

# TEMA 6: Árboles de Decisión

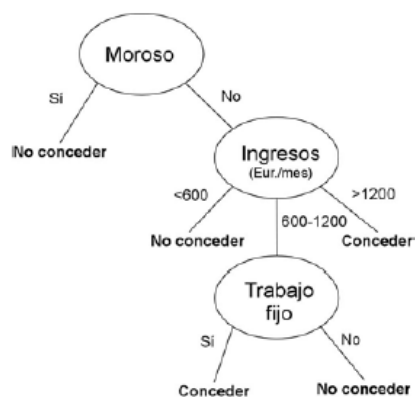
*Toma de decisiones en sistemas probabilísticos*

- Árboles de decisión
  - Planteamiento del problema
  - Ejemplo: Concesión de créditos
  - Entropía y Ganancia de Información
- Algoritmo ID3
  - Algoritmo recursivo
  - Aplicación al ejemplo
  - Consideración de atributos numéricos
  - Atributos con un gran número de valores

## Árboles de decisión

### Características:

- Estructura para **clasificación** de vectores de atributos.
- Establece **en qué orden** testar los atributos para conseguir la clasificación del vector de entrada.
- Para componer dicho orden se eligen primero aquellos atributos que **mejor ganancia de información** prometen a efectos de descubrir la clase del vector de entrada.
- Es interesante **aprenderlos** a partir de un conjunto de vectores



## Ejemplo “Concesión de créditos”

Cliente	Moroso	Antigüedad (años)	Ingresos (Eur./mes)	Trab.fijo	Conceder
1	sí	>5	600-1200	sí	no
2	no	<1	600-1200	sí	sí
3	sí	1-5	>1200	sí	no
4	no	>5	>1200	no	sí
5	no	<1	>1200	sí	sí
6	sí	1-5	600-1200	sí	no
7	no	1-5	>1200	sí	sí
8	no	<1	<600	sí	no
9	no	>5	600-1200	no	no
10	no	1-5	<600	no	no

### Aprendizaje:

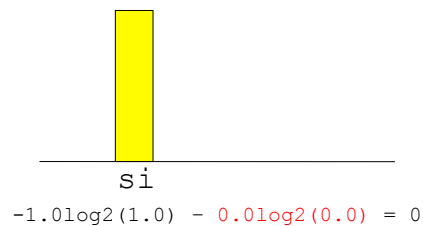
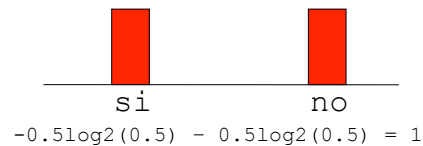
- ¿Por qué atributo **comenzar** primero?
- Esquema voraz: Elegir uno y **filtrar recursivamente**.

## Entropía

### Definición:

- Medida del **grado de incertidumbre** asociado a una distribución de probabilidad.
- En una **distribución uniforme**, todos los valores son igualmente probables  $P_i = 1/N$  y por tanto la **entropía es máxima**, lo cual indica máxima incertidumbre.
- Por el contrario, en una **distribución pico** en la que  $P_i = 1$  y  $P_j = 0$ , para todo  $j \neq i$  la entropía es mínima lo cual indica mínima incertidumbre o sea **máxima información**.

$$E(S) = \sum_{i \in C} -p_i \log_2 p_i$$



## Entropía condicionada

### Definición:

- Entropía de la distribución de Y **condicionada** a X.
- Una entropía condicionada **menor** que  $E(Y)$  indica que el conocimiento de X **mejora la información** que se dispone sobre Y

$$E(Y | X) = \sum_j \text{Prob}(X = v_j) E(Y | X = v_j)$$

$v_i$	$\text{Prob}(X = v_j)$	$E(Y   X = v_j)$
Math	0.5	1
History	0.25	0
CS	0.25	0

X	Y
Math	Yes
Hist.	No
CS	Yes
Math	No
Math	No
CS	Yes
Hist.	No
Math	Yes

$$E(Y) = 1$$

$$E(Y|X) = 0.5 * 1 + 0.25 * 0 + 0.25 * 0$$

$$E(Y|X) = 0.5$$

## Ganancia de información

### Definición:

- Medida de **cuanto ayuda el conocer** el valor de una variable aleatoria  $X$  para conocer el verdadero valor de otra  $Y$ .
- En nuestro caso,  $X$  es un **atributo** de un ejemplo dado mientras que  $Y$  es la **clase** a la que pertenece el ejemplo.
- Una alta ganancia implica que el atributo  $X$  permite **reducir la incertidumbre de la clasificación** del ejemplo de entrada.

