

# Aprendizaje inductivo de reglas de decisión

## Tratamiento Inteligente de la Información y Aplicaciones

Juan A. Botía

Departamento de Ingeniería de la Información y las Comunicaciones  
Universidad de Murcia

14 de noviembre de 2005

# Guión de la clase

Introducción

AQ

CN2

Algoritmos más recientes, otros enfoques

## Aprendizaje de reglas IF-THEN

- ▶ Se trata de inducir conceptos, a partir de los datos.
- ▶ Michalski lo denomina “aprendizaje inductivo de conceptos” [11]
- ▶ Concepto definido por una o más reglas
- ▶ Restringido a conjuntos de datos  $D$  tal que  $C = \{\oplus, \ominus\}$ .
- ▶ Buscamos compromiso completitud-consistencia
  - ▶ Errores de medida, clasificación, transmisión
  - ▶ Inconsistencias

## Morfología del modelo

Para un problema de clasificación  $n$ -ario,  $\{c_1, c_2, \dots, c_n\}$

$R_1$	$antecedente_1$	then	$c_1$
$R_2$	$antecedente_2$	then	$c_1$
	...		
$R_k$	$antecedente_k$	then	$c_1$
$R_{k+1}$	$antecedente_{k+1}$	then	$c_2$
	...		
$R_m$	$antecedente_1$	then	$c_{n-1}$
		else	$c_n$

## Morfología, cont.

- ▶ Es posible utilizar más de una regla por clase
- ▶ El índice  $i$  puede inducir un orden tal que  $R_{i-1}$  se evaluará justo antes que  $R_i$ : la predicción corresponderá a aquella que primero se dispare (*lista de reglas*[13])
- ▶ Un ejemplar determinado se clasificará por la clase menos probable siempre que no se haya emparejado con ninguno de los *antecedente<sub>k</sub>*,  $1 \leq k \leq n - 1$  anteriores

## Completitud

### Definición

Sea un conjunto de datos  $D$ , en el cual todas los ejemplares están etiquetadas con valores en un conjunto  $C = \{c_1, c_2, \dots, c_k\}$ . Sea  $R$  una regla que describe el concepto relacionado con  $c_i$ . Sea además,  $D_{c_i}$  el conjunto de instancias de  $D$  etiquetadas con  $c_i$ . Sea  $r$  el número de instancias de  $D_{c_i}$  clasificadas por  $R$ . Se define la completitud de la regla  $R$  (i.e. hipótesis) como

$$\frac{r}{|D_{c_i}|}$$

Es decir, el porcentaje de instancias de  $c_i$  cubiertas por la regla  $R$ .

## Consistencia

### Definición

Sea un conjunto de datos  $D$ , cuyas instancias etiquetadas con valores en el conjunto  $C = \{c_1, c_2, \dots, c_k\}$ . Sea  $R$  una regla que describe el concepto relacionado con  $c_i$ . Sea además,  $D_{c_i}$  el conjunto de instancias de  $D$  etiquetadas con  $c_i$ . Sea  $r'$  el número de instancias clasificadas de forma incorrecta. Se define la consistencia de la regla  $R$  como

$$\frac{|D_{c_i}| - r'}{|D_{c_i}|}$$

Es decir, el porcentaje de instancias de  $c_i$  cubiertas correctamente por la regla  $R$ .

## Completitud y Consistencia

- ▶ Una  $h$  que es completa y consistente con  $D$  cubre todos los ejemplos positivos, y ningún ejemplo negativo.
- ▶ Si  $D$  carece de ruido (i.e. errores), las  $h$  generadas por un  $L$  inductivo deben ser completas y consistentes
- ▶ En  $D$  del mundo real buscaremos  $h$  con valores de compromiso para los dos parámetros



## Completitud y Consistencia (II)

- ▶ Sea un conjunto de datos con 1000 ejemplos positivos (P) y 1000 ejemplos negativos (N)
- ▶ Sean dos reglas,  $r_1$  y  $r_2$ 
  - ▶  $r_1$  cubre 600 positivos (p) y 2 negativos (n)
  - ▶  $r_2$  cubre 950 p y 20 n
- ▶ Tenemos entonces que

