

Aprendizaje Automático  
Segundo Cuatrimestre de 2016

# Árboles de Decisión



DEPARTAMENTO  
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

# Aproximación de Funciones

## Ejemplo:

- Un amigo juega al tenis los sábados.
- Juega o no, dependiendo del estado del tiempo.
  - Cielo: {Sol, Nublado, Lluvia}
  - Temperatura: {Calor, Templado, Frío}
  - Humedad: {Alta, Normal}
  - Viento: {Fuerte, Débil}
- Queremos entender su criterio para jugar o no.
- **Función objetivo desconocida:**

Tenis : Cielo x Temperatura x Humedad x Viento  $\rightarrow$  {Sí, No}

- Para aprender esta función, registramos el comportamiento de nuestro amigo durante unos meses...

atributos

clase

instancias

Día	Cielo	Temperatura	Humedad	Viento	Tenis?
1	Sol	Calor	Alta	Débil	No
2	Sol	Calor	Alta	Fuerte	No
3	Nublado	Calor	Alta	Débil	Sí
4	Lluvia	Templado	Alta	Débil	Sí
5	Lluvia	Frío	Normal	Débil	Sí
6	Lluvia	Frío	Normal	Fuerte	No
7	Nublado	Frío	Normal	Fuerte	Sí
8	Sol	Templado	Alta	Débil	No
9	Sol	Frío	Normal	Débil	Sí
10	Lluvia	Templado	Normal	Débil	Sí
11	Sol	Templado	Normal	Fuerte	Sí
12	Nublado	Templado	Alta	Fuerte	Sí
13	Nublado	Calor	Normal	Débil	Sí
14	Lluvia	Templado	Alta	Fuerte	No

# Aproximación de Funciones

## Marco del problema:

- Conjunto de instancias posibles  $X$

Todas las combinaciones de Cielo x Temp x Humedad x Viento.

- Función objetivo desconocida  $f: X \rightarrow Y$

Tenis: Cielo x Temp x Humedad x Viento  $\rightarrow \{\text{Sí}, \text{No}\}$

- Conjunto de hipótesis  $H = \{ h \mid h : X \rightarrow Y \}$

Depende de la elección del algoritmo de aprendizaje.

## Entrada del algoritmo de aprendizaje:

- Ejemplos de entrenamiento  $\{ \langle x^{(i)}, y^{(i)} \rangle \}$  de la función  $f$ .

## Salida del algoritmo de aprendizaje:

- Hipótesis  $h \in H$  que mejor aproxima a la función  $f$ .

