SISTEMAS INTELIGENTES

T9: Árboles de Decisión

www.aic.uniovi.es/ssii

Índice

- Árboles de decisión para clasificación
  - Mecanismo de inducción: divide y vencerás
  - C4.5
  - Evitar el sobreajuste: poda
  - Tratamiento de atributos numéricos y missing
- Reglas:
  - A partir de árboles de decisión
  - Otros métodos
- Árboles de decisión para regresión

Historia

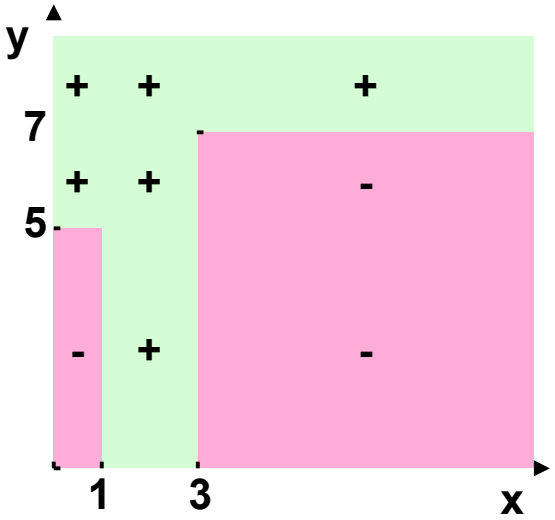
- Concept Learning System [Hunt *et al.* 66]
- CART [Breiman *et al.* 77]: Clasification and Regression Trees
- Árboles de decisión de Quinlan:
  - Clasificación
    - ID3: [Quinlan, 86]
    - C4.5: [Quinlan, 93]
  - Regresión
    - M5: [Quinlan, 95]
    - Cubist: Comercial, Rulequest

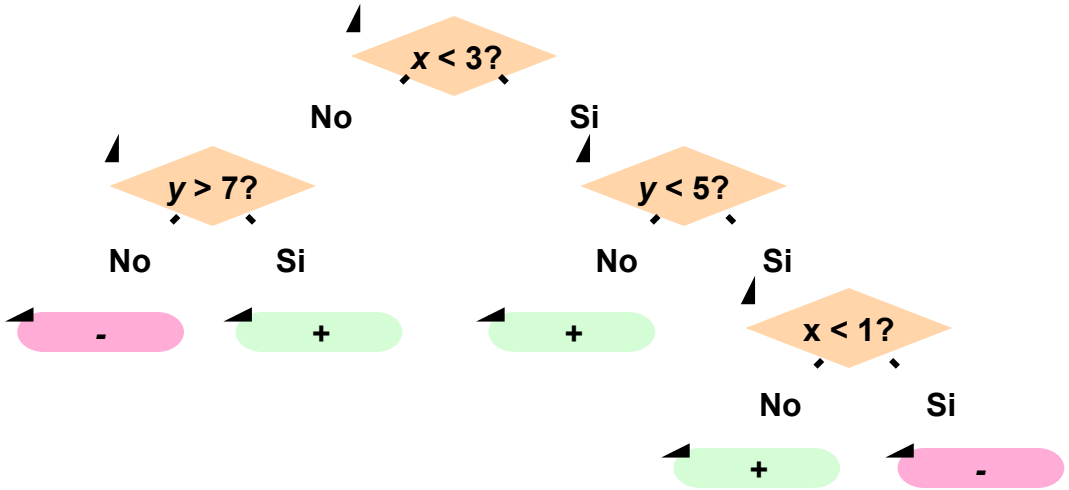
Árboles de decisión

- Los **nodos** están etiquetados con un **test**
  - At. discretos: ¿Cuál es el valor del atributo?
  - At. continuos: El valor del atributo ¿es menor o igual que un cierto valor?
- Las **hojas** están etiquetadas con la **predicción**
  - Clasificación: etiqueta de una clase
  - Regresión: una función dependiente de los atributos (continuos)

Regiones de decisión de los árboles

Dividen el espacio de ejemplos en rectángulos paralelos a los ejes y asignan una clase a cada uno de ellos





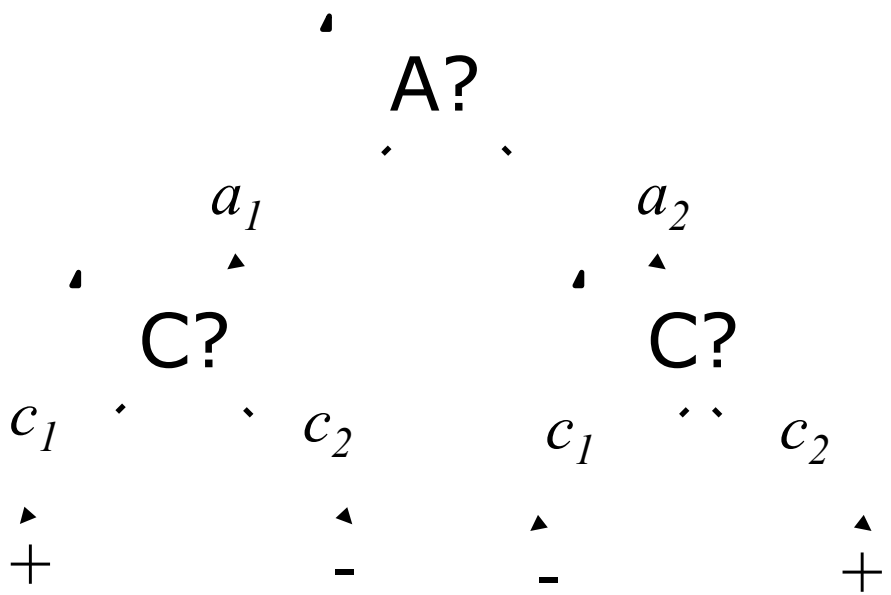
El mecanismo de inducción

- Partimos de un cjto de entrenamiento *E*:
  - ejemplos descritos por una serie de atributos
  - uno de ellos es la clase a predecir
- El procedimiento se basa en el paradigma **divide-y-vencerás**:
  - dividir *E* en subconjuntos homogéneos con respecto a la clase atendiendo a los valores de los atributos en los ejemplos
  - La clave del proceso es descubrir con que atributo (y con que test) se discriminan mejor los ejemplos disponibles

Divide y vencerás (I)

	A	B	C	Clase
1)	a_1	b_1	c_1	+
2)	a_1	b_1	c_2	-
3)	a_1	b_2	c_1	+
4)	a_1	b_3	c_1	+
5)	a_2	b_1	c_1	-
6)	a_2	b_1	c_2	+
7)	a_2	b_2	c_1	-
8)	a_2	b_3	c_1	-

Se divide el conjunto de observaciones, tratando de obtener subconjuntos homogéneos con respecto a la clase



	A	B	C	Clase
1)	a_1	b_1	c_1	+
2)	a_1	b_1	c_2	-
3)	a_1	b_2	c_1	+
4)	a_1	b_3	c_1	+

	A	B	C	Clase
5)	a_2	b_1	c_1	-
6)	a_2	b_1	c_2	+
7)	a_2	b_2	c_1	-
8)	a_2	b_3	c_1	-

Divide y vencerás (II)

- Procedimiento recursivo
  - **Si todos los ejemplos de *E* son de la misma clase**
    ⇒ hoja etiquetada con dicha clase
  - **Si *E* = ∅**
    ⇒ hoja etiquetada con la clase más abundante del nodo padre
  - **Si hay ejemplos de varias clases en *E***
    ⇒ nodo etiquetado con un test sobre los valores de un atributo y tantas ramas como posibles resultados tenga el test. Se divide *E* en subconjuntos disjuntos, uno por cada resultado del test
- Define un procedimiento voraz de escalada: reduce el número de errores en el cjto de entr.