Regla	Completitud	Consistencia
$r_1$	60 %	99.7 %
$r_2$	95 %	98 %

- ▶ El criterio de RIPPER [6] es maximizar

$$(p - n)/(P + N)$$

## AQ y CN2

- ▶ Los dos algoritmos más conocidos son la familia de algoritmos AQ [9, 12, 10, 8, 11] y CN2 [3, 4, 2]
- ▶ Funcionamiento procedimental parecido
  - ▶ Inducen reglas de forma iterativa.
  - ▶ Parten de una regla inicial  $R$ , creada a partir de  $D$ .
  - ▶  $R$  cubrirá muchos ejemplos positivos, y algún ejemplo negativo.
  - ▶ Repetimos proceso retirando de  $D$  ejemplos cubiertos por  $R$
  - ▶ Finalizamos hasta que todo ejemplo  $\oplus$  esté cubierto por alguna  $R_i$
  - ▶ Comportamiento de cubrimiento secuencial

## El algoritmo AQ

- ▶ AQ(n) induce un conjunto de reglas de decisión para cada etiqueta de clase que aparece en el conjunto  $D$ .
- ▶ Cada regla tiene la forma

Si  $\langle cover \rangle$  entonces predice  $\langle clase \rangle$ ,

- ▶  $\langle cover \rangle$  es una combinación lógica de tests sobre atributos del conjunto  $D$ .
- ▶ Un test básico en un atributo se denomina *selector*.

$\langle viento = suave \rangle$  ó  $\langle temperatura \leq 45 \rangle$

- ▶ Operadores  $\{=, \leq, >, \neq\}$
- ▶ Un *complejo* es un conjunto de selectores, y un *cover* es una disyunción de complejos.
- ▶ Una expresión cubre una instancia si se hace cierta para esa instancia.

## AQ: algoritmo

Sea  $D$  un cpto. de ejemplares tal que su correspondiente conjunto de clases es  $C = \{\oplus, \ominus\}$ .

Dividimos  $D$  en el conjunto  $Pos$  que contiene todos los ejemplares  $d_i \in D$  tal que  $c_i = \oplus$  y en el conjunto  $Neg$  que contiene todos los  $d_i \in D$  tal que  $c_i = \ominus$ .

Procedimiento  $AQR(Pos, Neg)$

1. Hacer  $Cover \leftarrow \emptyset$
2. Mientras que  $Cover$  no cubra todos los ejemplos positivos de  $Pos$ 
  - 2.1 Seleccionar un ejemplo positivo no cubierto por  $Cover$ , llamémosle  $Seed$
  - 2.2 Inicializar  $Star \leftarrow STAR(Seed, Neg)$
  - 2.3 Seleccionar complejo  $Best \in Star$  más simple
  - 2.4 Añadir  $Best$  como disyunción a  $Cover$
3. Devolver  $Cover$

## AQ: algoritmo ( y II)

### Procedimiento STAR(Seed, Neg)

- ▶  $Star \leftarrow \{\emptyset\}$  el complejo vacío
- ▶ Mientras que algún complejo en  $Star$  cubra ejemplos en  $Neg$ 
  1. Seleccionar un  $d \in Neg$ , cubierto por un complejo en  $Star$
  2. Especializar los complejos de  $Star$  para que no cubran a  $d$ 
    - 2.1 Sea  $Extension$  el conjunto de todos los selectores que cubren  $Seed$  y no  $d$
    - 2.2  $Star \leftarrow \{x \wedge y \mid x \in Star, y \in Extension\}$
    - 2.3 Eliminar todo complejo en  $Star$  subsumido en otros complejos en  $Star$
- ▶ Devolver  $Star$

## El objeto Star

### Definición

*Una STAR de un evento  $e$  bajo ciertas restricciones  $E$  es un conjunto de todas las posibles descripciones, no redundantes y alternativas, del evento  $e$  que no violan las restricciones de  $E$ .*

- ▶  $e$  se refiere a una instancia del conjunto  $D$ ,
- ▶  $E$  es el conjunto de ejemplos  $d \in Neg$  y
- ▶ una descripción es en realidad un complejo.
- ▶ STAR está formado por los selectores distintos que cubren algún ejemplo positivo (al menos SEED), y ninguno negativo.
- ▶ El eliminar ejemplos negativos se consigue mediante el bucle
- ▶ En cada iteración se especializa los complejos añadiendo en forma conjuntiva los tests que cumple al menos SEED y no el correspondiente  $d$ .

