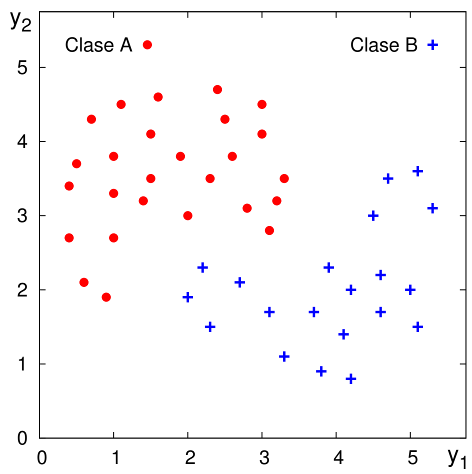
Árboles de clasificación

Alfons Juan
Jorge Civera
Albert Sanchis

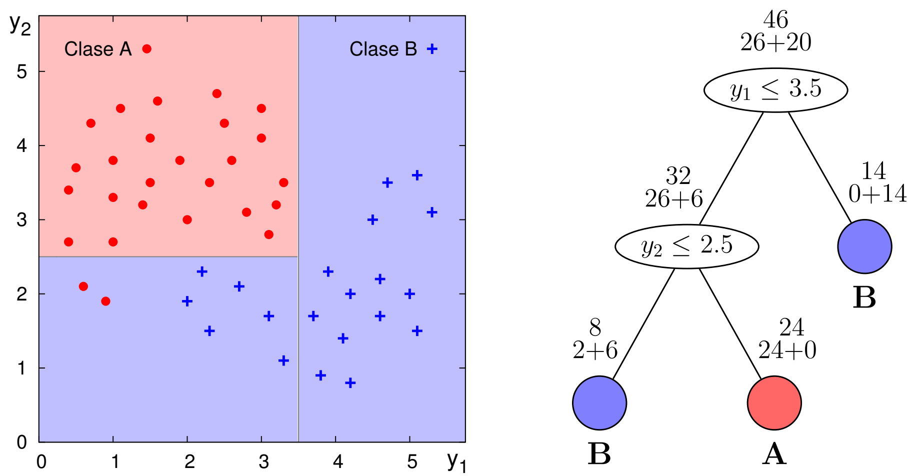
Departamento de Sistemas
Informáticos y Computación

Problema

- A partir de la muestra de aprendizaje que se muestra en la figura, aprende un árbol de clasificación T para clasificar objetos representados mediante vectores de características reales bidimensionales en dos posibles clases (A,B).

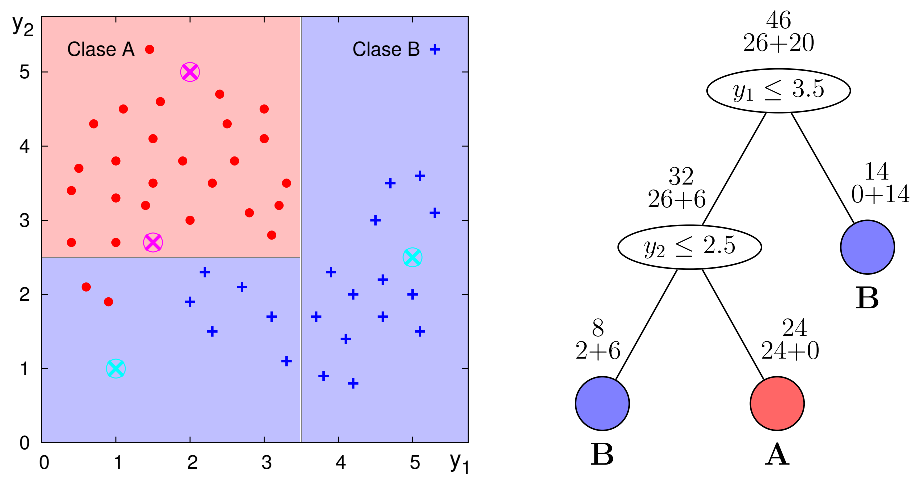


Árbol aprendido



Las regiones de decisión están formadas por bloques de forma rectangular, ya que las fronteras de decisión son siempre paralelas a los ejes.

Clasificación de nuevos datos



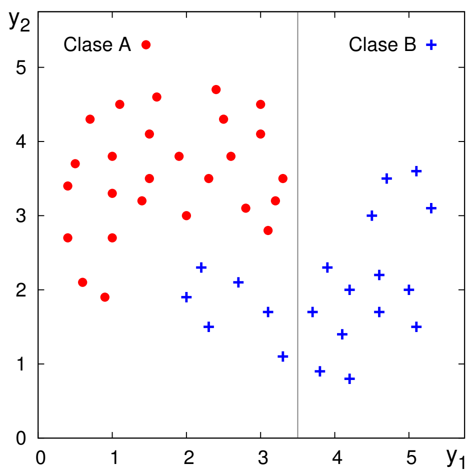
El árbol de decisión obtenido permite clasificar nuevos datos.