Construir un árbol: C4.5

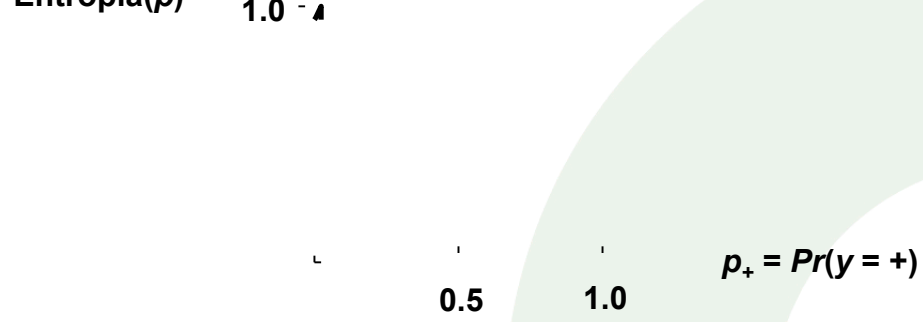
- Objetivo:
  - Construir un árbol de decisión tan pequeño como sea posible (Occam's razor)
  - Sujeto a: que sea consistente con los ejemplos de entrenamiento
- Obstáculos:
  - encontrar el árbol mínimo consistente es NP-duro
  - Algoritmo recursivo:
    - » búsqueda heurística greedy
    - » no se garantiza obtener el árbol óptimo
- Elemento clave: elegir el atributo para la siguiente condición
  - queremos atributos que dividan los ejemplos en conjuntos lo más puros posibles (de una sola clase) (casi nodos hoja)

Teoría de la información

- Ej: X tiene 4 posibles valores equiprobables ¿Cuántos bits hacen falta para transmitir X?
  - P(X=A)=P(X=B)=P(X=C)=P(X=D)=.25
  - Harán falta 2 bits, A = 00, B = 01, C = 10, D = 11
  - BAACBCDD = 01 00 00 10 01 10 11 11 (16 bits)
- P(X=A)=.5 P(X=B)=.25 P(X=C)=P(X=D)=.125
  - Necesitamos 1.75 bits por símbolo (con decimales!)
  - A = 0, B = 10, C = 110, D = 111
  - BAAABCAD = 10 0 0 0 10 110 0 111 (14 bits)
  - 1 x 0.5 + 2 x 0.25 + 3 x 0.125 + 3 x 0.125 = 1.75
- ¿Y cómo se calcula? **Entropía!**

Entropía (I) [Shannon, 48]

- Es una medida de incertidumbre, relacionada con:
  - Pureza: cómo de cerca está un cjto de pertenecer a una misma clase
  - Impureza (desorden): como de cerca está de la total incertidumbre
- La Entropía es una medida:
  - Directamente proporcional a la impureza, incertidumbre, irregularidad
  - Inversamente proporcional a la pureza, certidumbre, redundancia
- Ejemplo: supongamos dos clases { + , - }
  - Pureza óptima: que pertenezcan todos los ejemplos a una clase
    - » P(+) = 1 y P(-) = 0 o también P(-) = 1 y P(+) = 0
  - ¿Cuál es la distribución de probabilidad de menor pureza?
    - » P(+) = 0.5, P(-) = 0.5
- Función cóncava hacia abajo



Entropía (II)

$$Entropia(E) = -\sum_{j=1}^{k} p_j \cdot \log_2 p_j \quad \text{siendo k el número de clases}$$

- Si P(X=A)=.5 P(X=B)=.25 P(X=C)=P(X=D)=.125

$$Entropia(X) = -\sum_{j=1}^{k} p_j \cdot \log_2 p_j$$

$$Entropia(X) = -0.5 \cdot \log_2 0.5 - 0.25 \cdot \log_2 0.25 - 0.125 \cdot \log_2 0.125 - 0.125 \cdot \log_2 0.125 =$$
$$= -0.5 \cdot (-1) - 0.25 \cdot (-2) - 0.125 \cdot (-3) - 0.125 \cdot (-3) =$$
$$= 0.5 + 0.5 + 0.375 + 0.375 = 1.75 \text{ bits}$$

Muchos sistemas de Aprendizaje Automático utilizan la entropía para seleccionar hipótesis, ya que expresa la cantidad de información necesaria para transmitir la información que codifica una hipótesis

Selección del mejor test

C4.5 utiliza un concepto denominado **ganancia de información** que se basa en la entropía

$$\text{info}(E) = -\sum_{j=1}^{k} p_j \cdot \log_2 p_j = -\sum_{j=1}^{k} \frac{\text{frec}(C_j, E)}{|E|} \cdot \log_2 \left(\frac{\text{frec}(C_j, E)}{|E|} \right)$$

$$\text{info\_atrib}(E, X) = \sum_{i=1}^{n} \frac{|E_i|}{|E|} \cdot \text{info}(E_i)$$

$$\text{ganancia}(E, X) = \text{info}(E) - \text{info\_atrib}(E, X)$$

$$\text{ratio}(E, X) = \frac{\text{ganancia}(E, X)}{\text{info\_part}(E, X)}$$

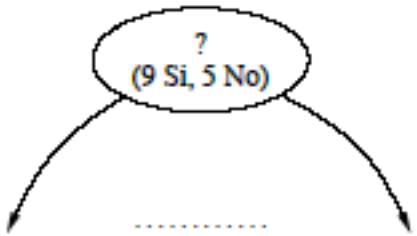
$$\text{info\_part}(E, X) = -\sum_{i=1}^{n} \frac{|E_i|}{|E|} \cdot \log_2 \left(\frac{|E_i|}{|E|} \right)$$

Un árbol paso a paso

	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
1)	Soleado	Alta	Alta	Flojo	No
2)	Soleado	Alta	Alta	Fuerte	No
3)	Nublado	Alta	Alta	Flojo	Si
4)	Lluvia	Moderada	Alta	Flojo	Si
5)	Lluvia	Baja	Normal	Flojo	Si
6)	Lluvia	Baja	Normal	Fuerte	No
7)	Nublado	Baja	Normal	Fuerte	Si
8)	Soleado	Moderada	Alta	Flojo	No
9)	Soleado	Baja	Normal	Flojo	Si
10)	Lluvia	Moderada	Normal	Flojo	Si
11)	Soleado	Moderada	Normal	Fuerte	Si
12)	Nublado	Moderada	Alta	Fuerte	Si
13)	Nublado	Alta	Normal	Flojo	Si
14)	Lluvia	Moderada	Alta	Fuerte	No

Un árbol paso a paso

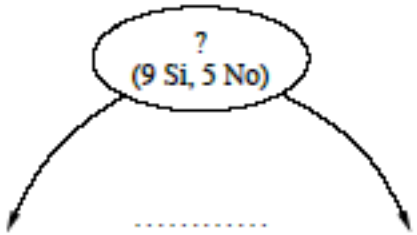
	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
1)	Soleado	Alta	Alta	Flojo	No
2)	Soleado	Alta	Alta	Fuerte	No
3)	Nublado	Alta	Alta	Flojo	Si
4)	Lluvia	Moderada	Alta	Flojo	Si
5)	Lluvia	Baja	Normal	Flojo	Si
6)	Lluvia	Baja	Normal	Fuerte	No
7)	Nublado	Baja	Normal	Fuerte	Si
8)	Soleado	Moderada	Alta	Flojo	No
9)	Soleado	Baja	Normal	Flojo	Si
10)	Lluvia	Moderada	Normal	Flojo	Si
11)	Soleado	Moderada	Normal	Fuerte	Si
12)	Nublado	Moderada	Alta	Fuerte	Si
13)	Nublado	Alta	Normal	Flojo	Si
14)	Lluvia	Moderada	Alta	Fuerte	No



$$\text{info}(E) = -\sum_{j=1}^{k} \frac{\text{frec}(C_j, E)}{|E|} \cdot \log_2 \left(\frac{\text{frec}(C_j, E)}{|E|} \right)$$

Un árbol paso a paso

	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
1)	Soleado	Alta	Alta	Flojo	No
2)	Soleado	Alta	Alta	Fuerte	No
3)	Nublado	Alta	Alta	Flojo	Si
4)	Lluvia	Moderada	Alta	Flojo	Si
5)	Lluvia	Baja	Normal	Flojo	Si
6)	Lluvia	Baja	Normal	Fuerte	No
7)	Nublado	Baja	Normal	Fuerte	Si
8)	Soleado	Moderada	Alta	Flojo	No
9)	Soleado	Baja	Normal	Flojo	Si
10)	Lluvia	Moderada	Normal	Flojo	Si
11)	Soleado	Moderada	Normal	Fuerte	Si
12)	Nublado	Moderada	Alta	Fuerte	Si
13)	Nublado	Alta	Normal	Flojo	Si
14)	Lluvia	Moderada	Alta	Fuerte	No

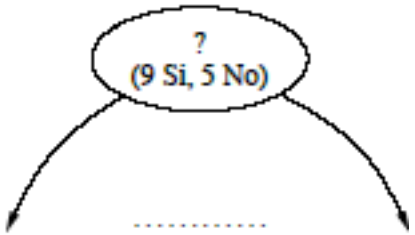


$$\text{info}(E) = -\sum_{j=1}^{k} \frac{\text{frec}(C_j, E)}{|E|} \cdot \log_2 \left(\frac{\text{frec}(C_j, E)}{|E|} \right)$$

$$\text{info}(E) = -\left[\frac{9}{14} \cdot \log_2 \left(\frac{9}{14} \right) + \frac{5}{14} \cdot \log_2 \left(\frac{5}{14} \right) \right] = 0{,}940$$