## AQ: búsqueda

- ▶ El tipo de búsqueda del algoritmo AQ es de lo genérico a lo específico
- ▶ De tipo *beam* (se mantienen  $k$  posibles soluciones)
- ▶ El mejor complejo se elige en base a una métrica concreta (compromiso completitud-consistencia)

## AQ: ejemplo (y 2)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No



## AQ: ejemplo (y 3)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
<b>D3</b>	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

► Seed  $\rightarrow D_3$

## AQ: ejemplo (y 4)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
<b>D1</b>	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
<b>D3</b>	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

▶ Seed  $\leftarrow D_3$

▶ d  $\leftarrow D_1$

## AQ: ejemplo (y 5)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
<b>D3</b>	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

- ▶ Seed  $\leftarrow D_3$
- ▶  $d \leftarrow D_1$
- ▶  $EXTENSION = \{Outlook = overcast\}$

## AQ: ejemplo (y 6)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
<b>D3</b>	<b>Overcast</b>	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

- ▶ Seed  $\leftarrow D_3$
- ▶  $d \leftarrow D_1$
- ▶ *EXTENSION*  $\leftarrow \{Outlook = overcast\}$
- ▶ *Star*  $\leftarrow \{Outlook = overcast\}$

## AQ: ejemplo (y 7)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

- ▶ Seed  $\leftarrow D_3$
- ▶  $d \leftarrow D_1$
- ▶  $EXTENSION \leftarrow \{Outlook = overcast\}$
- ▶  $Star \leftarrow \{Outlook = overcast\}$
- ▶  $Cover \leftarrow \{Outlook = overcast\}$

## AQ: ejemplo (y 8)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

## AQ: ejemplo (y 9)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

► Seed ← D4

## AQ: ejemplo (y 10)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

▶  $Seed \leftarrow D_4$

▶  $d \leftarrow D_1$



## AQ: ejemplo (y 11)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $Seed \leftarrow D_4$
- ▶  $d \leftarrow D_1$
- ▶  $EXTENSION \leftarrow \{Outlook = Rain, Temperature = Mild\}$
- ▶  $STAR = \{Outlook = Rain, Temperature = Mild\}$

## AQ: ejemplo (y 12)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

▶  $d \leftarrow D_6$

## AQ: ejemplo (y 13)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_6$
- ▶  $EXTENSION \leftarrow \{Temp = Mild, Humidity = High, Wind = Weak\}$

## AQ: ejemplo (y 14)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

►  $d \leftarrow D_6$

►  $EXTENSION \leftarrow \{Wind = Weak, Humidity = High, Wind = Weak\}$

$Star = \{$   
 $Outlook = Rain \wedge Temperature = Mild,$   
 $Outlook = Rain \wedge Humidity = High,$   
 $Outlook = Rain \wedge Wind = Weak,$   
 $Temp = Mild, Temp = Mild \wedge Humidity = High,$   
 $Temp = Mild \wedge Wind = Weak\}$

## AQ: ejemplo (y 15)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

▶  $d \leftarrow D_6$

▶  $EXTENSION \leftarrow \{Wind = Weak, Humidity = High, Wind = Weak\}$

$Star = \{$   
 $Outlook = Rain \wedge Temperature = Mild,$   
 $Outlook = Rain \wedge Humidity = High,$   
 $Outlook = Rain \wedge Wind = Weak,$   
 $Temp = Mild \wedge Humidity = High,$   
 $Temp = Mild \wedge Wind = Weak\}$

## AQ: ejemplo (y 16)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

▶  $d \leftarrow D_{14}$

## AQ: ejemplo (y 17)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_{14}$
- ▶  $EXTENSION \leftarrow \{Wind = Weak\}$