Construcción de un ADC a partir de una muestra de aprendizaje

Elementos necesarios en el proceso de construcción de un árbol de decisión:

1. Método para hacer particiones y para seleccionar la mejor; concretamente:
   - Condiciones o "preguntas" ("splits") admisibles para formar particiones.
   - Evaluación y optimización de la calidad de una partición

Primera partición



La primera partición se establece en base a la pregunta: ¿$y_1 \leq 3.5$?.

Evaluación de la calidad de una partición

- Para evaluar las particiones posibles se usa el concepto de *"impureza"*
- La impureza de un nodo t, $\mathcal{I}(t)$, se mide en función de las probabilidades estimadas de las clases en t.

Nodos:	$\hat{P}(A \mid t_i)$	$\hat{P}(B \mid t_i)$
t_1		
t_2		
t_3		

Probabilidad a posteriori de clase en el nodo t: $\hat{P}(c \mid t) = \dfrac{N_c(t)}{N(t)}$

- $N_c(t)$: número de datos en el nodo t de la clase c.
- $N(t)$: número de datos en el nodo t

Evaluación de la calidad de una partición

- Para evaluar las particiones posibles se usa el concepto de *"impureza"*
- La impureza de un nodo t, $\mathcal{I}(t)$, se mide en función de las probabilidades estimadas de las clases en t.

Nodos:	$\hat{P}(A \mid t_i)$	$\hat{P}(B \mid t_i)$
t_1	$26/46$	$20/46$
t_2	$26/32$	$6/32$
t_3	$0/14$	$14/14$

Probabilidad a posteriori de clase en el nodo t: $\hat{P}(c \mid t) = \dfrac{N_c(t)}{N(t)}$

- $N_c(t)$: número de datos en el nodo t de la clase c.
- $N(t)$: número de datos en el nodo t

Evaluación de la calidad de una partición

- La impureza de un nodo se calcula basándose en el concepto de *entropía*:

$$\mathcal{I}(t) \;=\; -\sum_{c=1}^{C} \hat{P}(c \mid t) \log_2 \hat{P}(c \mid t)$$

Nodos:	$\hat{P}(A \mid t_i)$	$\hat{P}(B \mid t_i)$	$\mathcal{I}(t_i)$
t_1	$26/46$	$20/46$	
t_2	$26/32$	$6/32$	
t_3	$0/14$	$14/14$	

Evaluación de la calidad de una partición

- La impureza de un nodo se calcula basándose en el concepto de *entropía*:

$$\mathcal{I}(t) \;=\; -\sum_{c=1}^{C} \hat{P}(c \mid t) \log_2 \hat{P}(c \mid t)$$

Nodos:	$\hat{P}(A \mid t_i)$	$\hat{P}(B \mid t_i)$	$\mathcal{I}(t_i)$
t_1	$26/46$	$20/46$	0.988
t_2	$26/32$	$6/32$	0.696
t_3	$0/14$	$14/14$	0.000

Evaluación de la calidad de una partición

- La calidad de una partición se mide mediante el *decremento de impureza*:

$$\Delta\mathcal{I}(t) \overset{\text{def}}{=} \mathcal{I}(t) - \hat{P}_t(L)\mathcal{I}(t_L) - \hat{P}_t(R)\mathcal{I}(t_R)$$

Nodos:	$\hat{P}(A \mid t_i)$	$\hat{P}(B \mid t_i)$	$\hat{P}_{ti}(L)$	$\hat{P}_{ti}(R)$	$\mathcal{I}(t_i)$	$\Delta\mathcal{I}(t_i)$
t_1	$26/46$	$20/46$			0.988	
t_2	$26/32$	$6/32$			0.696	
t_3	$0/14$	$14/14$			0.000	

Probabilidad de decisión por el hijo izquierdo de t : $\quad \hat{P}_t(L) = \dfrac{N(t_L)}{N(t)}$

Probabilidad de decisión por el hijo derecho de t : $\quad \hat{P}_t(R) = \dfrac{N(t_R)}{N(t)}$

- $N(t)$: número de datos en el nodo t

Evaluación de la calidad de una partición

- La calidad de una partición se mide mediante el *decremento de impureza*:

$$\Delta\mathcal{I}(t) \overset{\text{def}}{=} \mathcal{I}(t) - \hat{P}_t(L)\mathcal{I}(t_L) - \hat{P}_t(R)\mathcal{I}(t_R)$$