Un árbol paso a paso

	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
1)	Soleado	Alta	Alta	Flojo	No
2)	Soleado	Alta	Alta	Fuerte	No
3)	Nublado	Alta	Alta	Flojo	Si
4)	Lluvia	Moderada	Alta	Flojo	Si
5)	Lluvia	Baja	Normal	Flojo	Si
6)	Lluvia	Baja	Normal	Fuerte	No
7)	Nublado	Baja	Normal	Fuerte	Si
8)	Soleado	Moderada	Alta	Flojo	No
9)	Soleado	Baja	Normal	Flojo	Si
10)	Lluvia	Moderada	Normal	Flojo	Si
11)	Soleado	Moderada	Normal	Fuerte	Si
12)	Nublado	Moderada	Alta	Fuerte	Si
13)	Nublado	Alta	Normal	Flojo	Si
14)	Lluvia	Moderada	Alta	Fuerte	No

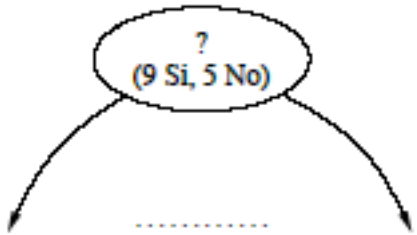


$$\text{info}(E) = -\sum_{j=1}^{k} \frac{\text{frec}(C_j, E)}{|E|} \cdot \log_2 \left(\frac{\text{frec}(C_j, E)}{|E|} \right)$$

$$\text{info}(E) = -\left[\frac{9}{14} \cdot \log_2 \left(\frac{9}{14} \right) + \frac{5}{14} \cdot \log_2 \left(\frac{5}{14} \right) \right] = 0{,}940$$

Un árbol paso a paso

	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
1)	Soleado	Alta	Alta	Flojo	No
2)	Soleado	Alta	Alta	Fuerte	No
3)	Nublado	Alta	Alta	Flojo	Si
4)	Lluvia	Moderada	Alta	Flojo	Si
5)	Lluvia	Baja	Normal	Flojo	Si
6)	Lluvia	Baja	Normal	Fuerte	No
7)	Nublado	Baja	Normal	Fuerte	Si
8)	Soleado	Moderada	Alta	Flojo	No
9)	Soleado	Baja	Normal	Flojo	Si
10)	Lluvia	Moderada	Normal	Flojo	Si
11)	Soleado	Moderada	Normal	Fuerte	Si
12)	Nublado	Moderada	Alta	Fuerte	Si
13)	Nublado	Alta	Normal	Flojo	Si
14)	Lluvia	Moderada	Alta	Fuerte	No



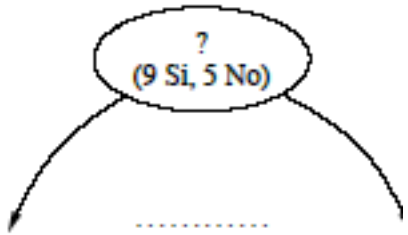
$$\text{info}(E_{Pron.=Soleado}) = -\left[\frac{2}{5}\cdot\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\cdot\log_2\left(\frac{3}{5}\right)\right] = 0{,}971$$

$$\text{info}(E_{Pron.=Nublado}) = -\left[\frac{4}{4}\cdot\log_2\left(\frac{4}{4}\right) + \frac{0}{4}\cdot\log_2\left(\frac{0}{4}\right)\right] = 0$$

$$\text{info}(E_{Pron.=Lluvia}) = -\left[\frac{3}{5}\cdot\log_2\left(\frac{3}{5}\right) + \frac{2}{5}\cdot\log_2\left(\frac{2}{5}\right)\right] = 0{,}971$$

Un árbol paso a paso

	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
1)	Soleado	Alta	Alta	Flojo	No
2)	Soleado	Alta	Alta	Fuerte	No
	.	.	.	.	
	.	.	.	.	
	.	.	.	.	
8)	Soleado	Moderada	Alta	Flojo	No
9)	Soleado	Baja	Normal	Flojo	Si
	.	.	.	.	
11)	Soleado	Moderada	Normal	Fuerte	Si
	.	.	.	.	



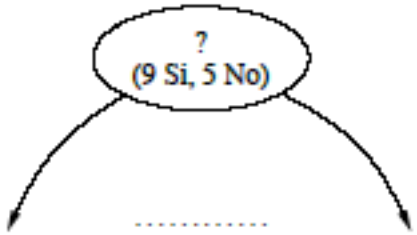
$$\text{info}(E_{Pron.=Soleado}) = -\left[\frac{2}{5}\cdot\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\cdot\log_2\left(\frac{3}{5}\right)\right] = 0{,}971$$

$$\text{info}(E_{Pron.=Nublado}) = -\left[\frac{4}{4}\cdot\log_2\left(\frac{4}{4}\right) + \frac{0}{4}\cdot\log_2\left(\frac{0}{4}\right)\right] = 0$$

$$\text{info}(E_{Pron.=Lluvia}) = -\left[\frac{3}{5}\cdot\log_2\left(\frac{3}{5}\right) + \frac{2}{5}\cdot\log_2\left(\frac{2}{5}\right)\right] = 0{,}971$$

Un árbol paso a paso

	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
3)	Nublado	Alta	Alta	Flojo	Si
7)	Nublado	Baja	Normal	Fuerte	Si
12)	Nublado	Moderada	Alta	Fuerte	Si
13)	Nublado	Alta	Normal	Flojo	Si



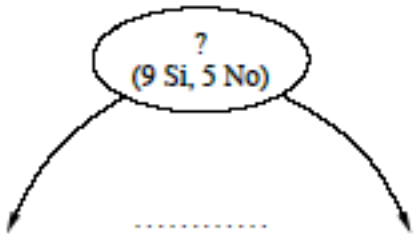
$$\text{info}(E_{Pron.=Soleado}) = -\left[\frac{2}{5}\cdot\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\cdot\log_2\left(\frac{3}{5}\right)\right] = 0{,}971$$

$$\text{info}(E_{Pron.=Nublado}) = -\left[\frac{4}{4}\cdot\log_2\left(\frac{4}{4}\right) + \frac{0}{4}\cdot\log_2\left(\frac{0}{4}\right)\right] = 0$$

$$\text{info}(E_{Pron.=Lluvia}) = -\left[\frac{3}{5}\cdot\log_2\left(\frac{3}{5}\right) + \frac{2}{5}\cdot\log_2\left(\frac{2}{5}\right)\right] = 0{,}971$$

Un árbol paso a paso

	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
4)	Lluvia	Moderada	Alta	Flojo	Si
5)	Lluvia	Baja	Normal	Flojo	Si
6)	Lluvia	Baja	Normal	Fuerte	No
10)	Lluvia	Moderada	Normal	Flojo	Si
14)	Lluvia	Moderada	Alta	Fuerte	No



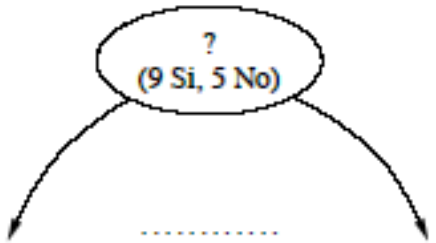
$$\text{info}(E_{Pron.=Soleado}) = -\left[\frac{2}{5}\cdot\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\cdot\log_2\left(\frac{3}{5}\right)\right] = 0{,}971$$

$$\text{info}(E_{Pron.=Nublado}) = -\left[\frac{4}{4}\cdot\log_2\left(\frac{4}{4}\right) + \frac{0}{4}\cdot\log_2\left(\frac{0}{4}\right)\right] = 0$$

$$\text{info}(E_{Pron.=Lluvia}) = -\left[\frac{3}{5}\cdot\log_2\left(\frac{3}{5}\right) + \frac{2}{5}\cdot\log_2\left(\frac{2}{5}\right)\right] = 0{,}971$$