## AQ: ejemplo (y 18)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

▶  $d \leftarrow D_{14}$

▶  $EXTENSION \leftarrow \{Wind = Weak\}$

Star =  $\{$  Outlook = Rain  $\wedge$  Temperature = Mild  $\wedge$  Wind = Weak,  
 Outlook = Rain  $\wedge$  Humidity = High  $\wedge$  Wind = Weak,  
 Outlook = Rain  $\wedge$  Wind = Weak  $\wedge$  Wind = Weak,  
 Temp = Mild  $\wedge$  Humidity = High  $\wedge$  Wind = Weak,  
 Temp = Mild  $\wedge$  Wind = Weak  $\wedge$  Wind = Weak  $\}$



## AQ: ejemplo (y 19)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_{14}$
- ▶  $EXTENSION \leftarrow \{Wind = Weak\}$

$Star = \{Outlook = Rain \wedge Temperature = Mild \wedge Wind = Weak,$   
 $Outlook = Rain \wedge Humidity = High \wedge Wind = Weak,$   
 $Temp = Mild \wedge Humidity = High \wedge Wind = Weak\}$

## AQ: ejemplo (y 20)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

▶  $d \leftarrow D_8$

## AQ: ejemplo (y 21)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_8$
- ▶  $EXTENSION \leftarrow \{Outlook = Rain\}$

## AQ: ejemplo (y 22)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

▶  $d \leftarrow D_8$

▶  $EXTENSION \leftarrow \{Outlook = Rain\}$

$Star = \{Outlook = Rain \wedge Temperature = Mild \wedge Wind = Weak \wedge Outlook = Rain,$   
 $Outlook = Rain \wedge Humidity = High \wedge Wind = Weak \wedge Outlook = Rain,$   
 $Temperature = Mild \wedge Humidity = High \wedge Wind = Weak \wedge Outlook = Rain\}$

## AQ: ejemplo (y 23)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_8$
- ▶  $EXTENSION \leftarrow \{Outlook = Rain\}$
- ▶  $Star \leftarrow \{Temperature = Mild \wedge Humidity = High \wedge Wind = Weak \wedge Outlook = Rain\}$

## AQ: ejemplo (y 24)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_8$
- ▶  $EXTENSION \leftarrow \{Outlook = Rain\}$
- ▶  $Star \leftarrow \{Temperature = Mild \wedge Humidity = High \wedge Wind = Weak \wedge Outlook = Rain\}$
- ▶  $Cover = \{Outlook = Overcast \vee (Temperature = Mild \wedge Humidity = High \wedge Wind = Weak \wedge Outlook = Rain)\}$

## AQ: ejemplo (y 25)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

▶  $Seed \leftarrow D_5$

▶  $d \leftarrow D_1$

## AQ: ejemplo (y 26)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $Seed \leftarrow D_5$
- ▶  $d \leftarrow D_1$
- ▶  $EXTENSION = \{Outlook = Rain, Temperature = Cool, Humidity = Normal\}$



## AQ: ejemplo (y 27)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $Seed \leftarrow D_5$
- ▶  $d \leftarrow D_1$
- ▶  $EXTENSION = \{Outlook = Rain, Temperature = Cool, Humidity = Normal\}$
- ▶  $STAR = \{Outlook = Rain, Temperature = Cool, Humidity = Normal\}$

## AQ: ejemplo (y 28)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

▶  $d \leftarrow D_6$

## AQ: ejemplo (y 29)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_6$
- ▶  $EXTENSION = \{Wind = Weak\}$

## AQ: ejemplo (y 30)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

▶  $d \leftarrow D_6$

▶  $EXTENSION = \{Wind = Weak\}$

$Star = \{Outlook = Rain \wedge Wind = Weak,$   
 $Temperature = Cool \wedge Wind = Weak,$   
 $Humidity = Normal \wedge Wind = Weak\}$

## AQ: ejemplo (y 31)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

## AQ: ejemplo (y 32)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

►  $d \leftarrow D_6$

►  $EXTENSION = \{Wind = Weak\}$

$Star = \{Outlook = Rain \wedge Wind = Weak,$   
 $Temperature = Cool \wedge Wind = Weak,$   
 $Humidity = Normal \wedge Wind = Weak\}$

