

Pontificia Universidad Católica de Chile  
Escuela de Ingeniería  
Departamento de Ciencia de la Computación



# IIC2613 – Inteligencia Artificial

## Árboles de decisión

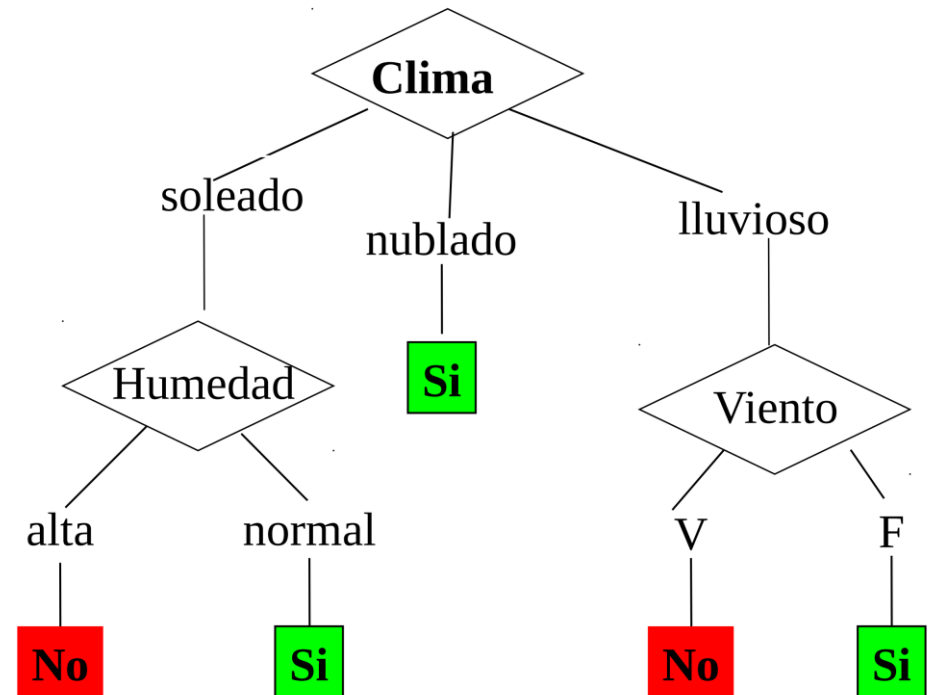
**Profesor:** Hans Löbel

¿Cómo solucionamos el siguiente problema de **clasificación**?

Clima	Temperatura	Humedad	Viento	Jugar?
soleado	alta	alta	F	No
soleado	alta	alta	V	No
nublado	alta	alta	F	Si
lluvioso	Agradable	alta	F	Si
lluvioso	frio	normal	F	Si
lluvioso	frio	normal	V	No
nublado	frio	normal	V	Si
soleado	Agradable	alta	F	No
soleado	frio	normal	F	Si
lluvioso	Agradable	normal	F	Si
soleado	Agradable	normal	V	Si
nublado	Agradable	alta	V	Si
nublado	alta	normal	F	Si
lluvioso	Agradable	alta	V	No

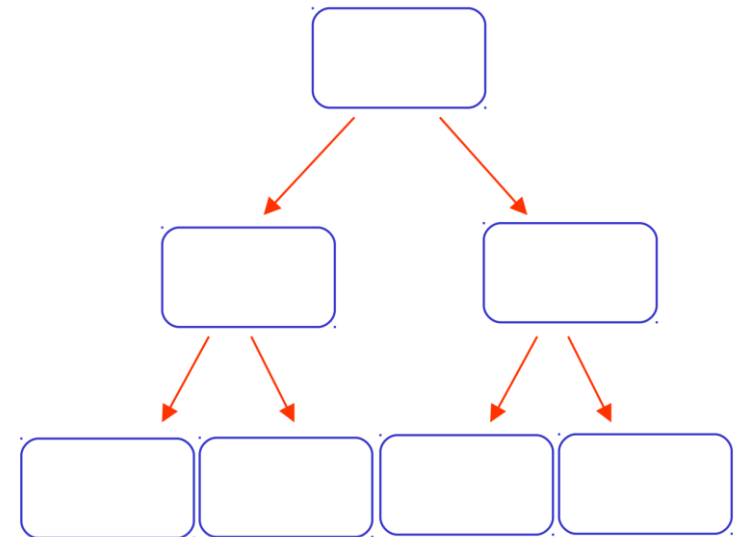
¿Cómo solucionamos el siguiente problema de **clasificación**?

Clima	Temperatura	Humedad	Viento	Jugar?
soleado	alta	alta	F	No
soleado	alta	alta	V	No
nublado	alta	alta	F	Si
lluvioso	Agradable	alta	F	Si
lluvioso	frio	normal	F	Si
lluvioso	frio	normal	V	No
nublado	frio	normal	V	Si
soleado	Agradable	alta	F	No
soleado	frio	normal	F	Si
lluvioso	Agradable	normal	F	Si
soleado	Agradable	normal	V	Si
nublado	Agradable	alta	V	Si
nublado	alta	normal	F	Si
lluvioso	Agradable	alta	V	No