$$IG(Y | X) = E(Y) - E(Y | X)$$

X	Y
Math	Yes
History	No
CS	Yes
Math	No
Math	No
CS	Yes
History	No
Math	Yes

$$E(Y) = 1$$

$$E(Y|X) = 0.5$$

$$IG(Y | X) = 1 - 0.5 = 0.5$$

## Algoritmo recursivo

```

Algoritmo ID3(ejemplos, atributos) {
  Si atributos = ∅ o MISMACLASE(ejemplos) {
    C ← CLASEMAYORITARIA(ejemplos)
    N ← CREARNODOHOJA(C)
  }
  Sino {
    a_max ← max_{A ∈ atributos} G(ejemplos, A)
    N ← CREARNODO(a_max)
    Para cada v_i ∈ VALORES(a_max) {
      ejemplos_{v_i} ← {elementos de ejemplos con valor v_i para a_max}
      AÑADIRHIJO(N, ID3(ejemplos_{v_i}, atributos - a_max))
    }
  }
  Devolver N
}

```

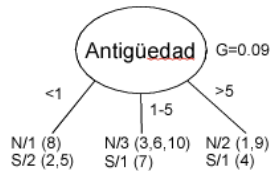
## Aplicación al ejemplo

### Entropía inicial:

- Aplicando la ecuación de entropía a los datos de entrada del ejemplo tenemos:

$$E(S) = -0.41 \log_2(0.4) - 0.61 \log_2(0.6) = 0.971$$

- Para cada atributo (Antigüedad, Moroso, Ingresos, Fijo), calculamos la ganancia de información que obtenemos al seleccionar cada uno de ellos



$$\text{Prob}(S<1)=0.3, \text{Prob}(S1-5)=0.4, \text{Prob}(S>5)=0.3$$

$$E(S<1) = -2/3 \log_2(2/3) - 1/3 \log_2(1/3) = 0.9183$$

$$E(S1-5) = -1/4 \log_2(1/4) - 3/4 \log_2(3/4) = 0.811$$

$$E(S>5) = -1/3 \log_2(1/3) - 2/3 \log_2(2/3) = 0.9183$$

$$E(S<1) * 0.3 = 0.2755$$

$$E(S1-5) * 0.4 = 0.3244$$

$$E(S>5) * 0.3 = 0.2755$$

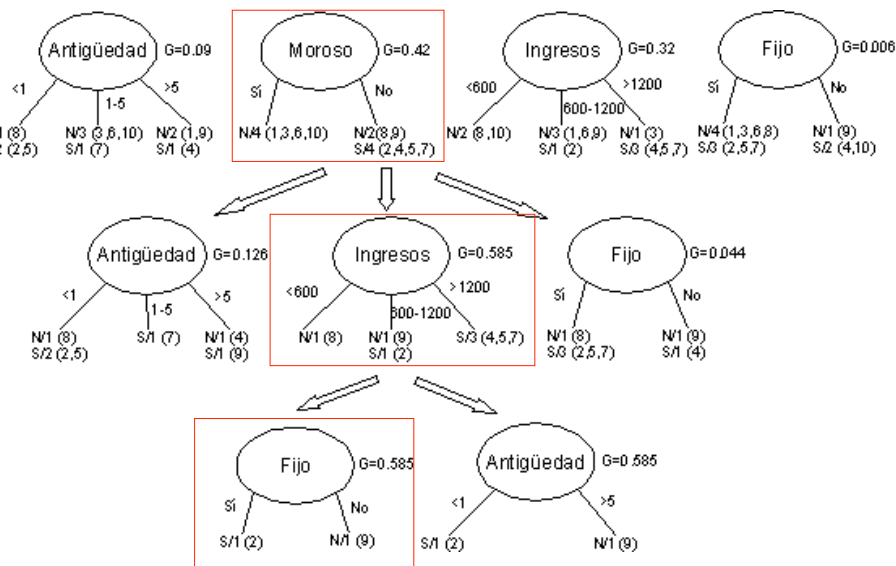
$$H(\text{Conceder} | \text{Antigüedad}) = 0.2755 + 0.3244 + 0.2755 = 0.8754$$

$$\text{Ganancia} = 0.971 - 0.8754 = 0.09$$

Árboles de decisión

9

## Aplicación al ejemplo



Árboles de decisión

10

## Extensiones del algoritmo

### Extensiones:

- **Atributos numéricos:** ID3 sólo trabaja con atributos discretos. Si se usan **atributos continuos** hay que **descomponerlos en rangos**. Para ello se ordenan los ejemplos según el valor y se toman como puntos límite los puntos medios de aquellos en que se cambie de clase.