$Cover = Outlook = Overcast \vee$   
 $(Temperature = Mild \wedge Humidity = High \wedge Wind = Weak \wedge Outlook = Rain) \vee$   
 $Outlook = Rain \wedge Wind = Weak$

## AQ: ejemplo (y 33)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

## AQ: ejemplo (y 34)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

► Seed ← D<sub>9</sub>



## AQ: ejemplo (y 35)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

▶  $Seed \leftarrow D_9$

▶  $d \leftarrow D_1$

## AQ: ejemplo (y 36)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $Seed \leftarrow D_9$
- ▶  $d \leftarrow D_1$
- ▶  $EXTENSION \leftarrow \{Temperature = Cool, Humidity = Normal\}$

## AQ: ejemplo (y 37)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $Seed \leftarrow D_9$
- ▶  $d \leftarrow D_1$
- ▶  $EXTENSION \leftarrow \{Temperature = Cool, Humidity = Normal\}$
- ▶  $STAR \leftarrow EXTENSION$

## AQ: ejemplo (y 38)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_2$
- ▶  $EXTENSION \leftarrow \{Wind = Weak\}$

## AQ: ejemplo (y 39)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_2$
- ▶  $EXTENSION \leftarrow \{Wind = Weak\}$
- ▶  $STAR \leftarrow \{Temp = Cool \wedge Wind = Weak, Humidity = Normal \wedge Wind = Weak\}$

## AQ: ejemplo (y 40)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_2$
- ▶  $EXTENSION \leftarrow \{Wind = Weak\}$
- ▶  $STAR \leftarrow \{Temp = Cool \wedge Wind = Weak, Humidity = Normal \wedge Wind = Weak\}$

Cover =  $Outlook = Overcast \vee$   
 $(Temperature = Mild \wedge Humidity = High \wedge Wind = Weak \wedge Outlook = Rain) \vee$   
 $(Outlook = Rain \wedge Wind = Weak) \vee$   
 $(Temp = Cool \wedge Wind = Weak)$

## AQ: ejemplo (y 41)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

## AQ: ejemplo (y 42)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

►  $Seed \leftarrow D_{11}$



## AQ: ejemplo (y 43)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

▶  $Seed \leftarrow D_{11}$

▶  $d \leftarrow D_1$

## AQ: ejemplo (y 44)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $Seed \leftarrow D_{11}$
- ▶  $d \leftarrow D_1$
- ▶  $EXTENSION = \{Temp = Mild, Humidity = Normal, Wind = Strong\}$

## AQ: ejemplo (y 45)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $Seed \leftarrow D_{11}$
- ▶  $d \leftarrow D_1$
- ▶  $EXTENSION = \{Temp = Mild, Humidity = Normal, Wind = Strong\}$
- ▶  $STAR \leftarrow EXTENSION$

## AQ: ejemplo (y 46)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

►  $d \leftarrow D_8$

## AQ: ejemplo (y 47)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_8$
- ▶  $EXENSION \leftarrow \{Humidity = Normal, Wind = Strong\}$

## AQ: ejemplo (y 48)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_8$
- ▶  $EXTENSION \leftarrow \{Humidity = Normal, Wind = Strong\}$

Star =  $\{Temperature = Mild \wedge Wind = Strong,$   
 $Temperature = Mild \wedge Humidity = Normal,$   
 $Humidity = Normal,$   
 $Humidity = Normal \wedge Wind = Strong,$   
 $Wind = Strong \wedge Humidity = Normal\}$

## AQ: ejemplo (y 49)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_8$
- ▶  $EXTENSION \leftarrow \{Humidity = Normal, Wind = Strong\}$

$Star = \{Temperature = Mild \wedge Humidity = Normal, Wind = Strong \wedge Humidity = Normal\}$

## AQ: ejemplo (y 50)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

►  $d \leftarrow D_6$



## AQ: ejemplo (y 51)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_6$
- ▶  $EXTENSION \leftarrow \{Outlook = Sunny, Temperature = Mild\}$