Un árbol paso a paso

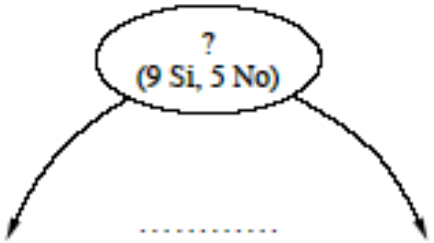
	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
1)	Soleado	Alta	Alta	Flojo	No
2)	Soleado	Alta	Alta	Fuerte	No
3)	Nublado	Alta	Alta	Flojo	Si
4)	Lluvia	Moderada	Alta	Flojo	Si
5)	Lluvia	Baja	Normal	Flojo	Si
6)	Lluvia	Baja	Normal	Fuerte	No
7)	Nublado	Baja	Normal	Fuerte	Si
8)	Soleado	Moderada	Alta	Flojo	No
9)	Soleado	Baja	Normal	Flojo	Si
10)	Lluvia	Moderada	Normal	Flojo	Si
11)	Soleado	Moderada	Normal	Fuerte	Si
12)	Nublado	Moderada	Alta	Fuerte	Si
13)	Nublado	Alta	Normal	Flojo	Si
14)	Lluvia	Moderada	Alta	Fuerte	No



$$\text{info\_atrib}(E, Pronóstico) = \frac{|E_{Pron.=Soleado}|}{|E|} \cdot \text{info}(E_{Pron.=Soleado}) +$$
$$+ \frac{|E_{Pron.=Nublado}|}{|E|} \cdot \text{info}(E_{Pron.=Nublado}) + \frac{|E_{Pron.=Lluvia}|}{|E|} \cdot \text{info}(E_{Pron.=Lluvia}) =$$
$$= \frac{5}{14} \cdot 0{,}971 + \frac{4}{14} \cdot 0 + \frac{5}{14} \cdot 0{,}971 = 0{,}694$$

Un árbol paso a paso

	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
1)	Soleado	Alta	Alta	Flojo	No
2)	Soleado	Alta	Alta	Fuerte	No
3)	Nublado	Alta	Alta	Flojo	Si
4)	Lluvia	Moderada	Alta	Flojo	Si
5)	Lluvia	Baja	Normal	Flojo	Si
6)	Lluvia	Baja	Normal	Fuerte	No
7)	Nublado	Baja	Normal	Fuerte	Si
8)	Soleado	Moderada	Alta	Flojo	No
9)	Soleado	Baja	Normal	Flojo	Si
10)	Lluvia	Moderada	Normal	Flojo	Si
11)	Soleado	Moderada	Normal	Fuerte	Si
12)	Nublado	Moderada	Alta	Fuerte	Si
13)	Nublado	Alta	Normal	Flojo	Si
14)	Lluvia	Moderada	Alta	Fuerte	No



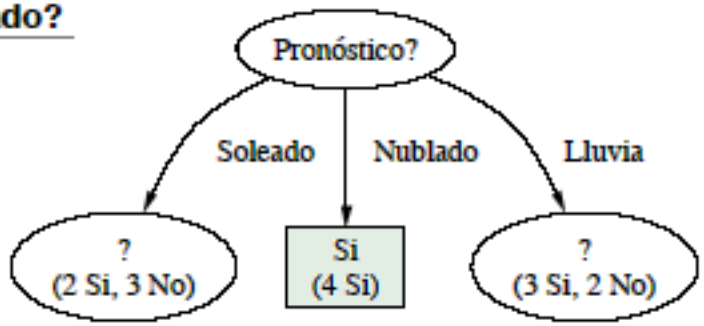
$$\text{ganancia}(E, Pronóstico) = \text{info}(E) - \text{info\_atrib}(E, Pronóstico) = 0{,}940 - 0{,}694 = 0{,}246$$

$$\text{info\_part}(E, Pronóstico) = -\left[\frac{5}{14}\log_2\left(\frac{5}{14}\right) + \frac{4}{14}\log_2\left(\frac{4}{14}\right) + \frac{5}{14}\log_2\left(\frac{5}{14}\right)\right] = 1{,}577$$

$$\text{ratio}(E, Pronóstico) = \frac{\text{ganancia}(E, Pronóstico)}{\text{info\_part}(E, Pronóstico)} = \frac{0{,}246}{1{,}577} = \boxed{0{,}156}$$

Un árbol paso a paso

	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
1)	Soleado	Alta	Alta	Flojo	No
2)	Soleado	Alta	Alta	Fuerte	No
3)	Nublado	Alta	Alta	Flojo	Si
4)	Lluvia	Moderada	Alta	Flojo	Si
5)	Lluvia	Baja	Normal	Flojo	Si
6)	Lluvia	Baja	Normal	Fuerte	No
7)	Nublado	Baja	Normal	Fuerte	Si
8)	Soleado	Moderada	Alta	Flojo	No
9)	Soleado	Baja	Normal	Flojo	Si
10)	Lluvia	Moderada	Normal	Flojo	Si
11)	Soleado	Moderada	Normal	Fuerte	Si
12)	Nublado	Moderada	Alta	Fuerte	Si
13)	Nublado	Alta	Normal	Flojo	Si
14)	Lluvia	Moderada	Alta	Fuerte	No



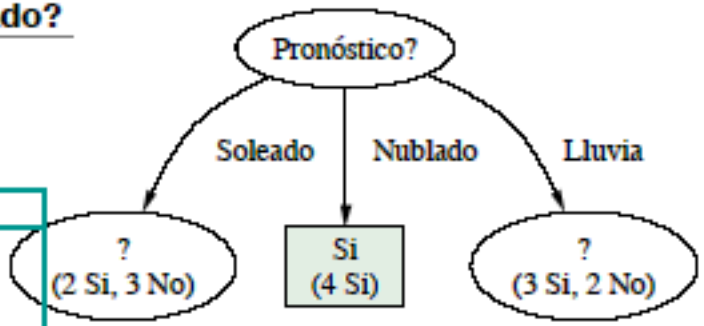
$$\text{ganancia}(E, \textit{Pronóstico}) = \text{info}(E) - \text{info\_atrib}(E, \textit{Pronóstico}) = 0{,}940 - 0{,}694 = 0{,}246$$

$$\text{info\_part}(E, \textit{Pronóstico}) = -\left[\frac{5}{14}\log_2\left(\frac{5}{14}\right) + \frac{4}{14}\log_2\left(\frac{4}{14}\right) + \frac{5}{14}\log_2\left(\frac{5}{14}\right)\right] = 1{,}577$$

$$\text{ratio}(E, \textit{Pronóstico}) = \frac{\text{ganancia}(E, \textit{Pronóstico})}{\text{info\_part}(E, \textit{Pronóstico})} = \frac{0{,}246}{1{,}577} = \boxed{0{,}156}$$

Un árbol paso a paso

	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
1)	Soleado	Alta	Alta	Flojo	No
2)	Soleado	Alta	Alta	Fuerte	No
8)	Soleado	Moderada	Alta	Flojo	No
9)	Soleado	Baja	Normal	Flojo	Si
11)	Soleado	Moderada	Normal	Fuerte	Si



$$\text{info}(E') = -\left[\frac{2}{5}\cdot\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\cdot\log_2\left(\frac{3}{5}\right)\right] = 0{,}971$$

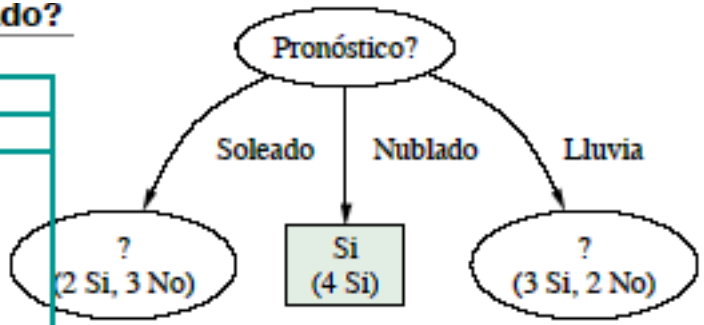
$$\text{info\_atrib}(E', Humedad) =$$

$$= \frac{|E'_{Hum.=Normal}|}{|E'|}\cdot\text{info}(E'_{Hum.=Normal}) + \frac{|E'_{Hum.=Alta}|}{|E'|}\cdot\text{info}(E'_{Hum.=Alta}) =$$

$$= \frac{2}{5}\left[\frac{2}{2}\log_2\left(\frac{2}{2}\right) + \frac{0}{2}\log_2\left(\frac{0}{2}\right)\right] + \frac{3}{5}\left[\frac{0}{3}\log_2\left(\frac{0}{3}\right) + \frac{3}{3}\log_2\left(\frac{3}{3}\right)\right] = 0$$

Un árbol paso a paso

	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
1)	Soleado	Alta	Alta	Flojo	No
2)	Soleado	Alta	Alta	Fuerte	No
8)	Soleado	Moderada	Alta	Flojo	No
9)	Soleado	Baja	Normal	Flojo	Si
11)	Soleado	Moderada	Normal	Fuerte	Si