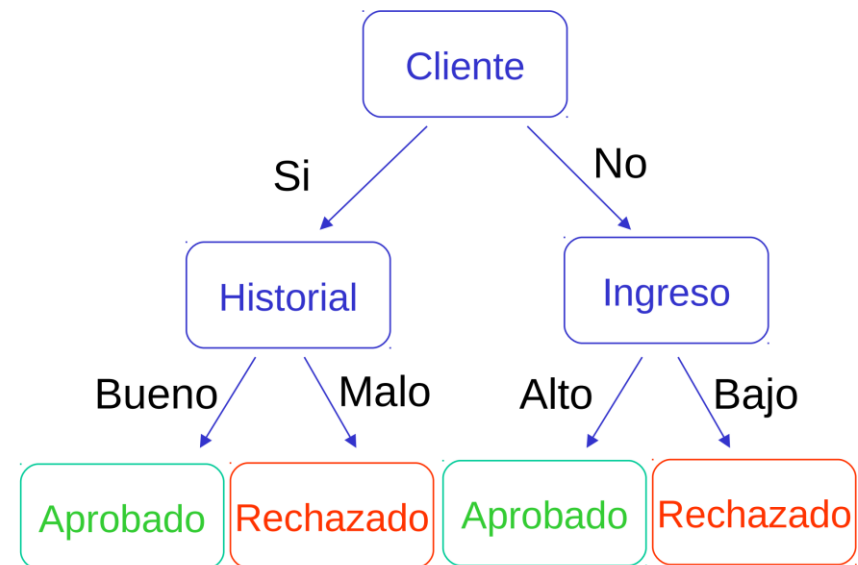
# Árboles de decisión pueden solucionar el caso anterior

- Técnica de aprendizaje supervisado.
- Pueden realizar clasificación y regresión.
- Pueden usarse sobre distintos tipos de variables (binaria, categórica, numérica).

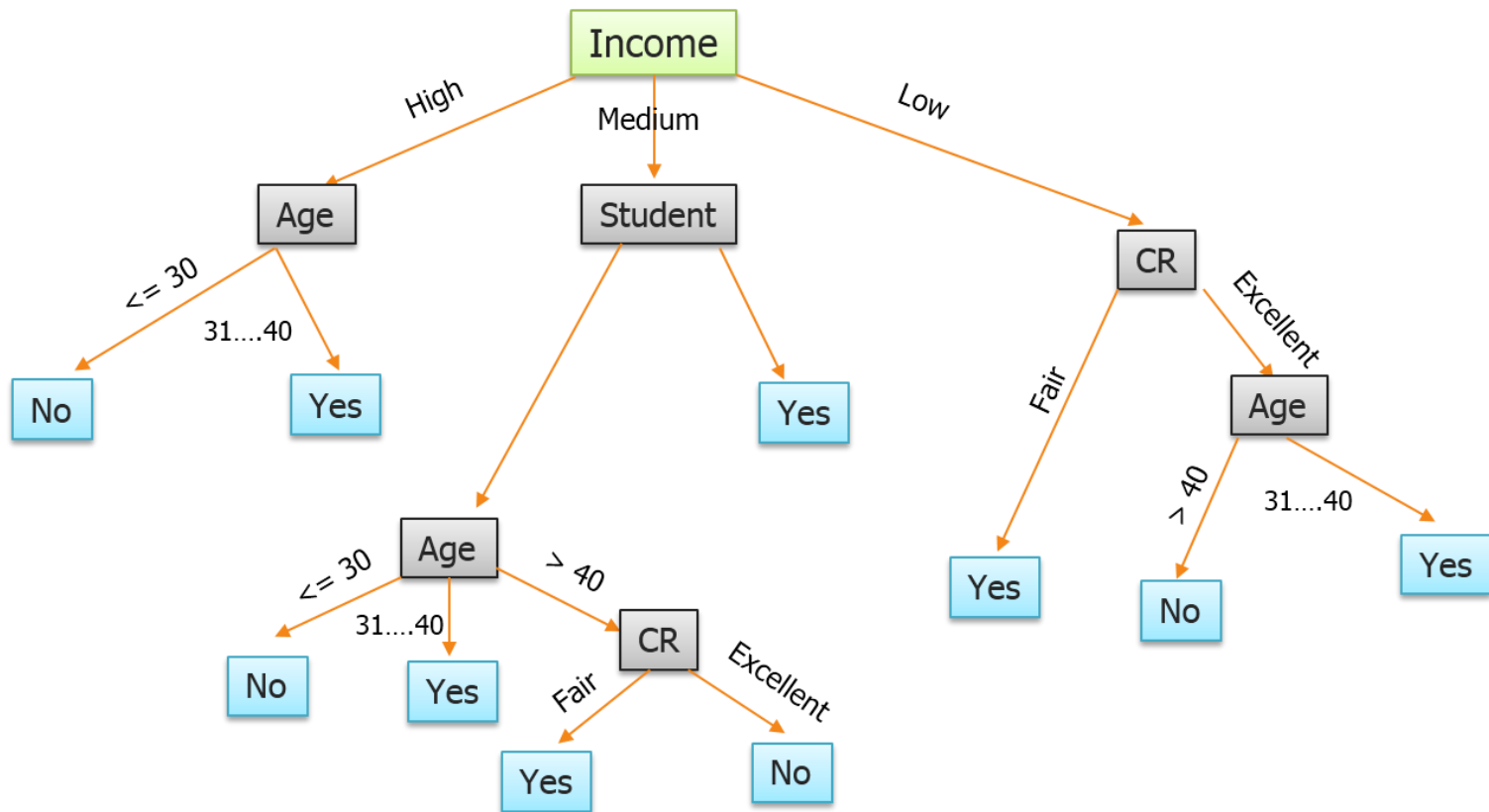


# Árboles de decisión son ampliamente utilizados en la práctica

- Cada nodo interno representa un atributo y cada nodo hoja representa una categoría.
- En cada nodo interno, se realiza un test en base a los valores del atributo.
- Aristas representan el resultado del test.
- Para clasificar un registro, se debe pasar **desde la raíz hasta alguna hoja**.