## AQ: ejemplo (y 52)

Data	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

►  $d \leftarrow D_6$

►  $EXTENSION \leftarrow \{Outlook = Sunny, Temperature = Mild\}$

$Star = \{Outlook = Sunny \wedge Temperature = Mild \wedge Humidity = Normal,$   
 $Outlook = Sunny \wedge Humidity = Normal \wedge Wind = Strong,$   
 $Temperature = Mild \wedge Temperature = Mild \wedge Humidity = Normal,$   
 $Temperature = Mild \wedge Humidity = Normal \wedge Wind = Strong\}$

## AQ: ejemplo (y 53)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_6$
- ▶  $EXTENSION \leftarrow \{Outlook = Sunny, Temperature = Mild\}$

$Star = \{Outlook = Sunny \wedge Temperature = Mild \wedge Humidity = Normal,$   
 $Outlook = Sunny \wedge Humidity = Normal \wedge Wind = Strong,$   
 $Temperature = Mild \wedge Humidity = Normal \wedge Wind = Strong\}$

## AQ: ejemplo (y 54)

Data	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D6	Rain	Cool	Normal	Strong	No
D8	Sunny	Mild	High	Weak	No
D11	Sunny	Mild	Normal	Strong	Yes
D14	Rain	Mild	High	Strong	No

- ▶  $d \leftarrow D_6$
- ▶  $EXTENSION \leftarrow \{Outlook = Sunny, Temperature = Mild\}$

Star =  $\{Outlook = Sunny \wedge Temperature = Mild \wedge Humidity = Normal,$   
 $Outlook = Sunny \wedge Humidity = Normal \wedge Wind = Strong,$   
 $Temperature = Mild \wedge Humidity = Normal \wedge Wind = Strong\}$

Cover =  $Outlook = Overcast \vee$   
 $(Temperature = Mild \wedge Humidity = High \wedge Wind = Weak \wedge Outlook = Rain) \vee$   
 $(Outlook = Rain \wedge Wind = Weak) \vee$   
 $(Temp = Cool \wedge Wind = Weak) \vee$   
 $(Outlook = Sunny \wedge Temperature = Mild \wedge Humidity = Normal)$

## AQ: heurísticas

- ▶ El mejor complejo es aquel que cubre más ejemplos
- ▶ Los peores complejos, cuando se está *recortando* STAR son aquellos que tienen un mínimo de ejemplos positivos cubiertos y negativos excluidos
- ▶ En caso de empate, se elige al más simple (menos selectores)
- ▶ Selección de SEED es aleatoria
- ▶ Selección de los  $d$  más cercanos a SEED (más similitud en valores de atributos)

## El algoritmo CN2

- ▶ CN2 → Clark y Niblet
- ▶ Rediseño de AQ → No se usa el concepto SEED
- ▶ El espacio de búsqueda se incrementa
- ▶ Búsqueda de tipo beam
- ▶ Mantenimiento de una lista ordenada de reglas [13]

Si  $\langle \text{complejo} \rangle$  entonces predice  $\langle \text{clase} \rangle$

- ▶ La última regla en la lista ordenada de reglas será siempre la regla por defecto, que predice la clase más probable.
- ▶ Se evalúan en el orden establecido hasta que una de ellas se cumple.

# Algoritmo

Procedimiento  $CN2(D)$ : RuleList

- ▶ Hacer  $RuleList \leftarrow \emptyset$
- ▶ Repetir
  1. Hacer  $MejorComp \leftarrow EncontrarMejorComp(D)$
  2. Si  $MejorComp \neq \emptyset$  entonces
    - 2.1 Sea  $D'$  el conjunto de ejemplos cubierto por  $MejorComp$
    - 2.2 Eliminar  $D'$  de  $D$ . Sea  $c_{D'}$  la clase más probable en  $D'$
    - 2.3 Añadir regla "Si  $MejorComp$  entonces  $class=c_{D'}$ " a  $RuleList$
- ▶ Hasta que  $MejorComp = \emptyset$  o  $D$  esté vacío