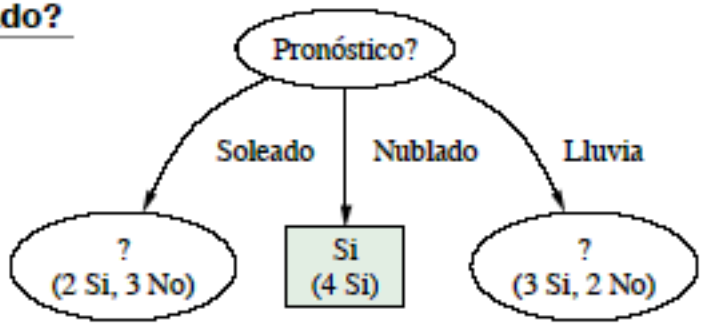


$$\text{info}(E') = -\left[\frac{2}{5}\cdot\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\cdot\log_2\left(\frac{3}{5}\right)\right] = 0{,}971$$

$$\text{info\_atrib}(E', Humedad) =$$
$$= \frac{|E'_{Hum.=Normal}|}{|E'|}\cdot\text{info}(E'_{Hum.=Normal}) + \frac{|E'_{Hum.=Alta}|}{|E'|}\cdot\text{info}(E'_{Hum.=Alta}) =$$
$$= \frac{2}{5}\left[\frac{2}{2}\log_2\left(\frac{2}{2}\right) + \frac{0}{2}\log_2\left(\frac{0}{2}\right)\right] + \frac{3}{5}\left[\frac{0}{3}\log_2\left(\frac{0}{3}\right) + \frac{3}{3}\log_2\left(\frac{3}{3}\right)\right] = 0$$

Un árbol paso a paso

	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
1)	Soleado	Alta	Alta	Flojo	No
2)	Soleado	Alta	Alta	Fuerte	No
8)	Soleado	Moderada	Alta	Flojo	No
9)	Soleado	Baja	Normal	Flojo	Si
11)	Soleado	Moderada	Normal	Fuerte	Si



$$\text{info}(E') = -\left[\frac{2}{5}\cdot\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\cdot\log_2\left(\frac{3}{5}\right)\right] = 0{,}971$$

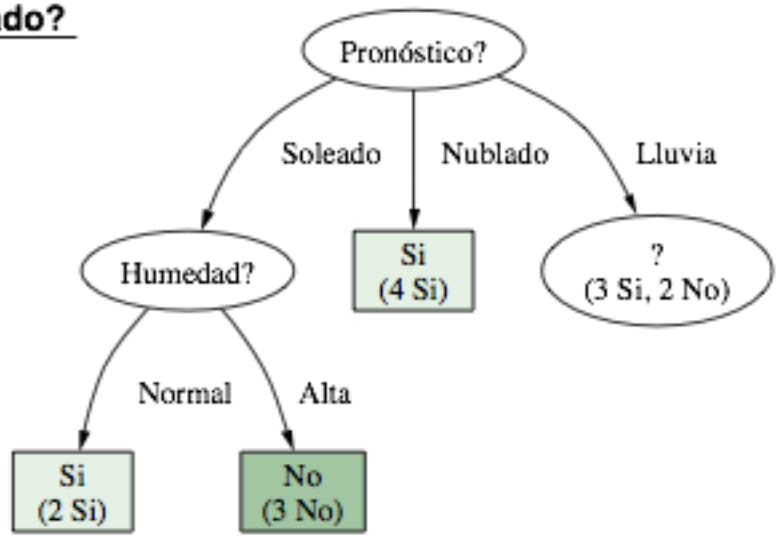
$$\text{ganancia}(E', \mathit{Humedad}) = 0{,}971 - 0 = 0{,}971$$

$$\text{info\_part}(E', \mathit{Humedad}) = -\left[\frac{2}{5}\cdot\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\cdot\log_2\left(\frac{3}{5}\right)\right] = 0{,}971$$

$$\text{ratio}(E', \mathit{Humedad}) = \frac{0{,}971}{0{,}971} = \boxed{1{,}0}$$

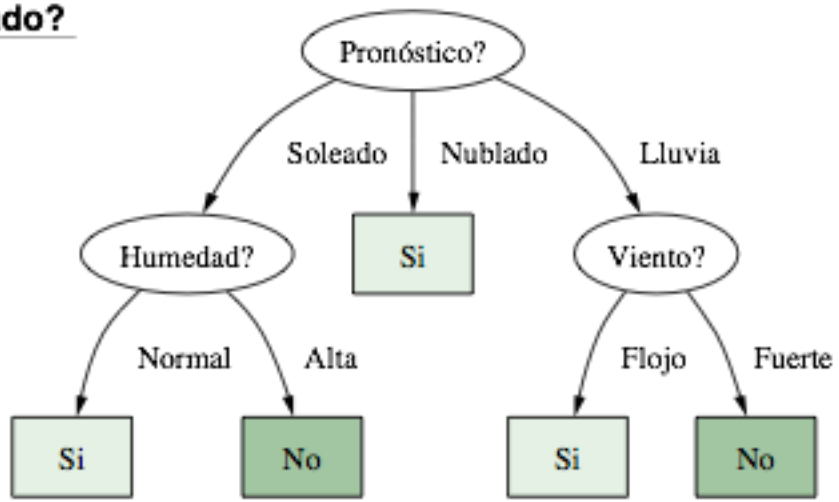
Un árbol paso a paso

	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
1)	Soleado	Alta	Alta	Flojo	No
2)	Soleado	Alta	Alta	Fuerte	No
8)	Soleado	Moderada	Alta	Flojo	No
9)	Soleado	Baja	Normal	Flojo	Si
11)	Soleado	Moderada	Normal	Fuerte	Si



Un árbol paso a paso

	Pronóstico	Temperatura	Humedad	Viento	Adecuado?
4)	Lluvia	Moderada	Alta	Flojo	Si
5)	Lluvia	Baja	Normal	Flojo	Si
6)	Lluvia	Baja	Normal	Fuerte	No
10)	Lluvia	Moderada	Normal	Flojo	Si
14)	Lluvia	Moderada	Alta	Fuerte	No



Mecanismos de poda (I)

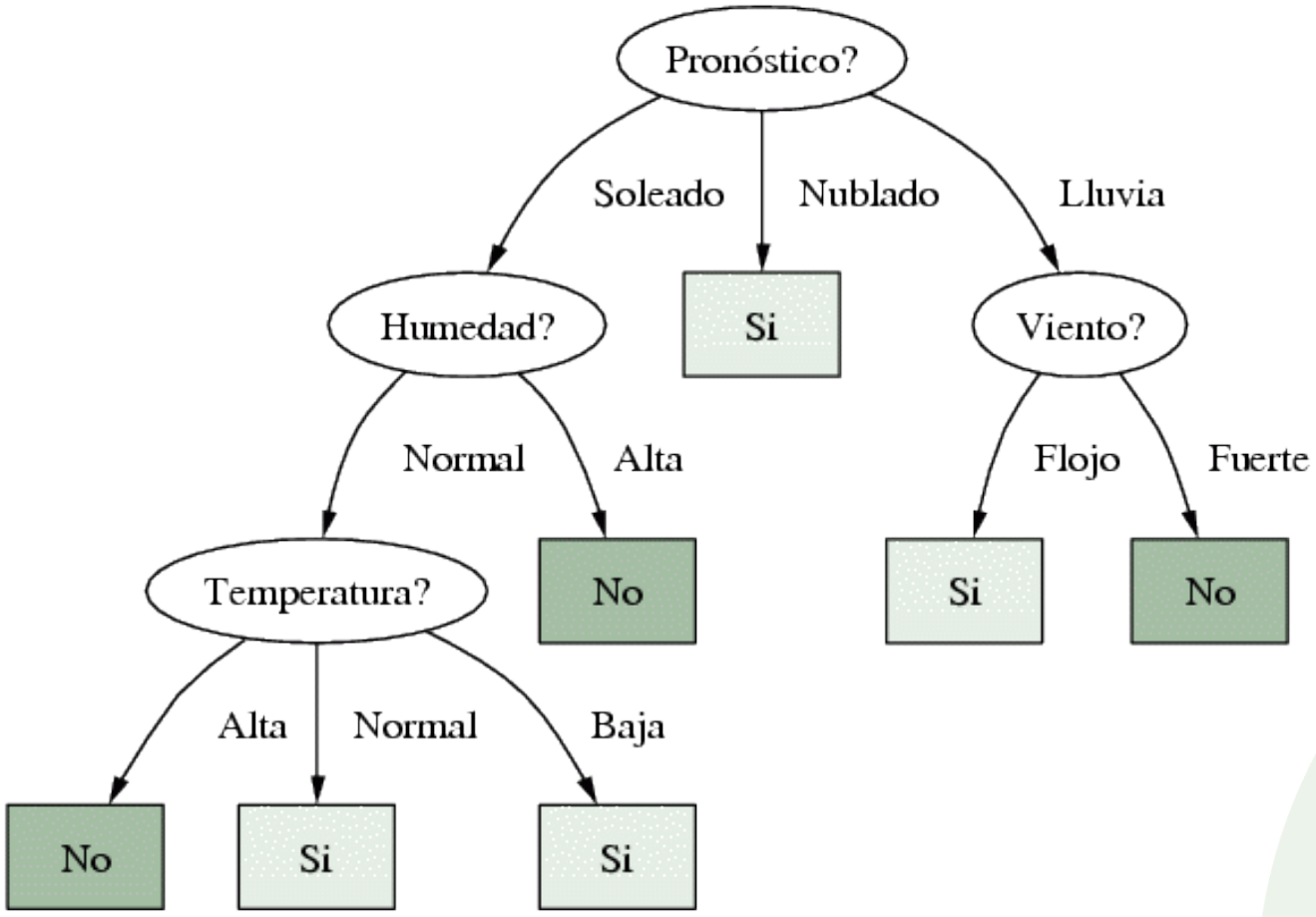
- Tratan de paliar el efecto del **sobreajuste**
- Ejemplo: añadimos al cjto. de entrenamiento el ejemplo ruidoso

