

Árboles de Decisión (DT)

Alfredo Cuesta Infante

E. T. S. Ingeniería Informática
Universidad Rey Juan Carlos

Master Univ. en Visión Artificial
Reconocimiento de Patrones

Intuición

Pureza de una división

Concatenar decisiones

Construcción del árbol

CART

Predicción

Intuición

Pureza de una división

Concatenar decisiones

Construcción del árbol

CART

Predicción

Intuición

Pureza de una división

Concatenar decisiones

Construcción del árbol

CART

Predicción

Pureza de una división

- ▶ Conjunto de datos: '0' vs. '1'
- ▶ 2 características:
($X_1 = W$, $x_2 = H$)
- ▶ En la figura sólo se muestran 30 ejemplos.
- ▶ Ordenar el conjunto por x_1
 - ▶ $\hat{t} = 1$ si $x_1 < 10$
→ 1 'impureza'
 - ▶ $\hat{t} = 1$ si $x_1 < 11$
→ 2 'impurezas'
 - ▶ $\hat{t} = 1$ si $x_1 < 12$
→ 3 'impurezas'
- ▶ Ordenar el conjunto por x_2
 - ▶ Cualquier selección tiene muchas 'impurezas'

x_1	x_2	t
15	17	0
11	19	1
2	18	1
13	17	0
10	18	0
14	17	0
16	17	0
11	18	0
3	18	1
7	10	1
15	18	0
3	19	1
1	18	1
4	18	1
1	8	1
17	13	0
2	19	1
2	17	1
3	18	1
3	17	1
12	19	0
16	17	0
2	17	1
12	17	0
13	17	0
3	19	1
8	19	1
4	18	1
16	16	0
2	18	1
9	19	1
2	19	1
9	19	1
7	17	1
9	17	1
14	18	0
11	18	0
1	17	1

x_1	x_2	t
1	18	1
1	8	1
1	17	1
2	18	1
2	19	1
2	17	1
2	17	1
2	18	1
2	19	1
3	18	1
3	19	1
3	18	1
3	17	1
3	19	1
4	18	1
4	18	1
7	10	1
7	17	1
8	19	1
9	19	1
9	19	1
9	17	1
10	18	0
11	19	1
11	18	0
11	18	0
12	19	0
12	17	0
13	17	0
13	17	0
14	17	0
14	18	0
15	17	0
15	18	0
16	17	0
16	16	0
17	13	0

x_1	x_2	t
1	8	1
7	10	1
17	13	0
16	16	0
15	17	0
13	17	0
14	17	0
16	17	0
2	17	1
3	17	1
16	17	0
2	17	1
12	17	0
13	17	0
7	17	1
9	17	1
1	17	1
2	18	1
10	18	0
11	18	0
3	18	1
15	18	0
1	18	1
4	18	1
3	18	1
4	18	1
2	18	1
14	18	0
11	18	0
11	19	1
3	19	1
2	19	1
12	19	0
3	19	1
8	19	1
9	19	1
2	19	1
9	19	1



— ??

Intuición

Pureza de una división

Concatenar decisiones

Construcción del árbol

CART

Predicción

Intuición

Concatenar decisiones

- ▶ Una vez dividida la tabla en dos, podemos seguir el proceso con cada una de las dos partes.
- ▶ Cada vez que se crea una división de la tabla (de una subtabla) en dos se crea un nodo del árbol de decisión y dos ramas
 - ▶ Si la condición del nodo se cumple se continúa por la rama de la izquierda
 - ▶ Si no, se continúa por la rama de la derecha
- ▶ Una rama acaba en una hoja cuando le asignamos una etiqueta

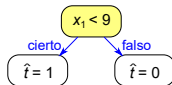


Figura: Árbol de decisión con 1 nodo y dos hojas. [Fuente: Original de A. Cuesta]

- ▶ El proceso podría seguir, en el caso más extremo, hasta tener una hoja por cada ejemplo = SOBREAJUSTE

Criterio de parada

- ▶ Para evitar el sobreajuste conviene imponer un criterio de parada
 - ▶ Por ej. la profundidad máxima del árbol

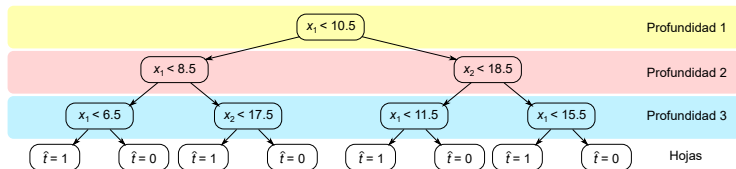


Figura: Árbol de decisión de profundidad 3 [Fuente: Original de A. Cuesta]