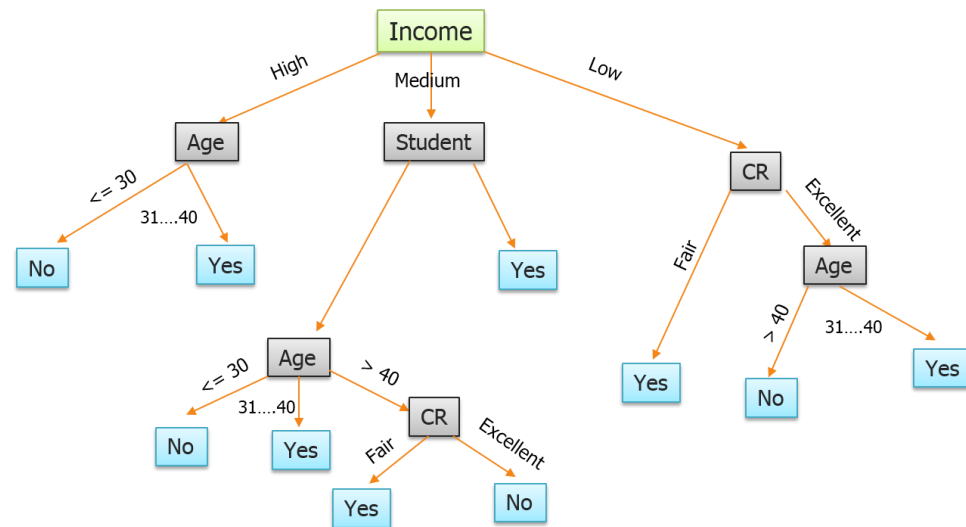


## ¿Cómo construimos un árbol de decisión?



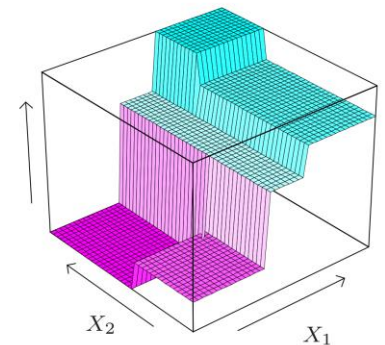
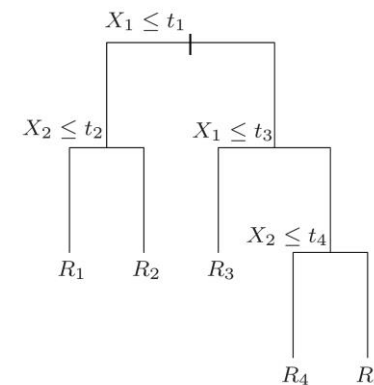
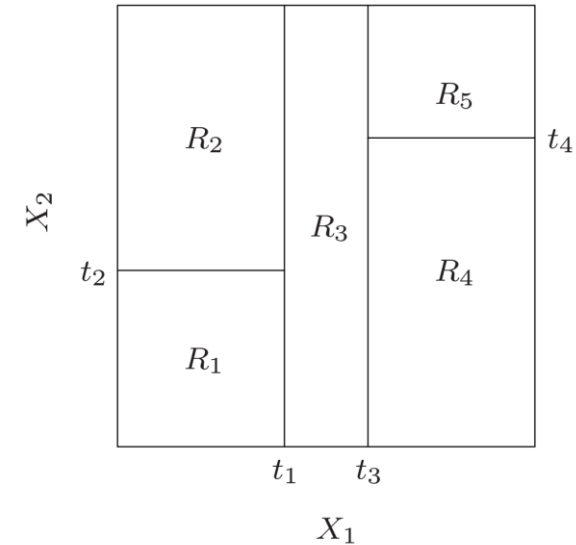
Podemos destilar esto en dos preguntas más específicas

1. ¿En qué orden realizo los tests?
2. ¿En que parte del dominio pongo el umbral de decisión? (atributos numéricos)



# Todo depende de cómo definamos lo que es **mejor**

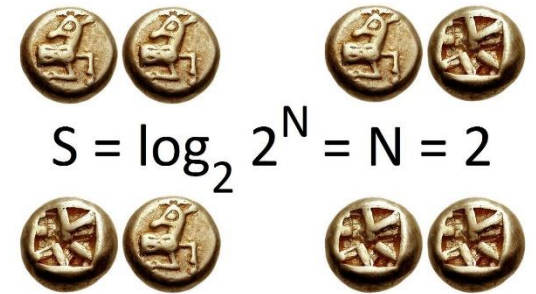
1. ¿Pertenece todos los registros a la misma clase?
  - Retornar marcando el nodo hoja con la clase respectiva.
2. ¿Tienen todos los registros el mismo valor para todos los atributos que determinan su clase?
  - Retornar marcando nodo hoja con la clase más común.
3. De lo contrario:
  - i. Seleccionar el atributo que **mejor** separa los registros de las distintas clases.
  - ii. Usar ese atributo como nodo raíz.
  - iii. Dividir el set de entrenamiento de acuerdo a este atributo y para cada rama resultante continuar la construcción del árbol en forma recursiva.