## Algoritmo (y II)

Procedimiento *EncontrarMejorComp(D) : MejorComp*

- ▶ Sea *Star* un conjunto con el complejo vacío,  $MejorComp \leftarrow \emptyset$  y *Selectores* el conjunto de todos los posibles selectores
- ▶ Mientras que  $Star \neq \emptyset$ 
  1. Especializar todo complejo en *Star* como sigue

1.1 Sea *NewStar* el conjunto

$$\{x \wedge y \mid x \in Star, y \in Selectores\}$$

- 1.2 Eliminar todo complejo en *NewStar* que bien se encuentra en *Star* ó es nulo (e.g.  $big = y \wedge big = n$ )
- 1.3 Para todo complejo  $Comp_i \in NewStar$  :  
Si  $Comp_i$  es significativo estadísticamente cuando se comprueba con *D* y además es mejor que *MejorComp*, dependiendo del criterio del usuario al comprobarlo con *D*, entonces hacer  $MejorComp = Comp_i$ .
- 1.4 Eliminar los peores complejos de *NewStar* hasta que su tamaño sea menor ó igual que un máximo definido por el usuario

2. Hacer  $Star \leftarrow NewStar$

- ▶ Devolver *MejorComp*



## CN2: heurísticas (I)

Evaluar complejo

$$Entropy = - \sum_{i=1}^c p_i \log(p_i)$$

- ▶ Siendo  $E'$  el conjunto de ejemplos cubiertos
- ▶  $n$  el número de clases diferentes en  $E'$
- ▶  $p_1, p_2, \dots, p_n$  las probabilidades de cada clase en  $E'$

## CN2: heurísticas (II)

### Significancia estadística

- ▶ Detecta correlaciones reales entre valores de atributos y clases
- ▶ Compara la distribución de probabilidad entre clases observada
- ▶ con la esperada si el complejo cubriera ejemplos aleatoriamente
- ▶ Estadístico

$$2 \sum_{i=1}^n f_i \log(f_i/e_i)$$

- ▶  $f_1, f_2, \dots, f_n$  es el conjunto de frecuencias observadas en el conjunto de ejemplos cubiertos
- ▶  $e_1, e_2, \dots, e_n$  es la probabilidad de las clases en el conjunto original
- ▶ Sigue una distribución  $\chi^2$  con  $n - 1$  grados de libertad

## Enfoques más recientes

¿Novedades?

- ▶ Por un lado, varios tipos posibles de control del crecimiento del conjunto de reglas aprendidas
- ▶ Por el otro, los tipos de heurísticas para determinar el mejor tests a aplicar para seguir haciendo crecer una determinada regla

## REP

### REP (Reduced Error Pruning) [1]

- ▶ Funcionamiento en dos fases (datos de aprendizaje y datos de poda)
  1. Se crea un conjunto de reglas que se sobreajusta a los datos usados para el aprendizaje
  2. Se poda el cjto. de reglas usando ejemplares que no participaron en el aprendizaje
    - ▶ Se pueden podar tanto selectores como reglas enteras
    - ▶ Se basa en el uso de operadores (i.e. borrar selector, borrar regla, etc) que se aplican a las reglas y se escoge el que mejor error de generalización obtiene en el nuevo conjunto de reglas
  3. Elevado coste computacional

## IREP

### IREP (Incremental Reduced Error Pruning) [7]

- ▶ Reduce la complejidad de REP
- ▶ Funciona similarmente e incorpora un esquema *divide y vencerás* (conjunto de entrenamiento y de poda para cada  $\{c_1, c_2, \dots, c_k\}$ )
- ▶ La poda trata de eliminar selectores según una métrica basada, como siempre, en la consistencia y completitud
- ▶ Su complejidad es bastante menor  $O(n \log n)$  siendo  $n$  el número de ejemplares a clasificar
- ▶ Crítica: su capacidad de generalización no aumenta con el número de ejemplos
- ▶ Se termina de crear reglas cuando la última regla construida presenta un error de clasificación excesivo (e.g. del 50%)
  - ▶ si cjto. total de reglas contiene un número de reglas alto y cubren pocos ejemplos cada una, el error de generalización estimado en el cjto. de poda tiene mucha varianza