Intuición

Pureza de una división

Concatenar decisiones

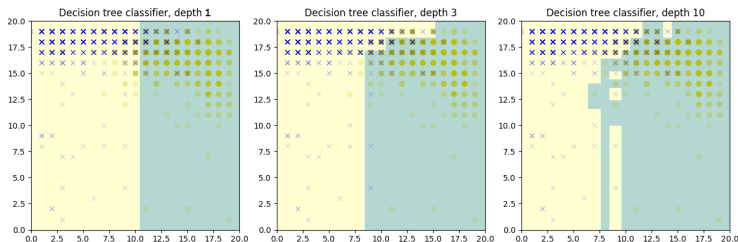
Construcción del árbol

CART

Predicción

Resultado

- Podemos ver la superficie de decisión con el mismo método que hemos utilizado para clasificación no lineal



Intuición

Pureza de una división

Concatenar decisiones

Construcción del árbol

CART

Predicción

Intuición

Pureza de una división

Concatenar decisiones

Construcción del árbol

CART

Predicción

CART

Classification and Regression Tree

- ▶ Método avaricioso (*greedy*) para construir un nodo
- ▶ Dada una característica x_k , y un umbral θ_k , dividimos el conjunto dado en dos:

$$\begin{cases} \text{A la "izquierda" aquellos } \mathbf{x} : x_k \leq \theta_k \\ \text{A la "derecha" el resto} \end{cases}$$

- ▶ Para elegir el mejor par (x_k, θ_k) necesitamos una función de coste
- ▶ En el algoritmo CART dicha función es:

$$J(x_k, \theta_k) = \frac{m_{\text{left}}}{m} G_{\text{left}} + \frac{m_{\text{right}}}{m} G_{\text{right}}, \quad (1)$$

- ▶ $m = m_{\text{left}} + m_{\text{right}}$ número de ejemplos que se están evaluando el nodo que estamos construyendo.
- ▶ m_{left} y m_{right} número de ejemplos en el subconjunto izquierdo y derecho respectivamente.
- ▶ G_{left} y G_{right} coeficientes de impureza de Gini de los subconjuntos izquierdo y derecho respectivamente.

Intuición

Pureza de una división

Concatenar decisiones

Construcción del árbol

CART

Predicción

Coeficiente de impureza de Gini

- ▶ Medida de la *impureza* de un conjunto de datos
- ▶ Suponiendo que hay n clases distintas,
y que dividimos el conjunto de datos en varios subconjuntos,
el coef. de Gini de una cada subconjunto es:

$$G_i = 1 - \sum_{j=1}^n p_{i,j}^2$$

- ▶ i subconjunto del que se quiere averiguar el coef. de Gini
- ▶ j una de las n clases
- ▶ $p_{i,j}$ proporción de ejemplos de la clase j en el subconjunto i
- ▶ Para clasificación binaria:

$$G_{\text{left}} = 1 - p_{\text{left},0}^2 - p_{\text{left},1}^2$$

$$G_{\text{right}} = 1 - p_{\text{right},0}^2 - p_{\text{right},1}^2$$

Intuición

Pureza de una división

Concatenar decisiones

Construcción del árbol

CART

Predicción

Otros algoritmos, otras medidas

- ▶ Hay otros algoritmos para construir el árbol: **ID3, C4.5, C5.0**
- ▶ También es posible utilizar la **Entropía**, en vez de Gini, para elegir el mejor par (x_k, θ_k)
 - ▶ **Entropía** \simeq medida de la capacidad de un subconjunto de subdividirse en dos diferentes.

En física un aumento de entropía se asocia con un aumento del desorden de las partículas y con la capacidad de obtener un trabajo.

- ▶ La expresión de la entropía, con la misma notación que Gini, es:

$$H_i = - \sum_{j=1}^n p_{i,j} \log(p_{i,j}) , \text{ siempre que } p_{i,j} > 0 ;$$

- ▶ y para clasificación binaria:

$$H_{\text{left}} = -p_{\text{left},0} \log(p_{\text{left},0}) - p_{\text{left},1} \log(p_{\text{left},1})$$

$$H_{\text{right}} = -p_{\text{right},0} \log(p_{\text{right},0}) - p_{\text{right},1} \log(p_{\text{right},1})$$

Intuición

Pureza de una división
Concatenar decisiones

Construcción del árbol
CART
Predicción

¿Cómo se produce la etiqueta de un nuevo ejemplo?

- ▶ El nuevo ejemplo entra en el árbol de decisión por el nodo raíz
- ▶ Si se satisface la condición que impone el nodo, se continúa por la derecha. En caso contrario por la izquierda
- ▶ El proceso se repite avanzando por los nodos hasta que se llega a una hoja, que asigna la etiqueta.

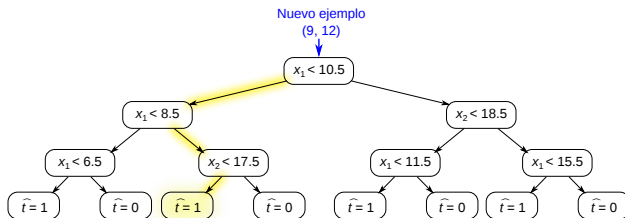


Figura: Clasificación con un árbol de decisión [Fuente: Original de A. Cuesta]

Intuición

Pureza de una división

Concatenar decisiones

Construcción del árbol

CART

Predicción