Todo depende de cómo definamos lo que es **mejor**

- Si objetivo es clasificar, es razonable que el **mejor atributo** separe mejor de acuerdo a las clases.
- Cuán homogéneo o impuro es un atributo, en función de las categorías.
- Dos maneras típicas de medir esto son:
  - Gini Index: desigualdad (inequidad) sobre distintas categorías.
  - **Information Entropy**: bits necesarios para codificar información.



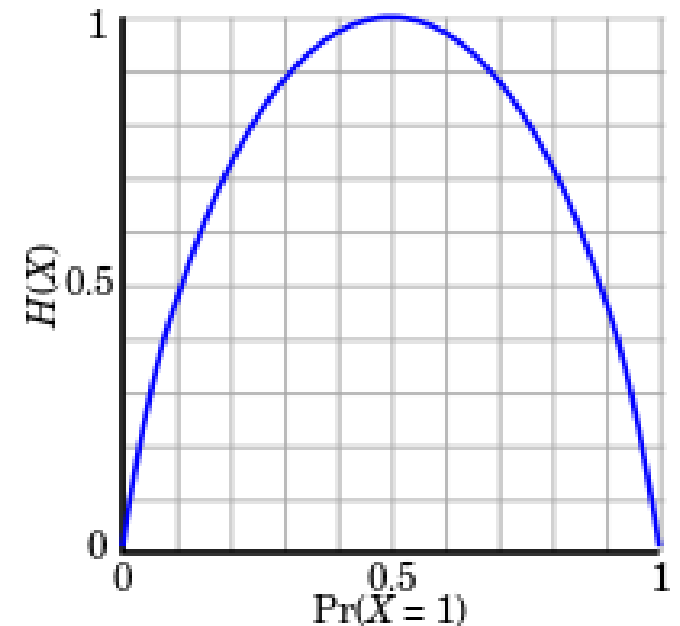
$$S = \log_2 2^N = N = 2$$

Entropía permite capturar de manera eficiente cuán homogénea es la distribución

- Intuitivamente, puede verse como un promedio ponderado de probabilidades de ocurrencia:

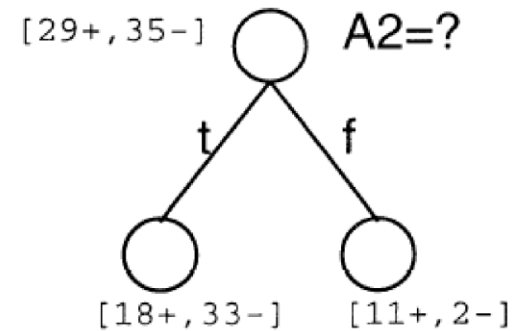
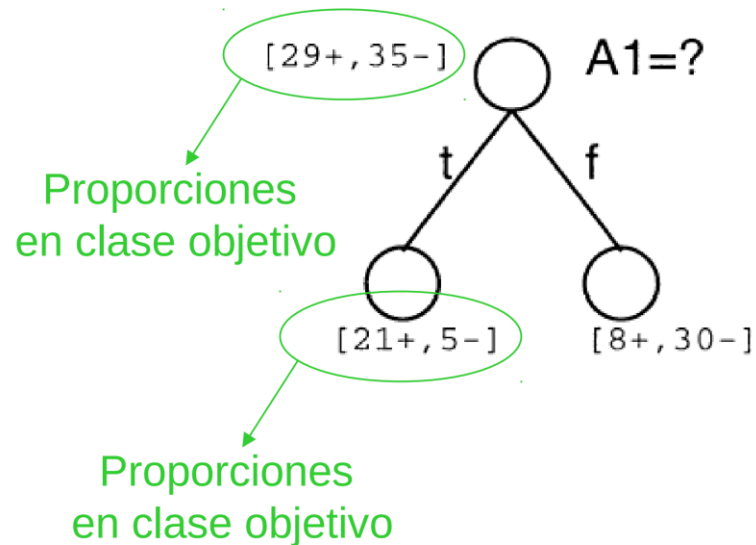
$$H(S) = - \sum_{c_i} p_i \log_2 p_i$$

- Por ejemplo:
  - 4 clases (A,B,C,D): 10 registros clase A, 20 clase B, 30 clase C, 40 clase D.
  - Entropía =  $-[(.1 \log .1) + (.2 \log .2) + (.3 \log .3) + (.4 \log .4)] = 1.85$ .



Elegimos el atributo que entrega la mayor **ganancia de información** (mayor reducción de entropía)