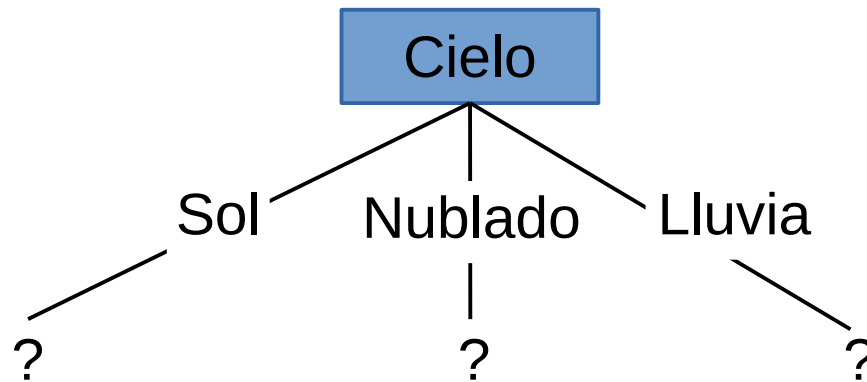
instancias

atributos

clase

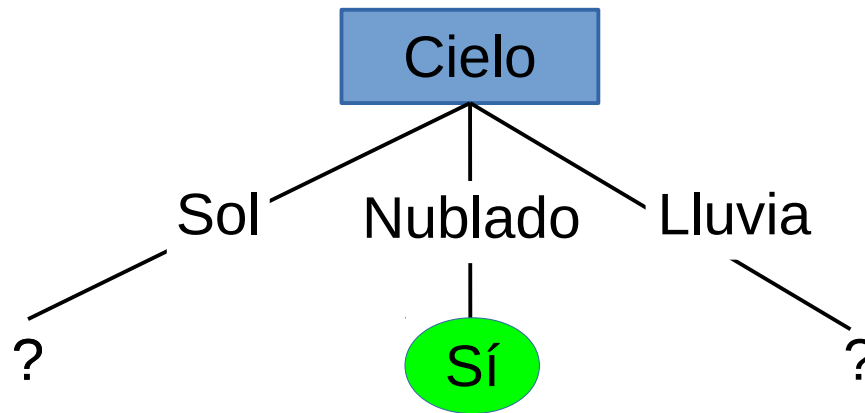
Día	Cielo	Temperatura	Humedad	Viento	Tenis?
1	Sol	Calor	Alta	Débil	No
2	Sol	Calor	Alta	Fuerte	No
3	Nublado	Calor	Alta	Débil	Sí
4	Lluvia	Templado	Alta	Débil	Sí
5	Lluvia	Frío	Normal	Débil	Sí
6	Lluvia	Frío	Normal	Fuerte	No
7	Nublado	Frío	Normal	Fuerte	Sí
8	Sol	Templado	Alta	Débil	No
9	Sol	Frío	Normal	Débil	Sí
10	Lluvia	Templado	Normal	Débil	Sí
11	Sol	Templado	Normal	Fuerte	Sí
12	Nublado	Templado	Alta	Fuerte	Sí
13	Nublado	Calor	Normal	Débil	Sí
14	Lluvia	Templado	Alta	Fuerte	No

# Construcción de un Árbol de Decisión



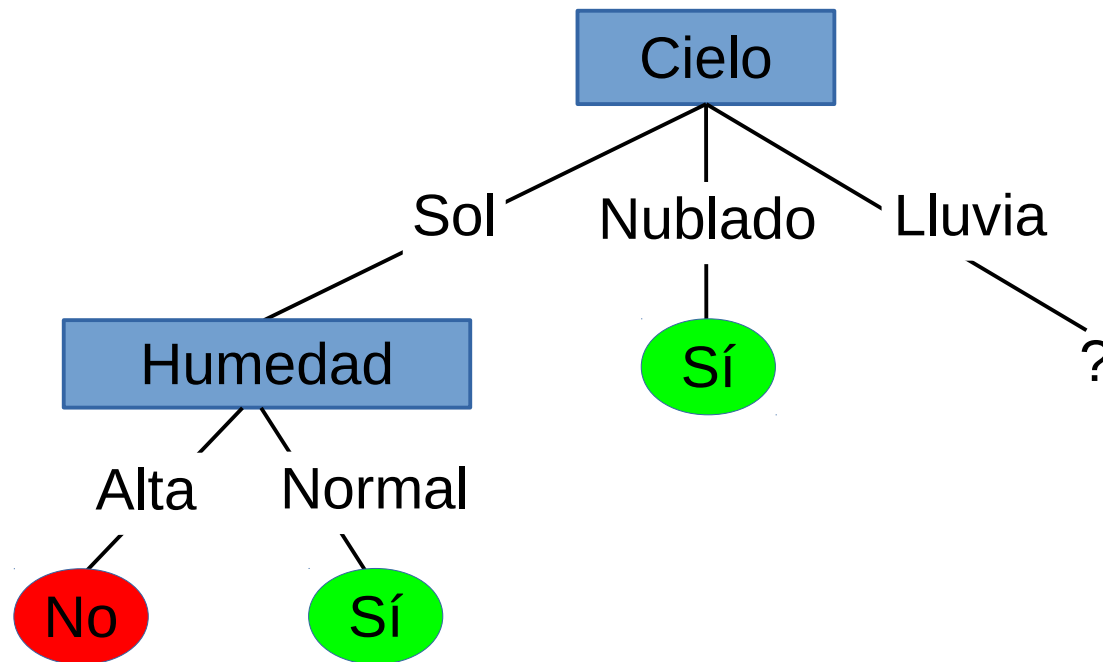
El atributo Cielo parece ser bueno para comenzar el árbol...