‘

Pronóstico=Soleado,
Temperatura=Alta,
Humedad=Normal,
Viento=Fuerte,
Adecuado?=No

Ajuste al ruido

- El mecanismos hasta ahora descrito induce un árbol que se ajustaría perfectamente a los datos, incluido el ruido

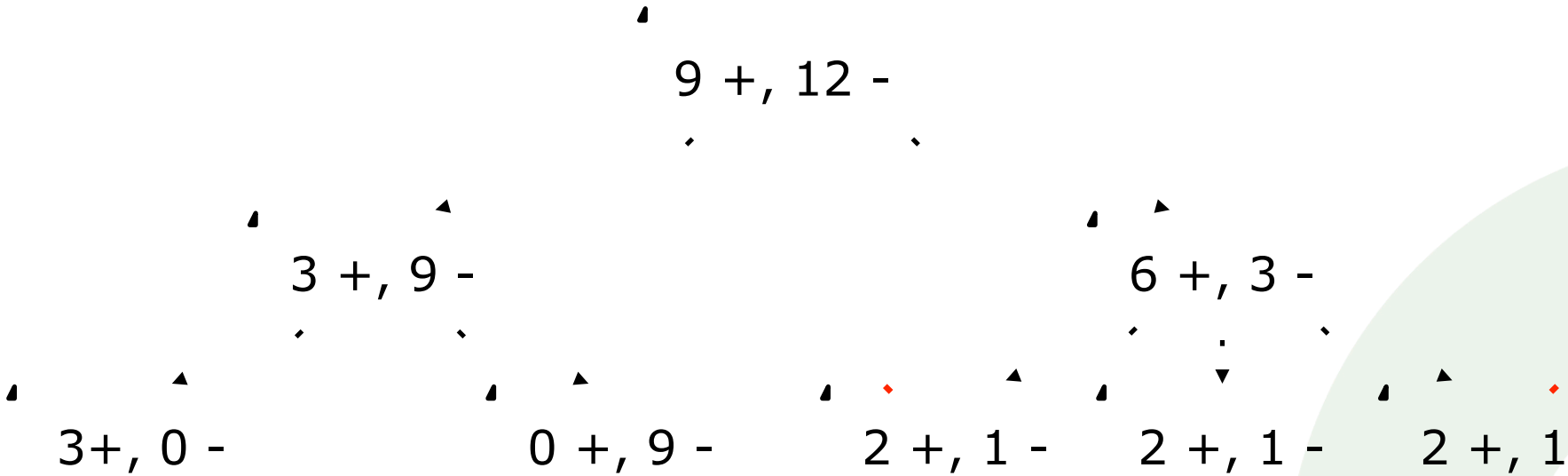


Mecanismos de poda (II)

- Pueden clasificarse en
  - técnicas de ***pre-poda***:
    - detener el crecimiento del árbol antes de que llegue a adaptarse perfectamente al conjunto de entrenamiento
  - técnicas de ***post-poda***:
    - permiten que el árbol se sobreajuste a los datos y luego se efectúa sobre él una poda
    - Tratan de compensar la falta de backtracking del proceso de inducción

La poda en C4.5 (I)

- Pre-poda (mínima)
  - la suma de errores de las ramas resultantes es mayor o igual que el error de una hoja (cuya predicción sería la clase mayoritaria)
  - tras la división no haya dos o más subconjuntos con al menos dos ejemplos



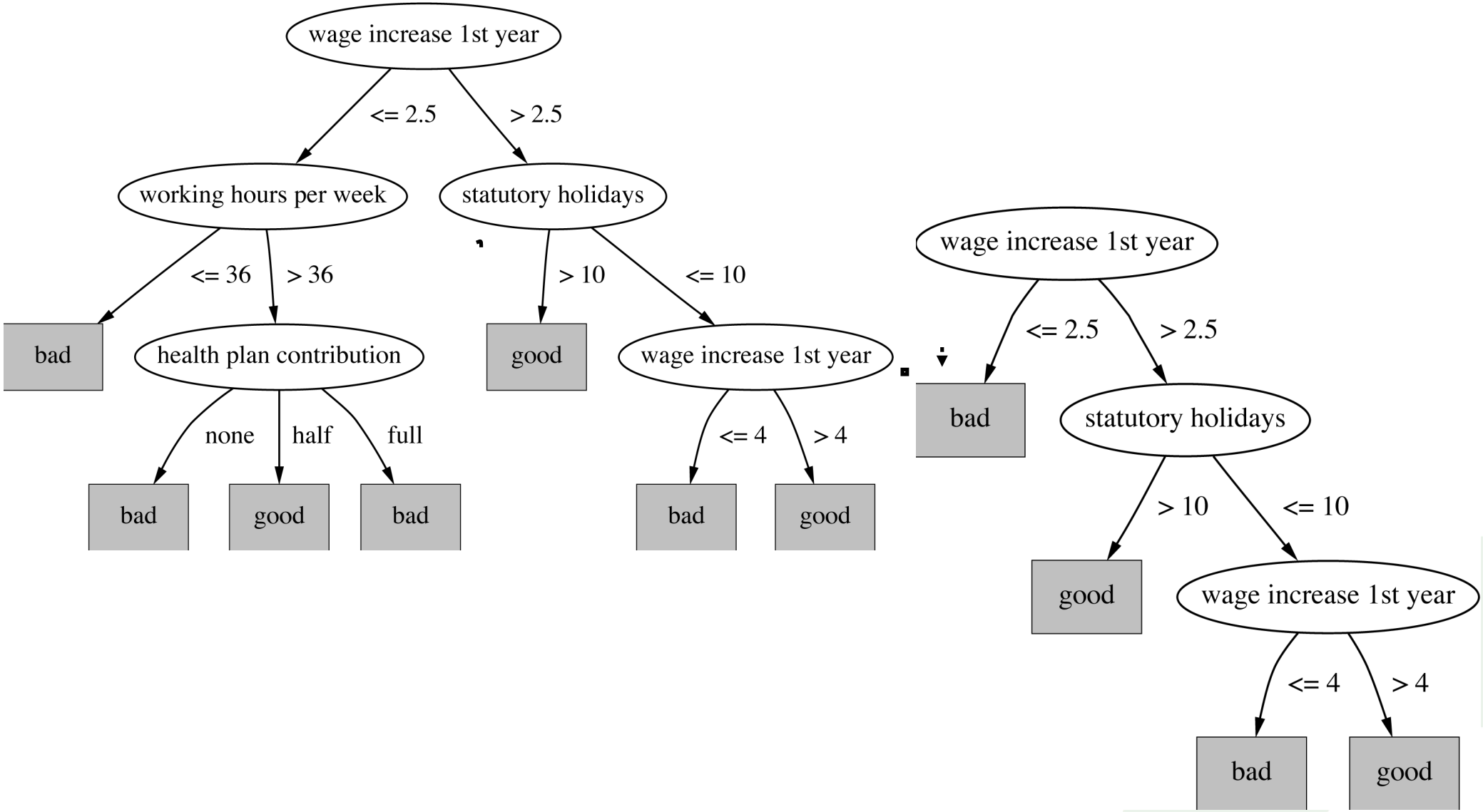
La poda en C4.5 (II)

- Post-poda
  - Se recorre el árbol desde las hojas hacia la raíz, sustituyendo algunos nodos si aumenta la **estimación de la precisión**
    - » Número de *errores estimados* (pesimista) en una hoja (n ejemplos, f fallos):

