https://media-exp3.licdn.com/dms/image/

RECONOCIMIENTO DE PATRONES EN IMÁGENES TICS 585

FACULTAD DE INGENIERÍA Y CIENCIAS
UNIVERSIDAD ADOLFO IBÁÑEZ

SEGUNDO SEMESTRE 2021

PROFESOR: MIGUEL CARRASCO

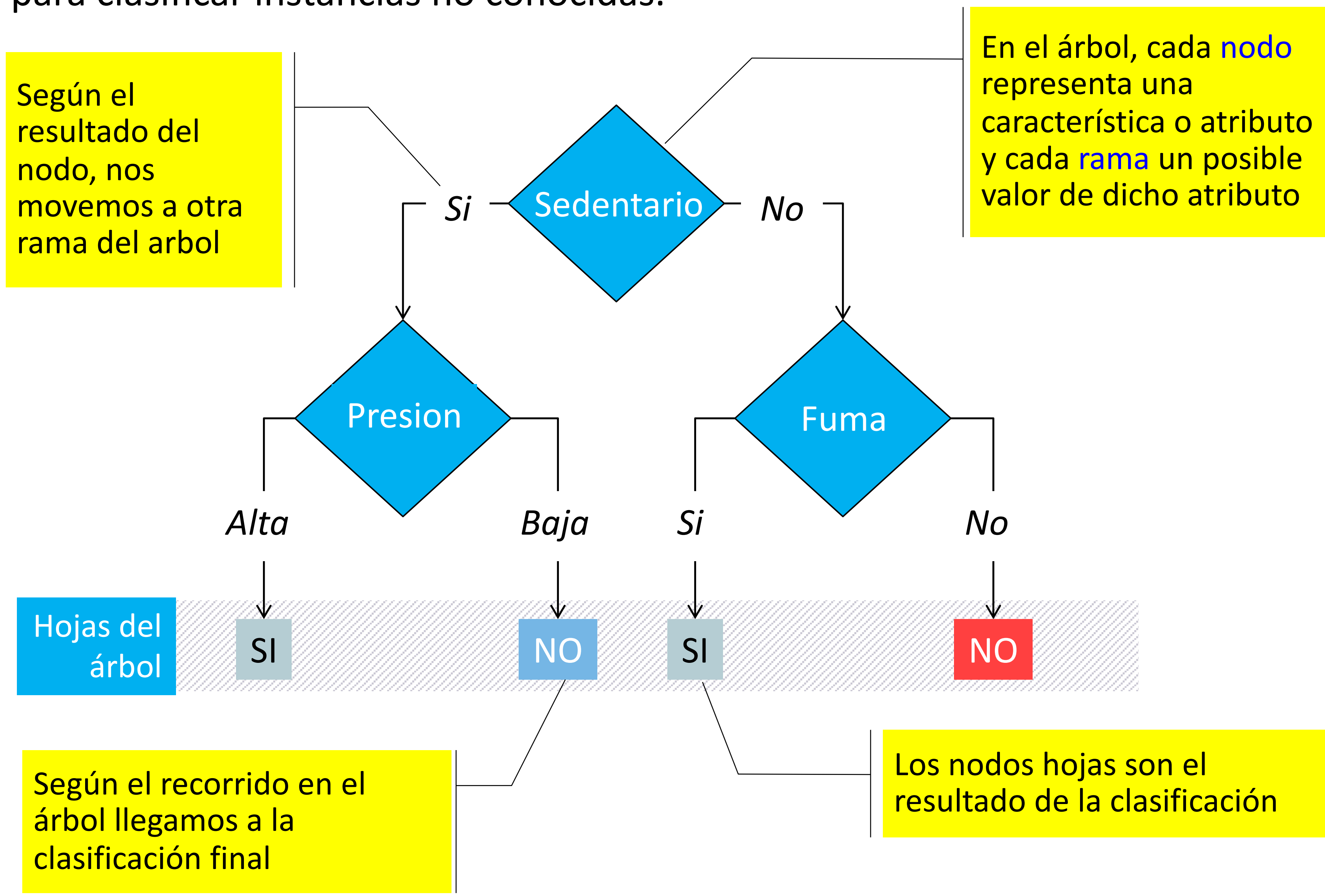
CLASIFICADORES SUPERVISADOS

Supervisados

- kNN
- Mahalanobis
- LDA
- **Árboles de decisión**
- Bayesiano
- Redes Neuronales

Árboles de decisión:

Es una técnica de clasificación supervisada que genera árboles de decisión para clasificar instancias no conocidas.



