\Delta \mathcal{I}(d, r) = \mathcal{I}(t) - \frac{N(L(t))}{N(t)} \mathcal{I}(L(t)) - \frac{N(R(t))}{N(t)} \mathcal{I}(R(t))$$

Ejemplo: decremento de impureza con $(d, r) = (1, 3.5)$



$$\begin{aligned} \Delta \mathcal{I}(1, 3.5) &= 0.81 - \frac{2}{4} \cdot 1 - \frac{2}{4} \cdot 0 \\ &= 0.81 - 0.50 \\ &= 0.31 \end{aligned}$$

Conclusiones

- Hemos visto qué es un árbol de clasificación y como aprenderlo con una técnica convencional