Ejemplo	8	10	6	2	1	9	3	5	4	7
Ingresos	450	530	650	800	850	1050	1250	1400	1600	3000
Crédito	no	no	no	no	sí	no	sí	sí	sí	sí

- **Atributos con gran número de valores.** Se forman grupos pequeños de ejemplos que pueden ser **homogéneos por casualidad**. Debe introducirse un elemento corrector que penalice atributos con un elevado número de valores (**ganancia normalizada**):

$$G_N(S, A) = \frac{G(S, A)}{\sum_{v_i \in V(A)} -p_{v_i} \log_2 p_{v_i}}$$

- **Sobre-entrenamiento.** Comprobación de capacidad de generalización.

Árboles de decisión

11

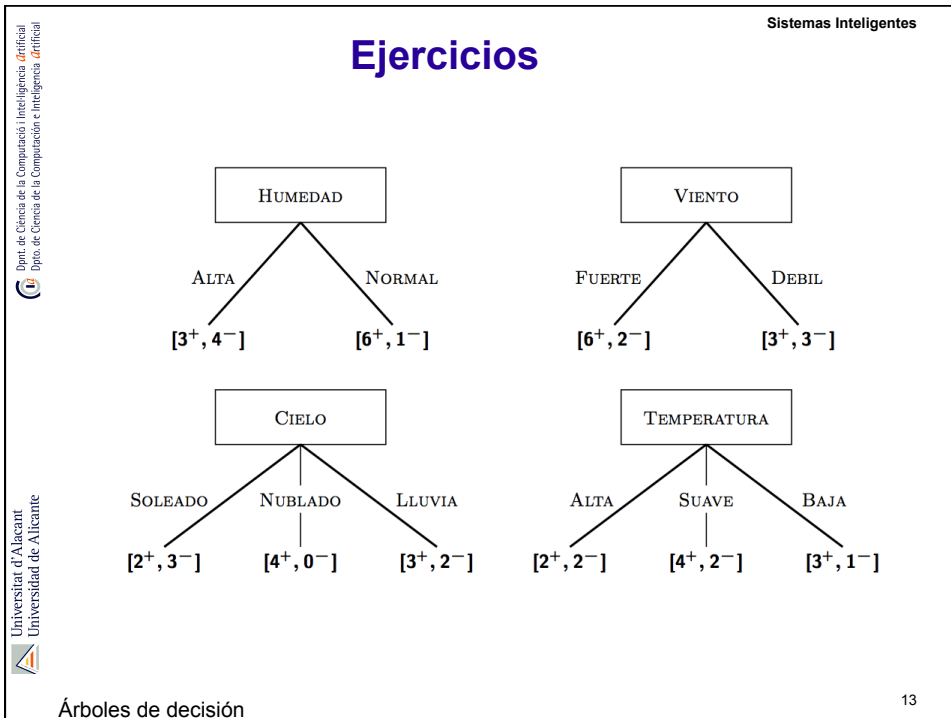
## Ejercicios

Objetivo: Dado el conjunto de entrenamiento, aprender el concepto “Días en los que se juega al tenis” obteniendo el árbol de decisión mediante el algoritmo ID3

EJ.	CIELO	TEMPERATURA	HUMEDAD	VIENTO	JUGAR TENIS
D <sub>1</sub>	SOLEADO	ALTA	ALTA	DÉBIL	-
D <sub>2</sub>	SOLEADO	ALTA	ALTA	FUERTE	-
D <sub>3</sub>	NUBLADO	ALTA	ALTA	DÉBIL	+
D <sub>4</sub>	LLUVIA	SUAVE	ALTA	DÉBIL	+
D <sub>5</sub>	LLUVIA	BAJA	NORMAL	DÉBIL	+
D <sub>6</sub>	LLUVIA	BAJA	NORMAL	FUERTE	-
D <sub>7</sub>	NUBLADO	BAJA	NORMAL	FUERTE	+
D <sub>8</sub>	SOLEADO	SUAVE	ALTA	DÉBIL	-
D <sub>9</sub>	SOLEADO	BAJA	NORMAL	DÉBIL	+
D <sub>10</sub>	LLUVIA	SUAVE	NORMAL	DÉBIL	+
D <sub>11</sub>	SOLEADO	SUAVE	NORMAL	FUERTE	+
D <sub>12</sub>	NUBLADO	SUAVE	ALTA	FUERTE	+
D <sub>13</sub>	NUBLADO	ALTA	NORMAL	DÉBIL	+
D <sub>14</sub>	LLUVIA	SUAVE	ALTA	FUERTE	-

