

1. Dado el siguiente conjunto de datos (entradas p_1 y p_2 , salida y):

p_1	p_2	y
1	0	1
0	1	0
0	0	1
1	1	0

- a) ¿Cuál es la entropía del conjunto de dato?
- b) Calcule la ganancia sobre el atributo p_2 .
- c) Grafique el árbol que resuelve el problema mediante el algoritmo ID3, puede resolverlo intuitivamente.

Soluciones

a) $-\frac{2}{4} \cdot \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} \cdot \log_2 \left(\frac{2}{4} \right) = 1.$

b) $1 - \left[\frac{2}{4} \cdot \left(-\frac{2}{2} \cdot \log_2 \frac{2}{2} - 0 \right) \right] - \left[\frac{2}{4} \cdot \left(-\frac{0}{2} \cdot \log_2 \frac{0}{2} - 0 \right) \right] = 1.$

- c) COMPLETAR.

2. Considerando el siguiente ejemplo de los Simpsons

Personaje	Longitud Pelo	Peso	Edad	Género
Homero	0	250	36	H
Bart	2	90	10	H
Abe	1	170	70	H
Otto	10	180	38	H
Kruty	6	200	45	H
Marge	10	150	34	M
Lisa	6	78	8	M
Maggie	4	20	1	M
Selma	8	160	41	M
Comic	8	290	38	?

- a) ¿Puede desarrollar un árbol de decisión que utilice sólo dos variables para determinar el género de un personaje en ese contexto? ¿Que valores de corte propondría para esas dos variables?
- b) Resolver en forma intuitiva primero y luego fundamentar con ganancia de información.

Soluciones

- a) Podemos considerar el peso y luego el pelo. Observemos que si el peso mayor a 160 todos los personajes son hombres, luego basta separar al individuo restante por ejemplo si la longitud del pelo es menor que 4.
- b) Observemos que:

$$\blacksquare \text{ entropia}(S) = -\frac{5}{9} \cdot \log_2\left(\frac{5}{9}\right) - \frac{4}{9} \cdot \log_2\left(\frac{4}{9}\right) \approx 0,99.$$

$$\blacksquare \text{ entropia}(\text{peso} > 160) = -\frac{4}{4} \cdot \log_2\left(\frac{4}{4}\right) - 0 = 0.$$

$$\blacksquare \text{ entropia}(\text{peso} \leq 160) = -\frac{1}{5} \cdot \log_2\left(\frac{1}{5}\right) - \frac{4}{5} \cdot \log_2\left(\frac{4}{5}\right) \approx 0,722.$$

$$\blacksquare \text{ entropia}(\text{peso} \leq 160 \wedge \text{pelo} < 4) = -\frac{1}{1} \cdot \log_2\left(\frac{1}{1}\right) - 0 = 0.$$

$$\blacksquare \text{ entropia}(\text{peso} \leq 160 \wedge \text{pelo} \geq 4) = -\frac{4}{4} \cdot \log_2\left(\frac{4}{4}\right) - 0 = 0.$$