# Construcción de un Árbol de Decisión



Las instancias con Cielo==Nublado son todas positivas.

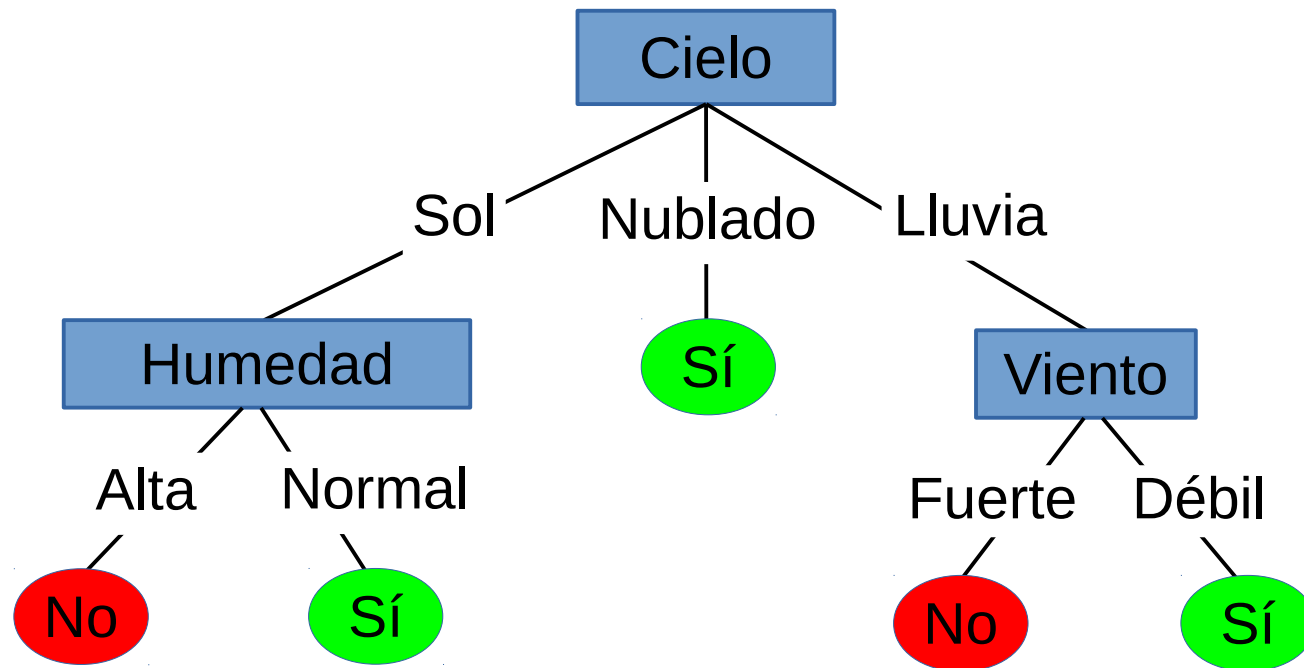
# Construcción de un Árbol de Decisión



Para las instancias con Cielo==Sol continuamos con el atributo Humedad, que separa (discrimina) perfectamente.