Árboles de decisión

12



Sistemas Inteligentes

## Ejercicios

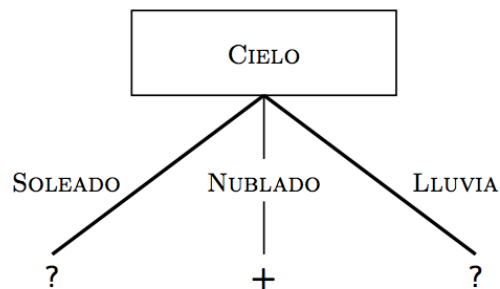
- Entropía inicial:  $\text{Ent}([9^+, 5^-]) = 0,94$
- Selección del atributo para el nodo raíz:
  - $\text{Ganancia}(D, \text{HUMEDAD}) = 0,94 - \frac{7}{14} \cdot \text{Ent}([3^+, 4^-]) - \frac{7}{14} \cdot \text{Ent}([6^+, 1^-]) = 0,151$
  - $\text{Ganancia}(D, \text{VIENTO}) = 0,94 - \frac{8}{14} \cdot \text{Ent}([6^+, 2^-]) - \frac{6}{14} \cdot \text{Ent}([3^+, 3^-]) = 0,048$
  - $\text{Ganancia}(D, \text{CIELO}) = 0,94 - \frac{5}{14} \cdot \text{Ent}([2^+, 3^-]) - \frac{4}{14} \cdot \text{Ent}([4^+, 0^-]) - \frac{5}{14} \cdot \text{Ent}([3^+, 2^-]) = 0,246$  (mejor atributo)
  - $\text{Ganancia}(D, \text{TEMPERATURA}) = 0,94 - \frac{4}{14} \cdot \text{Ent}([2^+, 2^-]) - \frac{6}{14} \cdot \text{Ent}([4^+, 2^-]) - \frac{4}{14} \cdot \text{Ent}([3^+, 1^-]) = 0,02$
- El atributo seleccionado es **CIELO**

Árboles de decisión

14

## Ejercicios

Árbol parcialmente construido:



## Ejercicios

- Selección del atributo para el nodo CIELO=SOLEADO
- $D_{\text{SOLEADO}} = \{D_1, D_2, D_8, D_9, D_{11}\}$  con entropía  $\text{Ent}([2^+, 3^-]) = 0,971$ 
  - $\text{Ganancia}(D_{\text{SOLEADO}}, \text{HUMEDAD}) = 0,971 - \frac{3}{5} \cdot 0 - \frac{2}{5} \cdot 0 = 0,971$  (mejor atributo)
  - $\text{Ganancia}(D_{\text{SOLEADO}}, \text{TEMPERATURA}) = 0,971 - \frac{2}{5} \cdot 0 - \frac{2}{5} \cdot 1 - \frac{1}{5} \cdot 0 = 0,570$
  - $\text{Ganancia}(D_{\text{SOLEADO}}, \text{VIENTO}) = 0,971 - \frac{2}{5} \cdot 1 - \frac{3}{5} \cdot 0,918 = 0,019$
- El atributo seleccionado es HUMEDAD

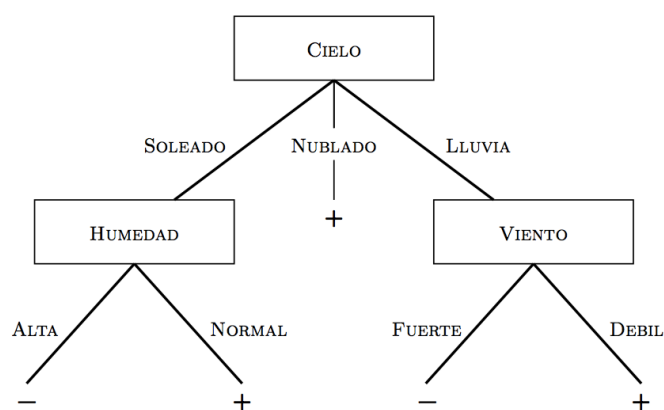


## Ejercicios

- Selección del atributo para el nodo CIELO=LLUVIA:
- $D_{LLUVIA} = \{D_4, D_5, D_6, D_{10}, D_{14}\}$  con entropía  $Ent([3^+, 2^-]) = 0,971$ 
  - $Ganancia(D_{LLUVIA}, HUMEDAD) = 0,971 - \frac{2}{5} \cdot 1 - \frac{3}{5} \cdot 0,918 = 0,820$
  - $Ganancia(D_{LLUVIA}, TEMPERATURA) = 0,971 - \frac{3}{5} \cdot 0,918 - \frac{2}{5} \cdot 1 = 0,820$
  - $Ganancia(D_{LLUVIA}, VIENTO) = 0,971 - \frac{3}{5} \cdot 0 - \frac{2}{5} \cdot 0 = 0,971$  (mejor atributo)
- El atributo seleccionado es VIENTO

## Ejercicios

- Árbol finalmente aprendido:



## Bibliografía

- Escolano et al. [Inteligencia Artificial](#). Thomson-Paraninfo 2003. Capítulo 4.
- Mitchel, [Machine Learning](#). McGraw Hill, Computer Science Series. 1997
- Cover, Thomas, [Information Theory](#). Wiley & Sons, New York 1991