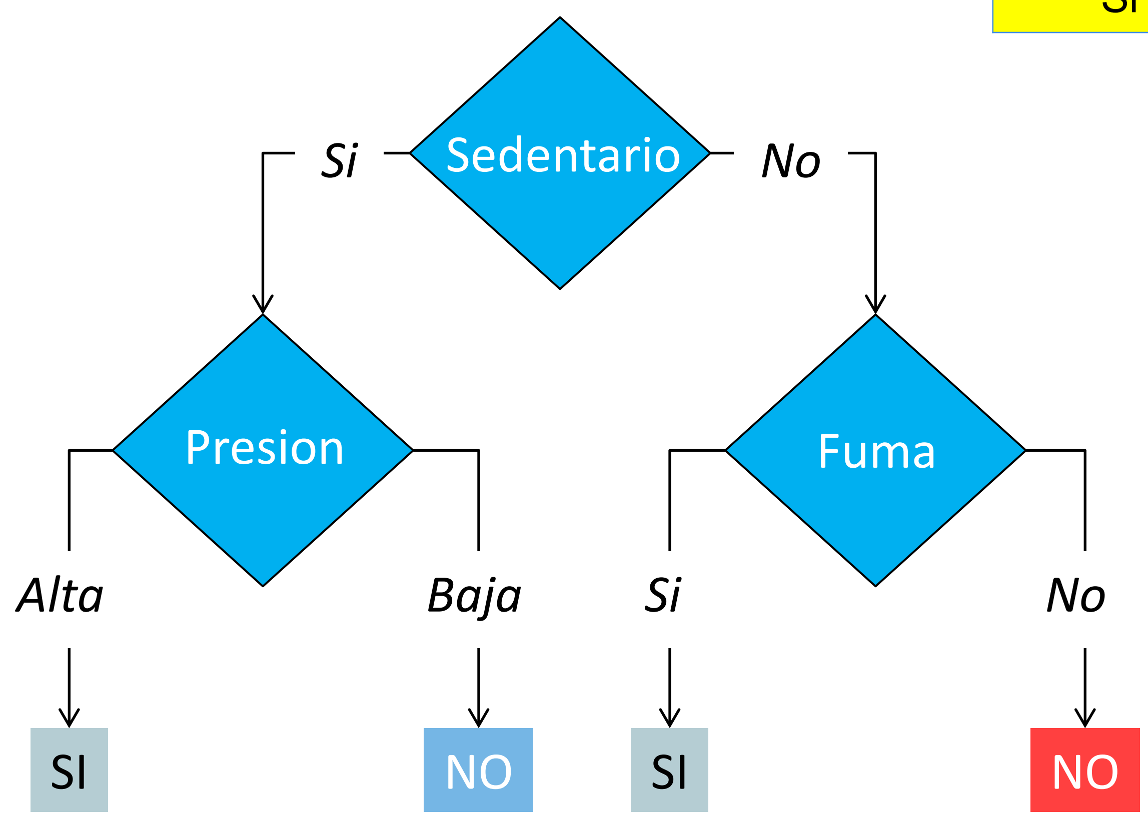
Supervisados

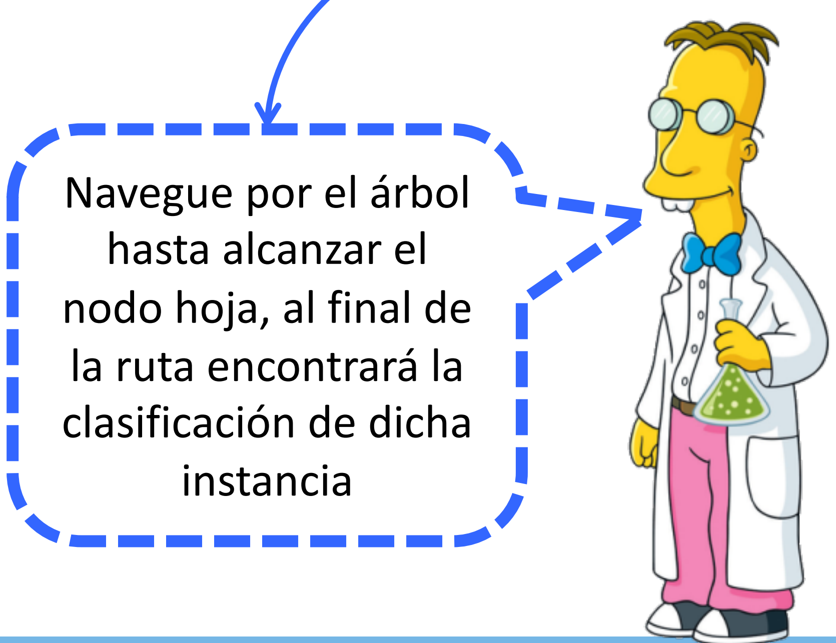
- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- **Clasificación**

  Una vez generado el árbol, para clasificar una nueva instancia recorremos el árbol desde el nodo raíz, evaluando la respuesta de cada nodo, hasta llegar al nodo hoja.

Características			Clase
Sedentario	Fumador	Presion	Riesgo
Si	Si	Alta	Si
Si	Si	Baja	No
Si	No	Baja	No
No	Si	Alto	Si
No	No	Baja	No
Si	Si	Alta	?





FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

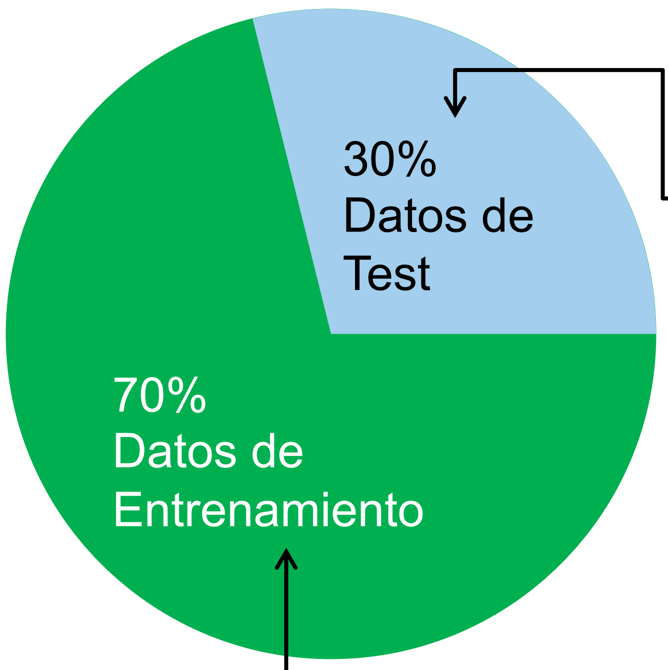
Bayesiano

Redes Neuronales

- **Construcción del árbol**

  La construcción del árbol cumple un principio inductivo. Esto significa que si el clasificador funciona bien para el conjunto de entrenamiento, suponemos que lo hará bien en nuevas instancias.

  En general el árbol funciona bien si las instancias de entrenamiento son representativas del problema. Para ello debemos seleccionar un conjunto de datos para entrenar y otro para testear (conjuntos disjuntos)



El conjunto de testing nos permite evaluar el modelo en instancias que no son conocidas para éste.

El conjunto de entrenamiento debe ser escogido aleatoriamente. De esta forma aseguramos que sea representativo.