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

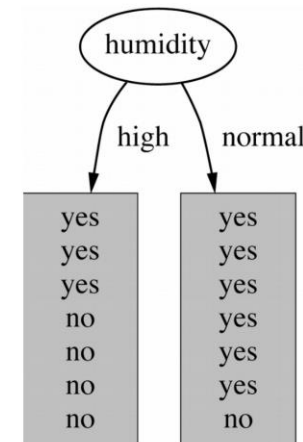
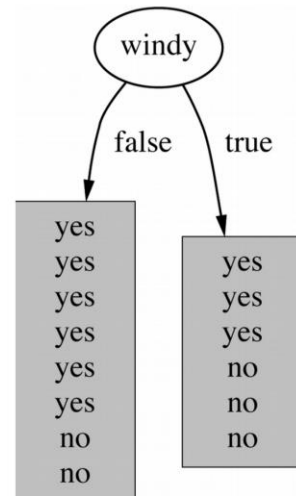
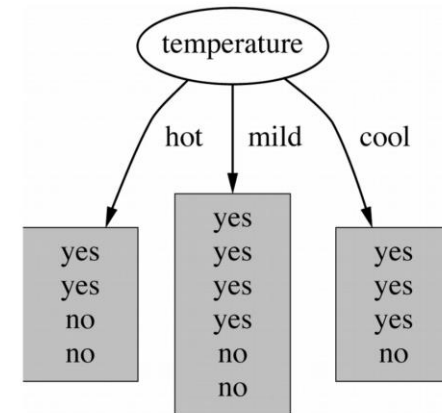
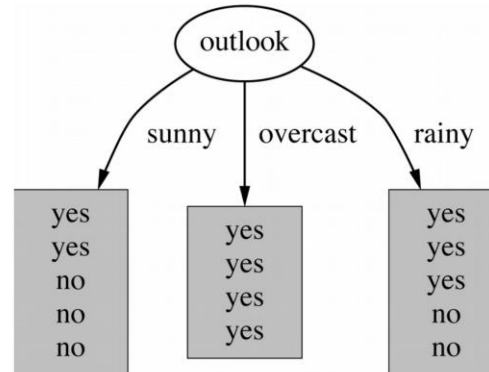


Elegimos el atributo que entrega la mayor **ganancia de información** (mayor reducción de entropía)

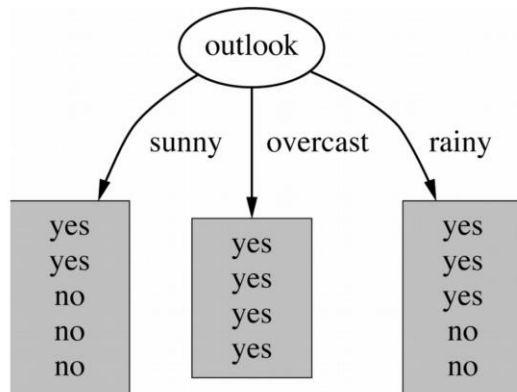
Day	Outlook	Temperature	Humidity	Wind	PlayTenn
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	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

Elegimos el atributo que entrega la mayor **ganancia de información** (mayor reducción de entropía)

Day	Outlook	Temperature	Humidity	Wind	PlayTenn
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



Elegimos el atributo que entrega la mayor **ganancia de información** (mayor reducción de entropía)



S:[9+,5-]  
E=0.940

**Outlook**

**Sunny**

**Overc.**

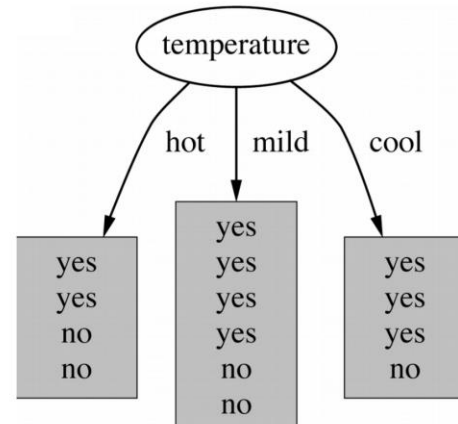
**Rainy**

[2+,3-]  
E=0.971

[4+,0-]  
E=0

[3+,2-]  
E=0.971

$$\text{Gain}(S, \text{Outlook}) = 0.940 - (5/14)0.971 - 0 - (5/14)0.971 = 0.266$$



temperature

hot

mild

cool

S:[9+,5-]  
E=0.940

**Temperat.**

**Hot**

**Mild**

**Cool**

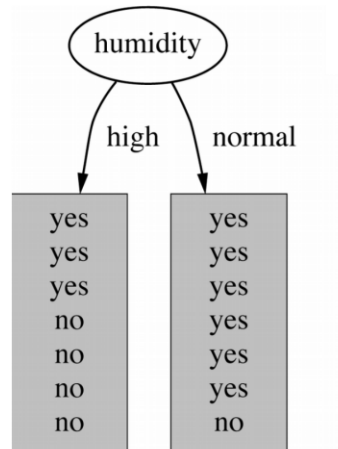
[2+,2-]  
E=1

[4+,2-]  
E=0.918

[3+,1-]  
E=0.811

$$\text{Gain}(S, \text{Temp.}) = 0.940 - (4/14) - (6/14)0.918 - (4/14)0.811 = 0.029$$

Elegimos el atributo que entrega la mayor **ganancia de información** (mayor reducción de entropía)



S:[9+,5-]  
E=0.940

**Humidity**

High

Normal

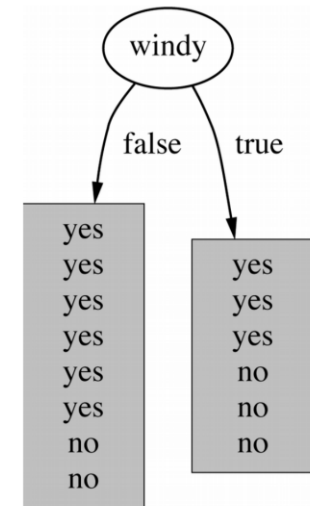
[3+,4-]

E=0.985

[6+,1-]

E=0.592

$$\text{Gain}(S, \text{Humidity}) = 0.940 - (7/14)0.985 - (7/14)0.592 = 0.151$$



S:[9+,5-]  
E=0.940

**Windy**

False

True

[6+,2-]