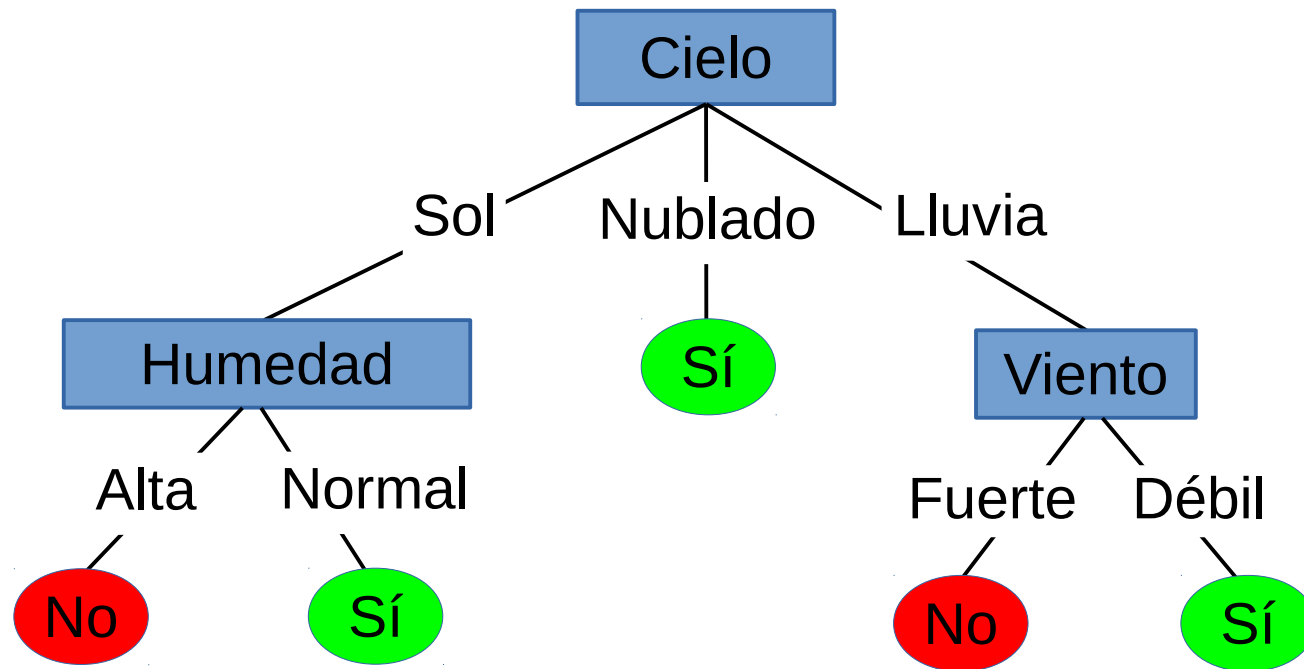


# Construcción de un Árbol de Decisión



Para las instancias con Cielo==Lluvia, el atributo Viento separa (discrimina) perfectamente.

# Árboles de Decisión

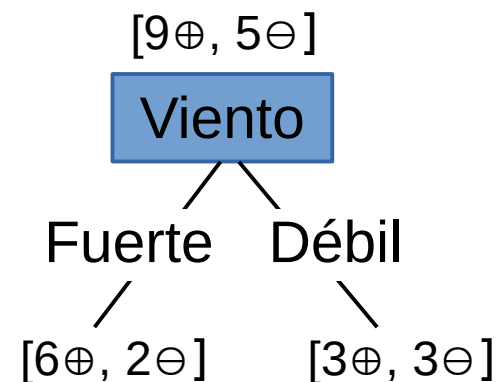
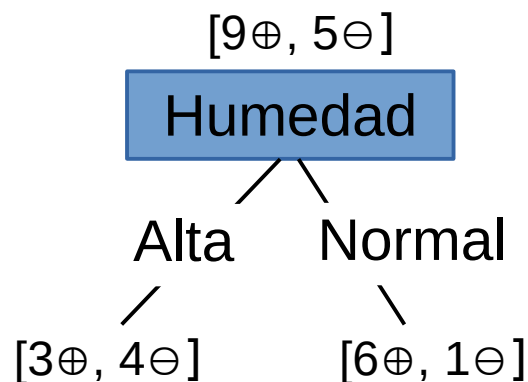