## IREP\* [5]

- ▶ Este nuevo algoritmo parte de IREP para cambiar la métrica de poda por una función y la condición de parada en la generación de reglas.
- ▶ Métrica: favorece la poda de reglas que cubran pocos ejemplos
- ▶ Condición de parada: basado en el MDL (*Minimum Description Length*)
  - ▶ consiste en intentar minimizar la expresión siguiente:

$$size(\text{cjto. reglas}) + size(\text{ejemplos mal clasificados}) \quad (1)$$

## RIPPER [5]

RIPPER equivale a *IREP*\* más un procedimiento de optimización de las reglas producidas

- ▶ Para cada regla se generan dos reglas (reemplazo y revisión) nuevas (se ha de escoger entre las tres)
- ▶ Reemplazo: se construye, partiendo de la regla original a la que se le añade algún test más y posteriormente se poda
- ▶ Revisión: análogamente

## SLIPPER

- ▶ Produce los modelos de más calidad
- ▶ Basado en el boosting: parte de una técnica *débil*
  - ▶ Aquella que consigue un error de predicción ligeramente mejor que el que se obtendría realizando las predicciones de manera aleatoria

y produce y combina varias hipótesis obtenidas a partir de diferentes muestreos del cjto. de datos

- ▶ Lo veremos en el ejemplo de aplicación



## Conclusiones

- ▶ Dos familias de algoritmos para inducción de reglas/listas de decisión
- ▶ Búsqueda de lo general a lo específico
- ▶ Compromiso entre consistencia y completitud
- ▶ Búsqueda beam en los complejos



C. A. Brunk and M. J. Pazzani.

An investigation of noise-tolerant relational concept learning algorithms.

Technical report, Department of Information and Computer Science, University of California, 1991.



P. Clark and R. Boswell.

Rule induction with CN2: Some recent improvements.

In *Proc. Fifth European Working Session on Learning*, pages 151–163, Berlin, 1991. Springer.



P. Clark and T. Niblett.

Induction in noisy domains.

In *Progress in Machine Learning—Proceedings of EWSL 87: 2nd European Working Session on Learning*, pages 11–30, Bled, Yugoslavia, 1987.



Peter Clark and Tim Niblett.

The cn2 induction algorithm.

*Machine Learning Journal*, 3(3):261–283, 1988.



P. Cohen and H. Levesque.

Communicative actions for artificial agents.

In Victor Lesser, editor, *Proceedings ICMAS-95*. AAAI, 1995.



W. Cohen.

Fast effective rule induction.

In *Proceedings of the Twelfth International Conference on Machine Learning*, pages 115–123, Lake Tahoe, CA, 1995.



Johannes Fürnkranz.

Separate-and-conquer rule learning.

*Artificial Intelligence Review*, 13(1):3–54, February 1999.



Kenneth A. Kaufman and Michalski Ryszard S.

Learning in an inconsistent world. rule selection in STAR/AQ18. mli 99-2.

Technical report, Machine Learning and Inference Laboratory.  
George Mason University, 1999.



Ryszard S. Michalksi.

On the quasi-minimal solution of the general covering problem.

*In Proceedings of the First International Simposium of Information Processing*, pages 125–128, Bled. Yugoslavia, 1969.



R. S. Michalski, L. Mozetic, J. R. Hong, and N. Lavrac.

The multi-purpose incremental learning system aq15 and its testing application to three medical domains.

*In AAAI Conference Proceedings*, 1986.



R.S. Michalski and K Kaufman.

Learning patterns in noisy data: The aq approach.

*In G. Paliouras, V. Karkaletsis, and Spyropoulos, editors, Machine Learning and Applications*. Springer-Verlag, 2001.



Ryszard S. Michalski.

A theory and methodology of inductive learning.

In R.S. Michalski, J.G. Carbonell, and T.M. Mitchell, editors,  
*Machine Learning: An Artificial Intelligence Approach*,  
volume 1, pages 83–129. Springer, 1983.



R. L. Rivest.

Learning decision lists.

*Machine Learning*, 2(3):229–246, 1987.