E=0.811

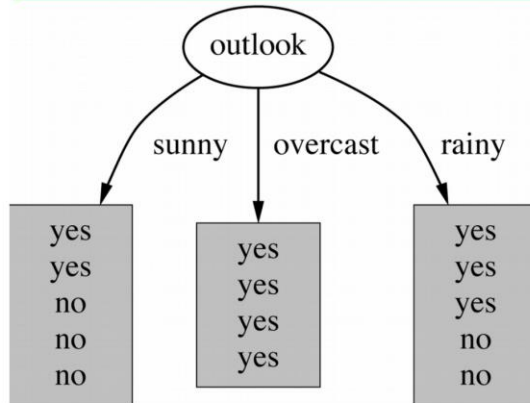
[3+,3-]

E=1

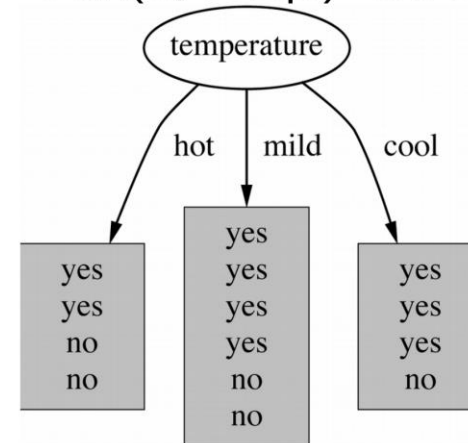
$$\text{Gain}(S, \text{Windy}) = 0.940 - (8/14)0.985 - (6/14)1 = 0.048$$

Elegimos el atributo que entrega la mayor **ganancia de información** (mayor reducción de entropía)

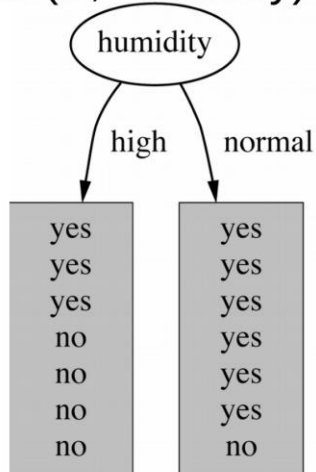
$$\text{Gain}(S, \text{Outlook})=0.266$$



