

Tema 8. Árboles de Clasificacin. Ejercicios

Abdelmalik Moujahid, Iñaki Inza, Pedro Larrañaga
Departamento de Ciencias de la Computación e Inteligencia Artificial
Universidad del País Vasco–Euskal Herriko Unibertsitatea

1. La tabla adjunta contiene 20 casos que constituyen el conjunto de entrenamiento para un clasificador basado en el algoritmo *ID3*. Las variables X_i con $i = 1, \dots, 4$ corresponden a las 4 variables predictoras y la variable C a la variable que tratamos de predecir.

<i>Caso</i>	X_1	X_2	X_3	X_4	C
1	1	2	1	1	1
2	2	2	1	1	1
3	1	2	1	1	1
4	2	2	1	1	1
5	1	1	1	1	1
6	2	2	1	1	1
7	1	1	1	1	1
8	2	2	1	1	1
9	1	1	2	2	1
10	2	1	2	2	1
11	1	2	1	2	2
12	2	2	1	2	2
13	1	2	2	2	2
14	2	2	2	2	2
15	1	2	1	2	2
16	2	2	2	2	2
17	1	2	1	2	2
18	2	1	2	1	2
19	1	1	2	1	2
20	2	2	2	1	2

El problema de clasificación supervisada se relaciona con la predicción de si un alumno aprobará la asignatura $C = \text{Métodos Matemáticos en Ciencias de la Computación}$ en su primera convocatoria, teniendo en cuenta si aprobó o no en primera convocatoria las asignaturas: $X_1 = \text{Análisis Matemático}$, $X_2 = \text{Cálculo}$, $X_3 = \text{Optimización}$, y $X_4 = \text{Probabilidad y Estadística}$. El valor 1 en cualquiera de las 5 variables anteriores indica que la asignatura en cuestión fué aprobada en la primera convocatoria, mientras que el valor 2 hace alusión a que se necesitó de más de una convocatoria para aprobarla.

Obtener el árbol de clasificación a partir del algoritmo *ID3* justificando en cada paso la elección de la variable nodo por medio de la ganancia en información. No es necesario llevar a cabo el proceso de poda previa.

2. Supongamos que una persona aficionada a jugar al tenis ha ido guardando la información de 14 días –en los que se plateó ir a jugar al tenis– relacionada con

el *Tiempo Atmosférico* (X_1), la *Temperatura* (X_2), la *Humedad* (X_3), y el *Viento* (X_4) obteniendo la tabla adjunta.

<i>Dia</i>	X_1	X_2	X_3	X_4	<i>Tenis</i>
1	soleado	calor	alta	debil	no
2	soleado	calor	alta	fuerte	no
3	cubierto	calor	alta	debil	si
4	lluvioso	templada	alta	debil	si
5	lluvioso	frio	normal	debil	si
6	lluvioso	frio	normal	fuerte	no
7	cubierto	frio	normal	fuerte	si
8	soleado	templada	alta	debil	no
9	soleado	frio	normal	debil	si
10	lluvioso	templada	normal	debil	si
11	soleado	templada	normal	fuerte	si
12	cubierto	templada	alta	fuerte	si
13	cubierto	calor	normal	debil	si
14	lluvioso	templada	alta	fuerte	no

Obtener los árboles de clasificación correspondientes a los procedimientos *ID3* y *C4.5*.