- Supervisados

kNN

Mahalanobis

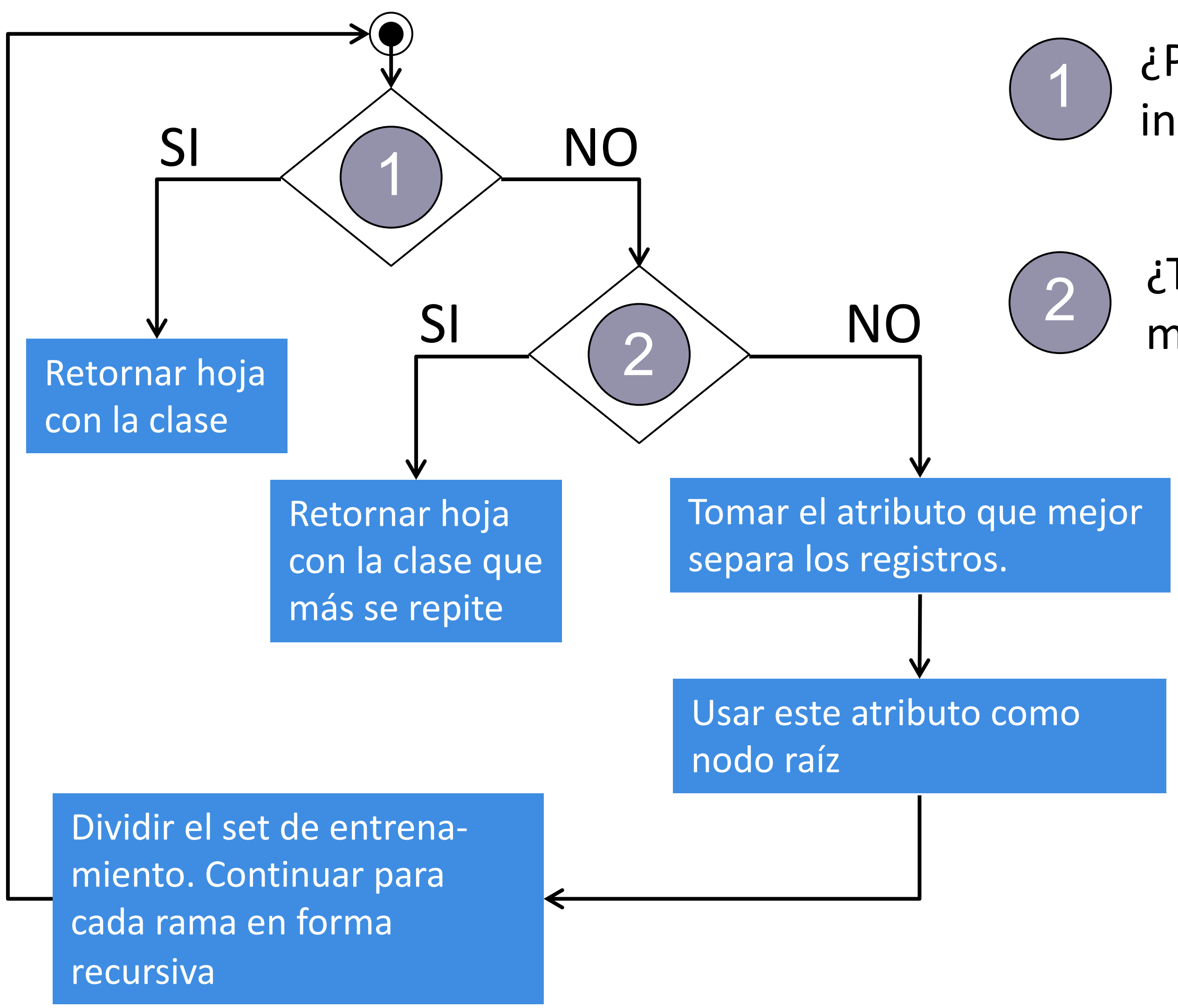
LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- **Construcción del árbol:**
  Una forma para construir el árbol de decisión consiste en el siguiente pseudoalgoritmo:



1 ¿Pertenecen todas las instancias a la misma clase?

2 ¿Tiene todas las instancias el mismo valor ?

FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- Construcción del árbol (Algoritmo ID3. Quinlan, 1986)
  Para determinar que atributo seleccionar, empleamos un índice denominado Ganancia de Información. Ésta mide la reducción esperada de entropía dado una característica.