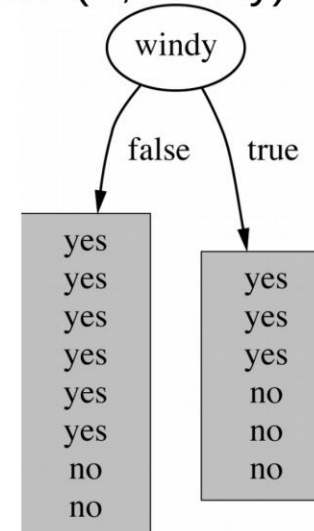
$$\text{Gain}(S, \text{Temp.})=0.029$$



$$\text{Gain}(S, \text{Humidity})=0.151$$



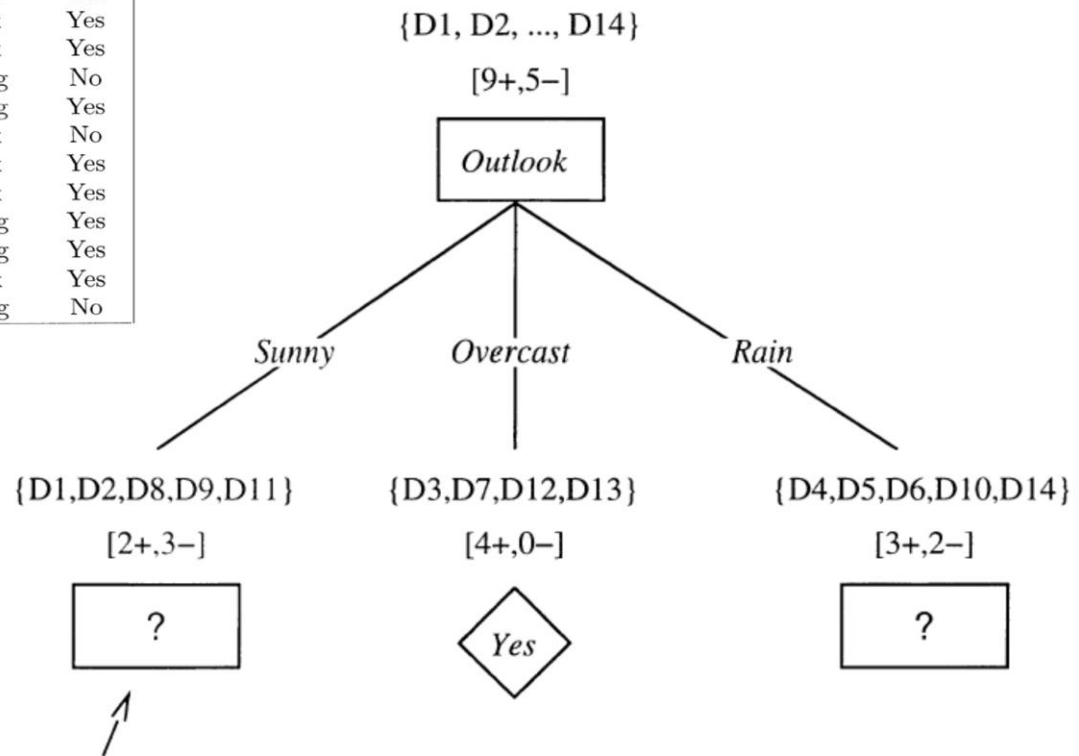
$$\text{Gain}(S, \text{Windy})=0.048$$





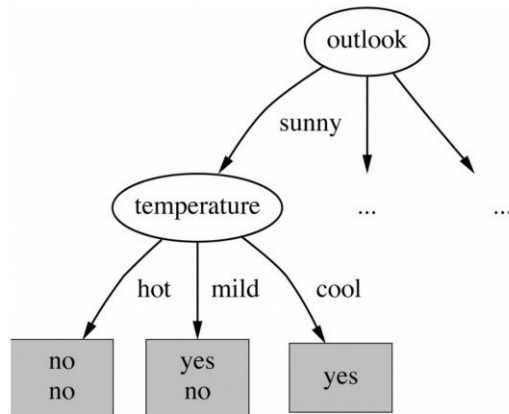
Elegimos el atributo que entrega la mayor **ganancia de información** (mayor reducción de entropía)

Day	Outlook	Temperature	Humidity	Wind	PlayTenn
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

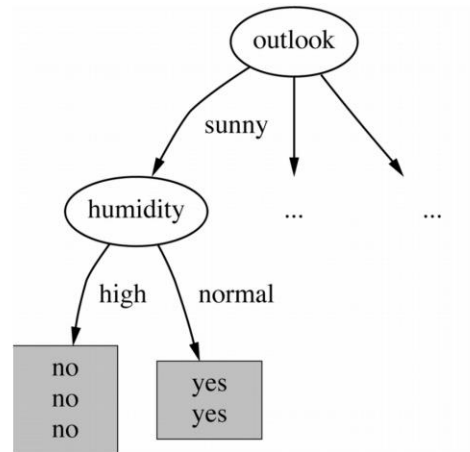


¿Cuál atributo?

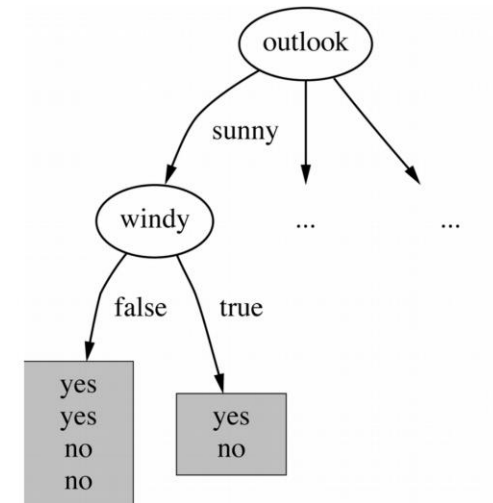
Elegimos el atributo que entrega la mayor **ganancia de información** (mayor reducción de entropía)



Gain(S,Temp.)=?

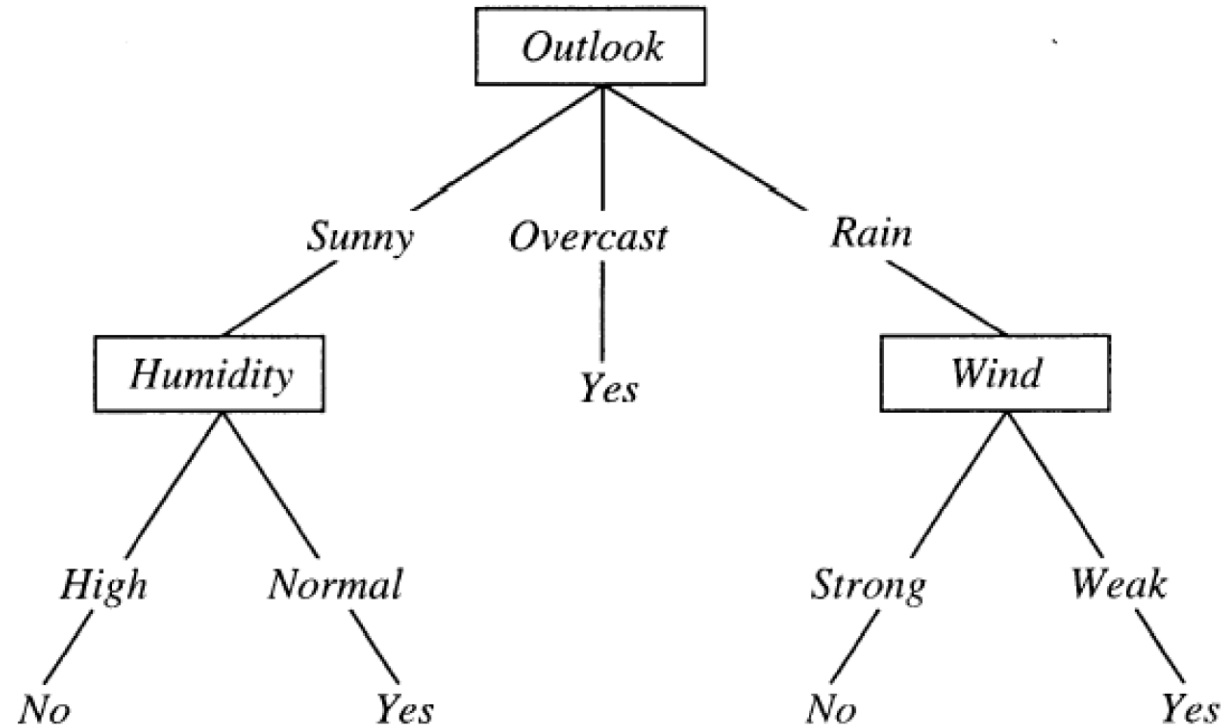


Gain(humid.,Temp.)=?