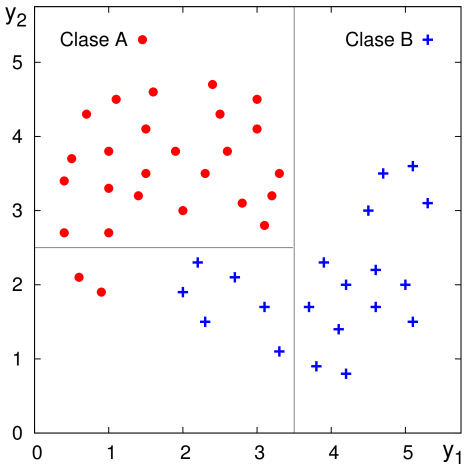
Nodos:	$\hat{P}(A \mid t_i)$	$\hat{P}(B \mid t_i)$	$\hat{P}_{ti}(L)$	$\hat{P}_{ti}(R)$	$\mathcal{I}(t_i)$	$\Delta\mathcal{I}(t_i)$
t_1	$26/46$	$20/46$	$32/46$	$14/46$	0.988	0.504
t_2	$26/32$	$6/32$			0.696	
t_3	$0/14$	$14/14$			0.000	

Probabilidad de decisión por el hijo izquierdo de t : $\quad \hat{P}_t(L) = \dfrac{N(t_L)}{N(t)}$

Probabilidad de decisión por el hijo derecho de t : $\quad \hat{P}_t(R) = \dfrac{N(t_R)}{N(t)}$

- $N(t)$: número de datos en el nodo t

Segunda partición



En el nodo de la izquierda se procede a una segunda partición con la pregunta: ¿$y_2 \leq 2.5$?.

Evaluación de la segunda partición

Nodos:	$\hat{P}(A \mid t_i)$	$\hat{P}(B \mid t_i)$	$\hat{P}_{ti}(L)$	$\hat{P}_{ti}(R)$	$\mathcal{I}(t_i)$	$\Delta\mathcal{I}(t_i)$
t_1	$26/46$	$20/46$	$32/46$	$14/46$	0.988	0.504
t_2	$26/32$	$6/32$			0.696	
t_3	$0/14$	$14/14$			0.000	
t_4						
t_5						

Evaluación de la segunda partición

Nodos:	$\hat{P}(A \mid t_i)$	$\hat{P}(B \mid t_i)$	$\hat{P}_{ti}(L)$	$\hat{P}_{ti}(R)$	$\mathcal{I}(t_i)$	$\Delta\mathcal{I}(t_i)$
t_1	$26/46$	$20/46$	$32/46$	$14/46$	0.988	0.504
t_2	$26/32$	$6/32$	$8/32$	$24/32$	0.696	0.493
t_3	$0/14$	$14/14$			0.000	
t_4	$2/8$	$6/8$			0.811	
t_5	$24/24$	$0/24$			0.000	

Criterios de suficiente "pureza" en nodos terminales

Un nodo t es *terminal* si el máximo decremento de impureza posible es demasiado pequeño:

$$\max_{\substack{1 \leq j \leq D \\ -\infty < r < +\infty}} \Delta\mathcal{I}(j, r, t) \; < \; \epsilon$$

donde ϵ es una constante pequeña a determinar empíricamente.

Otro posible criterio es exigir que los nodos terminales sean totalmente puros.

Nodos:	$\hat{P}(A \mid t_i)$	$\hat{P}(B \mid t_i)$	$\hat{P}_{ti}(L)$	$\hat{P}_{ti}(R)$	$\mathcal{I}(t_i)$	$\Delta\mathcal{I}(t_i)$
t_1	$26/46$	$20/46$	$32/46$	$14/46$	0.988	0.504
t_2	$26/32$	$6/32$	$8/32$	$24/32$	0.696	0.493
t_3 (terminal)	$0/14$	$14/14$			0.000	
t_4	$2/8$	$6/8$			0.811	
t_5 (terminal)	$24/24$	$0/24$			0.000	

Asignación de etiquetas de clase a nodos terminales

A cada nodo terminal se asigna la clase de la mayoría de sus elementos:

$$c^\star(t) = \operatorname*{argmax}_{1 \le c \le C} \hat{P}(c \mid t), \quad \forall t \in \tilde{T}$$

Nodos:	$\hat{P}(A \mid t_i)$	$\hat{P}(B \mid t_i)$	$\hat{P}_{ti}(L)$	$\hat{P}_{ti}(R)$	$\mathcal{I}(t_i)$	$\Delta\mathcal{I}(t_i)$	$c^\star(t)$
t_1	$26/46$	$20/46$	$32/46$	$14/46$	0.988	0.504	
t_2	$26/32$	$6/32$	$8/32$	$24/32$	0.696	0.493	
t_3 (terminal)	$0/14$	$14/14$			0.000		B
t_4	$2/8$	$6/8$			0.811		
t_5 (terminal)	$24/24$	$0/24$			0.000		A

Ejercicio: Continúa el ejercicio procediendo a una nueva partición del nodo t_4 con la pregunta: ¿$y_1 \le 1.5$?.