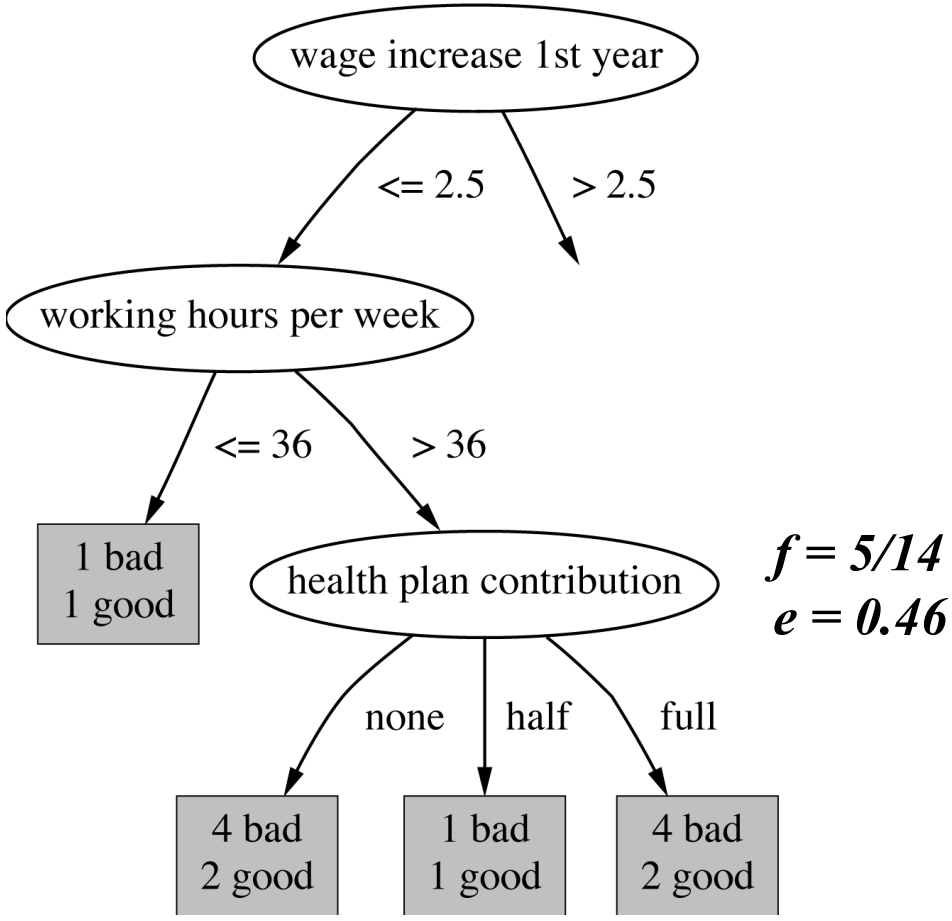
$$n \cdot IC^{\alpha}_{\text{sup}}\left(n, \frac{f}{n}\right)$$

    - » Para un subárbol se utiliza la suma de errores estimados de sus descendientes

La poda en C4.5 (III)



La poda en C4.5 (IV)



Error estimado combinado (6:2:6): 0.51

Valores desconocidos (missing)

- Inducción:
  - Se calcula el ratio de ganancia sobre los ejemplos con valor conocido,
  - se añade una rama ficticia que agrupa, si hay, los ejemplos con valor desconocido y
  - se usa un esquema de ***ponderación***, asociando a cada ejemplo con valor desconocido la probabilidad de pertenecer a cada subconjunto
- Clasificación:
  - Se utiliza el mismo esquema de ***ponderación***

Atributos de tipo continuo

- Test de la forma: Valor ≤ t
  - un conjunto E se divide en dos subconjuntos: $E_{\leq t}$ y $E_{>t}$
- Si un atributo continuo toma en el conjunto E los valores ordenados $\{v_1, v_2, \ldots, v_n\}$
  - hay $n\text{-}1$ umbrales posibles: $t_i=(v_i+v_{i+1})/2$ con $i=1..n\text{-}1$
  - Optimización: sólo se consideran los umbrales en los que se produce un cambio de clase
    - » Ej: { 1_+ 2_+ 2_+ 3_- 3_- 3_- 4_+ 4_+ 4_+ } Umbrales: 2.5 3.5
- C4.5 selecciona como umbral el t_i que maximiza un criterio de calidad:

$$\text{ratio de ganancia} - \frac{\log_2(n-1)}{|E|}$$

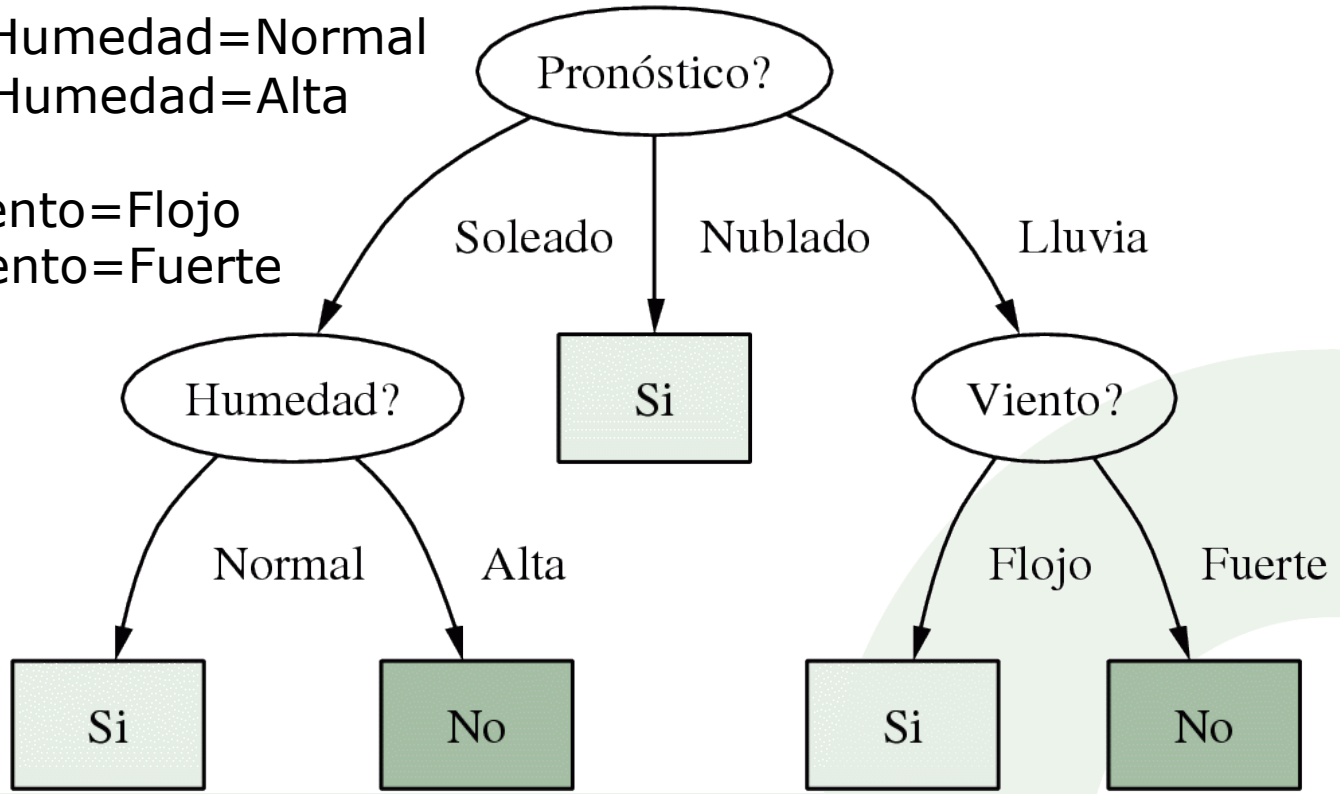
C4.5: Conclusiones

- El espacio de hipótesis es completo, no existe el riesgo que la función objetivo no se encuentre en el espacio de hipótesis
- Sólo mantiene una hipótesis mientras explora el espacio de hipótesis posible
- Las técnicas de post-poda permiten realizar un backtracking que puede evitar que el algoritmo seleccione un árbol que se encuentre en un mínimo local
- Usa todos los ejemplos en cada paso de búsqueda, en contraposición a técnicas incrementales, lo que le hace menos sensible al ruido
- Sesgo inductivo:
  - preferencia por los árboles pequeños (sesgo por preferencia)
  - coloca atributos más informativos cerca del nodo raíz
  - no presenta sesgos por restricción

De árboles a reglas (c4.5-rules)

- Proceso trivial: se construye una regla por cada camino desde el nodo raíz a cada hoja