- $f : \langle X_1, \dots, X_n \rangle \rightarrow Y$
- Cada nodo interno evalúa un atributo discreto  $X_i$
- Cada rama corresponde a un valor para  $X_i$
- Cada hoja predice un valor de  $Y$

# Inducción *Top-Down* de Árboles de Decisión (ID3<sup>(a)</sup> y C4.5<sup>(b)</sup>, Quinlan)

- 1)  $A \leftarrow$  el “mejor” atributo para *nodo\_actual*.
- 2) Asignar  $A$  como atributo de decisión del *nodo\_actual*.
- 3) Para cada valor de  $A$ , crear un nuevo hijo del *nodo\_actual*.
- 4) Clasificar (*repartir*) las instancias en los nuevos nodos (según el valor de  $A$ ).
- 5) Si las instancias están clasificadas perfectamente: FIN.  
Si no: iterar sobre los nuevos nodos.

¿Cuál es el  
“mejor”  
atributo?



(a) J.R. Quinlan, “Induction of Decision Trees”, Machine Learning, 1(1):81-106, 1986.

(b) J.R. Quinlan, “Simplifying Decision Trees”, Intl. Journal of Human-Computer Studies, 51(2):497-510, 1999.

# Entropía

- Entropía de una muestra  $S$  con respecto a una variable objetivo  $Y$ :

$$H_Y(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

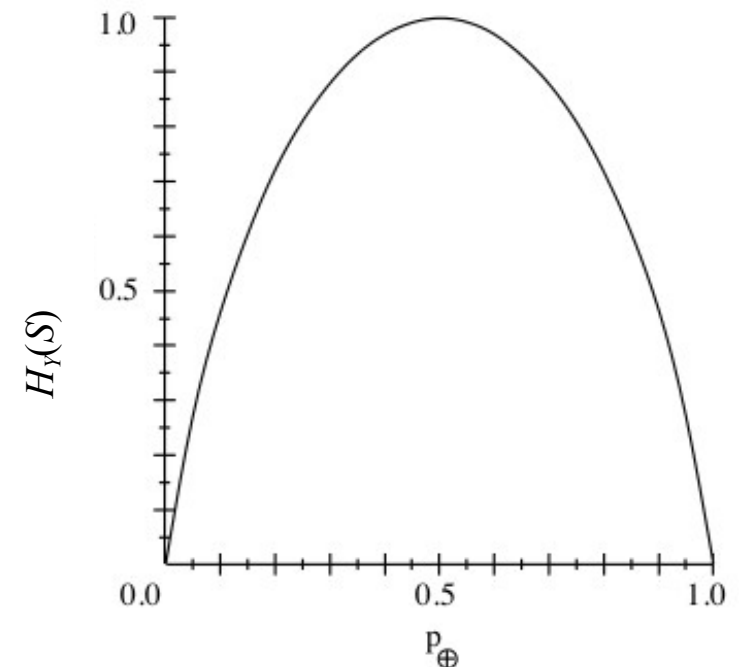
$c$  : cantidad de valores posibles de  $Y$

$p_i$  : probabilidad en  $S$  de que  $Y$  tome valor  $i$

- La entropía mide el **grado de impureza** de  $S$  respecto de  $Y$ .