Gain(S,Windy)=?

Elegimos el atributo que entrega la mayor **ganancia de información** (mayor reducción de entropía)



¿Qué pasa con IG si hay muchos posibles valores para los atributos?

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

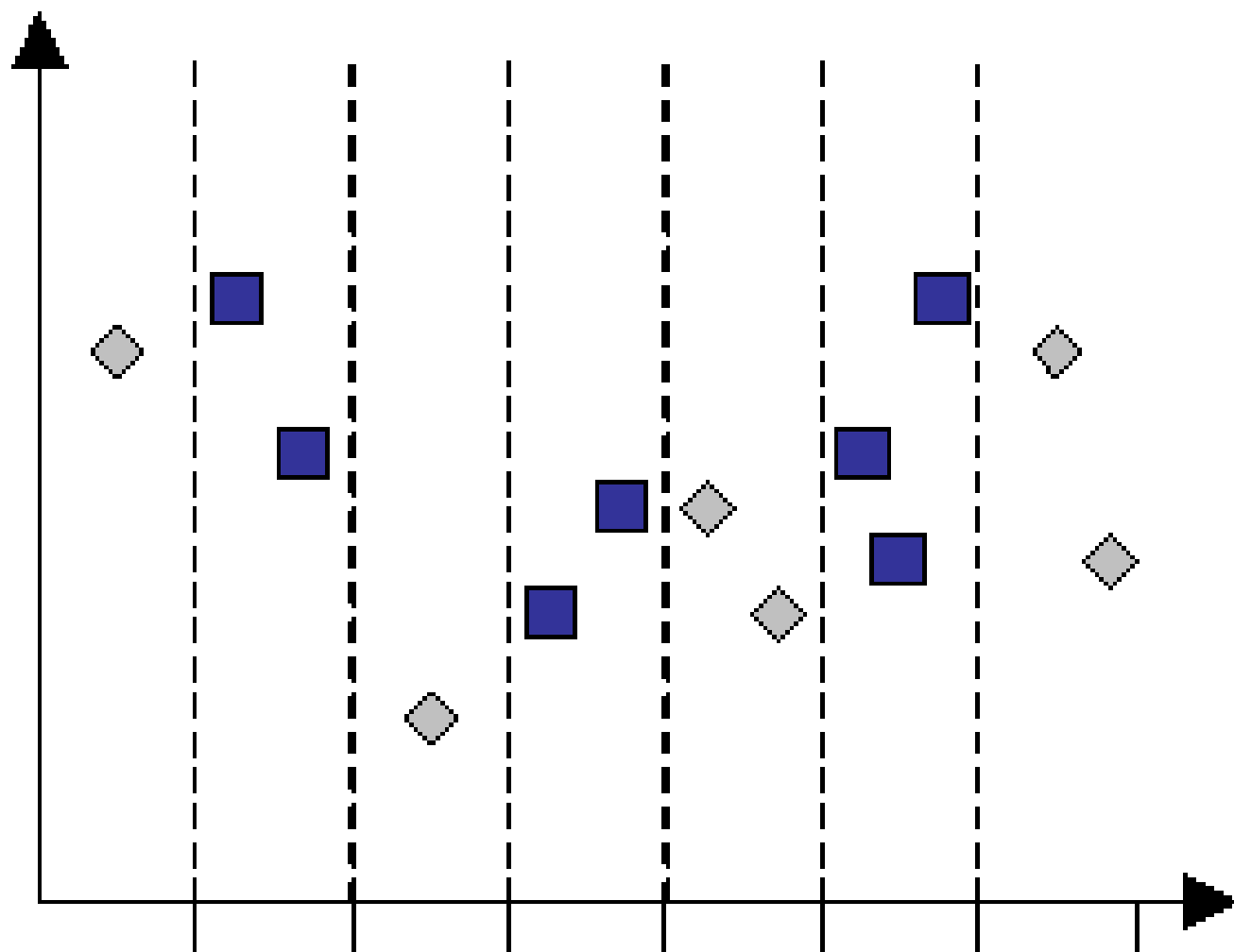
Si IG anda mal, se puede usar **Gain Ratio**

$$GainRatio(S, A) \equiv \frac{Gain(S, A)}{SplitInformation(S, A)}$$

$$SplitInformation(S, A) \equiv - \sum_{i=1}^c \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$$

## Quedan aún dos preguntas relevantes

- ¿Cómo funciona este algoritmo si tenemos variables numéricas?
  1. Fuerza bruta
  2. Ordenar por dimensión, y evaluar split en cada cambio de categoría.
- ¿Cómo se puede hacer regresión usando árboles de este tipo?



## Quedan aún dos preguntas relevantes

1. ¿Cómo funciona este algoritmo si tenemos variables numéricas?
  - Fuerza bruta
  - Ordenar por dimensión, y evaluar split en cada cambio de categoría.
2. ¿Cómo se puede hacer regresión usando árboles de este tipo?
  - En vez de medir IG, se calcula la desviación cuadrática con respecto a la media.
  - Construcción recursiva sigue la misma idea que para clasificación.