Reglas:
Si ← Pronóstico=Soleado ^ Humedad=Normal
No ← Pronóstico=Soleado ^ Humedad=Alta
Si ← Pronóstico=Nublado
Si ← Pronóstico=Lluvia ^ Viento=Flojo
No ← Pronóstico=Lluvia ^ Viento=Fuerte



Post-proceso de reglas (I)

- **Eliminación de antecedentes:**
- Estimación pesimista de la probabilidad de fallo:

$$I^{\alpha}_{\sup}\left(n, \frac{f}{n}\right)$$

- Es preferible R' (R sin el antecedente X) si:

$$I^{\alpha}_{\sup}\left(n_R, \frac{f_R}{n_R}\right) \geq I^{\alpha}_{\sup}\left(n_{R'}, \frac{f_{R'}}{n_{R'}}\right)$$

Post-proceso de reglas (II)

- **Eliminación de reglas:**
- Se minimiza el coste de codificación (MDL) de cada subconjunto S de reglas (uno por clase):

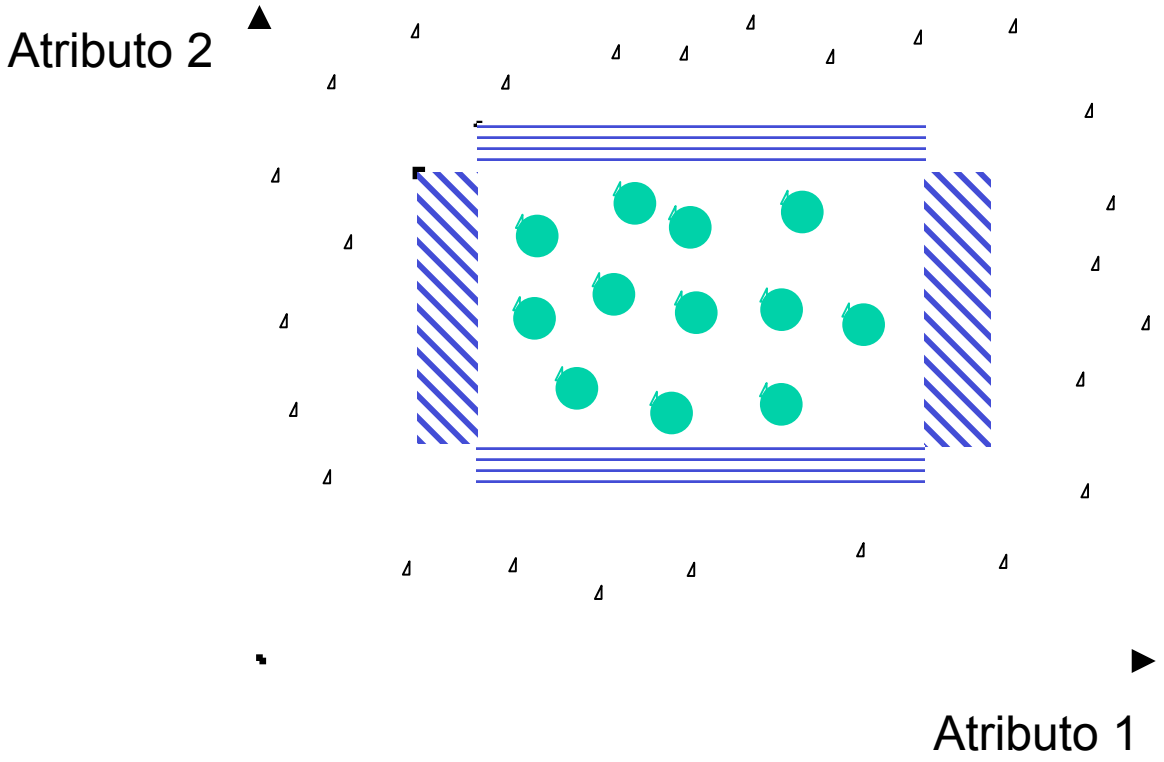
$$C(S) = C_X(S) + 0.5 \cdot C_T(S)$$

$$C_T(S) = \sum_{i=1}^{|S|} C(R_i) - \log_2(|S|!)$$

$$C_X(S) = \log_2 \binom{r}{fp} + \log_2 \binom{|E| - r}{fn}$$

- siendo fp los falsos positivos, fn falsos negativos y r el nº de ejemplos cubiertos por todas las reglas de S
- Se ordenan de menor a mayor número de falsos positivos
- Finalmente, se añade una **regla por defecto** cuya clase será la mayoritaria entre los ejemplos no cubiertos o, en caso de empate, la de mayor frecuencia en el conjunto de entrenamiento

Falsos positivos y falsos negativos



- ***hipótesis***
- ***función objetivo***
- ***falsos negativos***
- ***falsos positivos***

Conjunto de reglas final

No ← Pronóstico=Soleado ^ Humedad=Alta
No ← Pronóstico=Lluvia ^ Viento=Fuerte
Si ← Pronóstico=Nublado
Si ← Humedad=Normal
Si ← Pronóstico=Lluvia ^ Viento=Flojo
Si ← (regla por defecto)

Mecanismos de inducción de reglas

- Paradigma **Separa y Vencerás**
  - En cada iteración se trata de inducir una regla que explique los ejemplos no explicados por otras reglas inferidas anteriormente
    - » IREP, RIPPER, etc…
- **Generalización de instancias**
  - Reglas muy específicas (por ejemplo las que cubren un sólo caso) son generalizadas mediante el ajuste/eliminación de antecedentes
    - » RISE, INNER, etc…

Árboles de regresión: M5 [Quinlan 95]

- Muchas veces las clases son continuas
- Clásicamente, se ha utilizado la regresión cuando esto ocurría, pero los modelos obtenidos eran numéricos
- M5 genera árboles de decisión similares a los producidos por c4.5
- Es una variación de CART (Breiman et al., 84)
  - Las hojas en CART son valores numéricos en lugar de modelos lineales
  - CART elige aquél atributo que maximice la reducción esperada en varianza o en desviación absoluta

Características de M5

- **Heurística**: minimizar la variación interna de los valores de la clase dentro de cada subconjunto
- **Medida concreta**: elegir aquél atributo que maximice la reducción del error, de acuerdo a la siguiente formula:

$$\Delta error = sd(E) - \sum_{i=1}^{k} \frac{|E_i|}{|E|} \cdot sd(E_i)$$

  - E es el conjunto de ejemplos en el nodo a dividir,
  - E_i son los ejemplos con valor i del atributo a considerar, y
  - $sd(C)$ es la desviación típica de los valores para los ejemplos en C
- **Hojas**: se calcula un modelo lineal utilizando regresión estándar en función de los valores de los atributos (numéricos)
- **Criterio de parada en cada nodo**: pocos ejemplos, o poca variación de los valores de la clase

Estimación del error

- Para estimar el error en posteriores instancias calcula la media del error residual producido al clasificar con el modelo creado, m, cada instancia del cjto de test T:

$$error(T,m) = \frac{1}{n}\sum_{i \in T} \left\| c(i) - c(m,i) \right\|$$

- siendo $n = |T|$, $c(i)$ es la clase de la instancia i, y $c(m,i)$ es la clasificación con el modelo m de la instancia i
- Como esto subestima el error en posteriores instancias, se multiplica por

$$\frac{n+v}{n-v}$$

- siendo v el número de atributos en el modelo m
- Esto incrementa el error en modelos construidos con muchos parámetros y pocas instancias

Otros procesos

- **Construcción de modelos lineales**: se calculan para cada nodo del árbol, considerando sólo los atributos que aparecen en su subárbol como test o en modelos lineales
- **Simplificación de los modelos lineales**: en cada modelo lineal se eliminan atributos, utilizando escalada, para reducir el error estimado. Esto, normalmente, hace que aumente el error residual, pero también reduce el factor por el que luego se multiplica. Puede llegar a dejar sólo una constante
- **Poda**: cada nodo interno del árbol tiene ahora un modelo simplificado lineal y un modelo subárbol. Se elige aquél que minimice el error. Si es el modelo lineal, el subárbol se elimina
- **Suavizar el árbol**: se tienen en cuenta los demás modelos desde el nodo hoja al nodo raíz

Ejemplo

Trabajo	Pelota	Va a clase	Examen	Nota
4	Sí	Sí	6	3.6
6	No	Sí	4	6.5
8	No	No	5	6.5
10	Sí	No	6	6
…	…	…	…	…

Pelota

Sí — No

Nota=0.4*Trabajo+0.5*Examen-1

Va a clase

Sí — No

Nota=0.7*Trabajo+0.7*Examen+0.5

Nota=0,5*Trabajo+0,5*Examen