- Ejemplo:  $c=2$

$$H_Y(S) = -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}$$



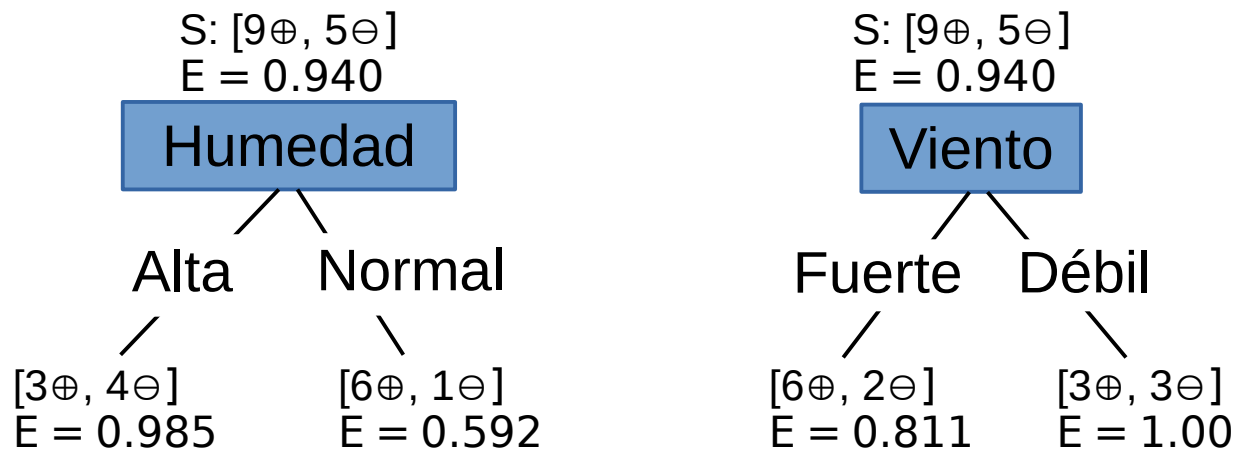
# Ganancia de Información

- Es la **reducción de entropía** de la muestra  $S$  (respecto de la variable objetivo  $Y$ ), después de clasificar las instancias según  $A$ .

$$Ganancia(S, A) = H(S) - \sum_{v \in Valores(A)} \frac{|S_v|}{|S|} H(S_v)$$

$Valores(A)$  : conjunto de valores posibles del atributo  $A$ .

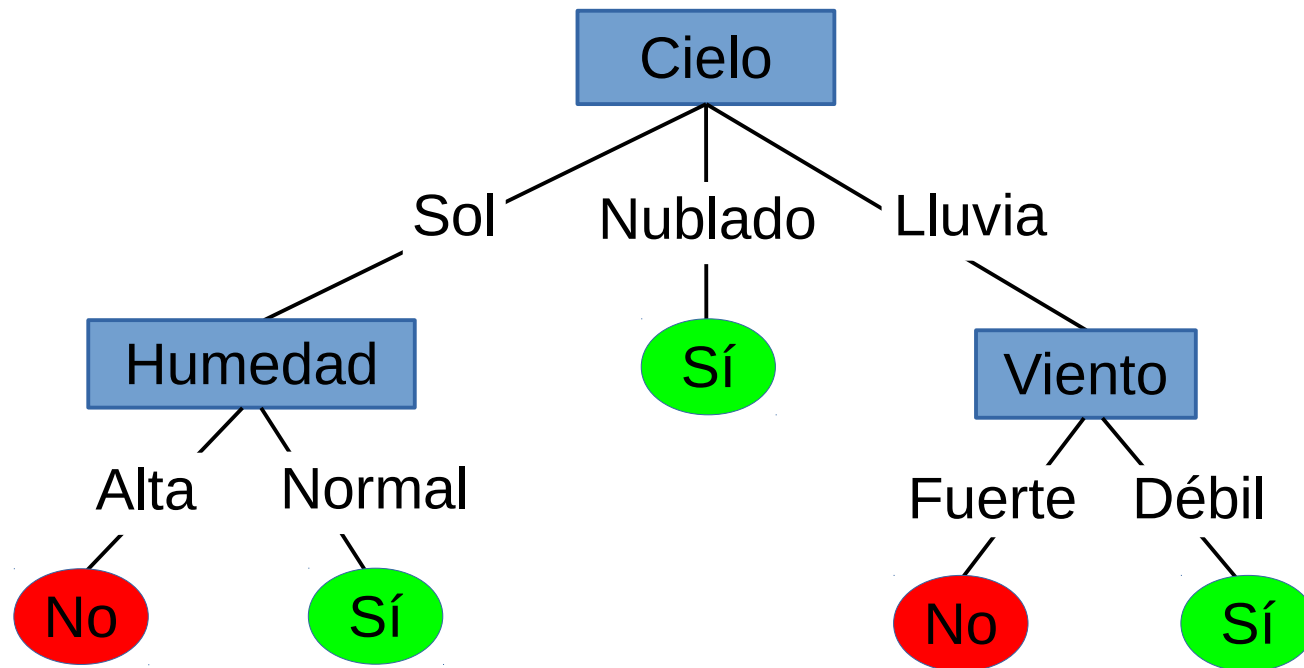
$$S_v = \{ s \in S \mid A(s) = v \}$$



$$Ganancia(S, Humedad) = .940 - (7/14) .985 - (7/14) .592 = \mathbf{0.151}$$

$$Ganancia(S, Viento) = .940 - (8/14) .811 - (6/14) 1.00 = \mathbf{0.048}$$

# Inducción *Top-Down* de Árboles de Decisión (ID3 y C4.5, Quinlan)



- Entonces, en cada nodo elegimos el atributo con mayor ganancia de información (el más *informativo*).
- Otra métrica: *Gain Ratio*.
  - Corrige preferencia de *Ganancia* por atributos con muchos (demasiados) valores.

# Inducción *Top-Down* de Árboles de Decisión (ID3 y C4.5, Quinlan)

- **Complejidad temporal** ( $n$ : #instancias,  $m$ : #atributos)
  - Construcción:  $O(n m^2)$  peor caso<sup>(a)</sup>,  $O(n m)$  promedio<sup>(b)</sup>
  - Consulta:  $O(m)$
- **Espacio de hipótesis:**
  - Espacio *completo* de funciones de valores discretos. En principio puede construirse cualquier árbol.
- **Sesgo inductivo:**
  - Construcción de árboles cada vez más complejos.
  - Hill-climbing sin backtracking (converge a máx. local).
  - Atributos más informativos  $\rightarrow$  cerca de la raíz.

---

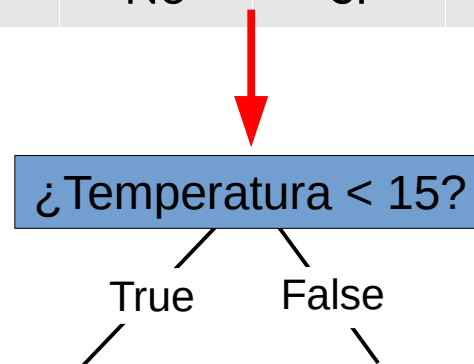
(a) P. E. Utgoff. "Incremental induction of decision trees". Machine Learning, 4(2):161–186, 1989

(b) J. W. Shavlik, R. J. Mooney, and G. Towell. "Symbolic and neural learning algorithm: An experimental comparison". Machine Learning, 6:111–143, 1991.

# Atributos Numéricos

- ¿Qué pasa si tenemos un atributo numérico  $A$ ?
- Buscamos un umbral  $c$ , para discriminar según  $A < c$ .
- ¿Cómo elegir  $c$ ?
  - 1) Ordenar las instancias según  $A$ .
  - 2) Buscar la forma de partir la lista que maximice la ganancia de información.

Temperatura:	10	12	18	21	28	31
Tenis:	No	No	Sí	Sí	Sí	No





# Para pensar...

- ¿Cuán robustos son los Árboles de Decisión ante...

## ...atributos faltantes?

- Instancias de entrenamiento con valores indefinidos en algunos atributos.
- Ej.: datos clínicos de un paciente incompletos.

## ...datos ruidosos?

- Instancias de entrenamiento mal clasificadas.
- Ej.: Errores cometidos en las anotaciones manuales.

# Resumen

- Árboles de decisión: construcción y consulta.
- Métricas para evaluar atributos basadas en la Teoría de la Información.
- Espacio de hipótesis. Sesgo inductivo. Complejidad temporal.
- Atributos discretos y numéricos.
- Robustez ante datos faltantes y ruidosos.