Características o Atributos — Clase

Día	Clima	Temperatura	Humedad	Viento	¿Juego Tenis?
1	Soleado	Calor	Alta	Débil	NO
2	Soleado	Calor	Alta	Fuerte	NO
3	Cubierto	Calor	Alta	Débil	SI
4	Lluvia	Media	Alta	Débil	SI
5	Lluvia	Fria	Normal	Débil	SI
6	Lluvia	Fria	Normal	Fuerte	NO
7	Cubierto	Fria	Normal	Fuerte	SI
8	Soleado	Media	Alta	Débil	NO
9	Soleado	Fria	Normal	Débil	SI
10	Lluvia	Media	Normal	Débil	SI
11	Soleado	Media	Normal	Fuerte	SI
12	Cubierto	Media	Alta	Fuerte	SI
13	Cubierto	Calor	Normal	Débil	SI
14	Lluvia	Media	Alta	Fuerte	NO

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- **PRIMERO:**
Para cada característica o atributo, calculamos la tabla clases correspondientes al atributo seleccionado.

Día	Clima	¿Juego Tenis?
1	Soleado	NO
2	Soleado	NO
3	Cubierto	SI
4	Lluvia	SI
5	Lluvia	SI
6	Lluvia	NO
7	Cubierto	SI
8	Soleado	NO
9	Soleado	SI
10	Lluvia	SI
11	Soleado	SI
12	Cubierto	SI
13	Cubierto	SI
14	Lluvia	NO

CLASES =

ATRIBUTO =

RANGOS =

CLIMA	SI	NO	Σ
Soleado	2	3	5
Nublado	4	0	4
Lluvioso	3	2	5
Σ	9	5	14

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- **SEGUNDO:**
  Calculamos la entropía de la clase. La Entropía es el máximo número de bits requeridos para codificar la clase de un ejemplo. A mayor valor de Entropía, significa que un suceso es menos probable.

  - Entropía:

$$H(S) = -\sum_{C_i}^{n} p_i \log_2(p_i)$$

$H(S)$ = Entropía del conjunto S

c_i = Subcojunto de registros de la clase i-ésima

p_i = Proporción de registros de la clase c_i el conjunto S

  - Entropía del atributo *clima*

$$H(\text{Clima}) = -\frac{9}{14} \times \log_2\left(\frac{9}{14}\right) - \frac{5}{14} \times \log_2\left(\frac{5}{14}\right) = 0.94$$

	CLASES	
CLIMA	SI	NO
Soleado	2	3
Cubierto	4	0
Lluvioso	3	2
Σ	9	5

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- **SEGUNDO:**
  Calculamos la efectividad de los subvalores según un atributo a través de la suma parcial de la Entropía según sus rangos definidos.

  - Efectividad:

$$H(S,X)=\sum_{i=1}^{n}\frac{|S_i|}{|S_X|}H(S_i)$$

$H(S,X)$ = Entropía del conjunto S en el valor posible X
$|S_i|$ = Número de registros de clase i-$ésima$
$|S_X|$ = Número total de registros del subconjunto X

  - Efectividad del atributo *clima* en cada rango

$$H(\text{Clima, Soleado})=-\frac{2}{5}\times\log_2\left(\frac{2}{5}\right)-\frac{3}{5}\times\log_2\left(\frac{3}{5}\right)=0.971$$

$$H(\text{Clima, Cubierto})=-\frac{4}{4}\times\log_2\left(\frac{4}{4}\right)-0\times\log_2\left(\frac{0}{4}\right)=0$$

$$H(\text{Clima, Lluvioso})=-\frac{3}{5}\times\log_2\left(\frac{3}{5}\right)-\frac{2}{5}\times\log_2\left(\frac{2}{5}\right)=0.971$$

	CLASES		
CLIMA	SI	NO	Σ
Soleado	2	3	5
Cubierto	4	0	4
Lluvioso	3	2	5

Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- **TERCERO:**
  Calculamos la *ganancia de información* de cada característica (o atributo). La ganancia de información es la reducción esperada en entropía al seleccionar un atributo

  - Ganancia de Información:

$$G(S,X) = H(S) - \sum_{i=1} \frac{|S_i|}{|S|} \times H(S,X)$$

S_i = Subconjunto de registros en X
$|S|$ = Número total de registros de clase S

CLIMA	SI	NO	Σ	H
Soleado	2	3	5	0.971
Cubierto	4	0	4	0
Lluvioso	3	2	5	0.971

La ganancia de información nos indica qué atributo es el que provee mayor información, es decir, qué suceso es más probable que ocurra.

$$G(\text{Clima}) = H(\text{Clima}) - \left(\frac{5}{14} H(\text{Clima, Sol}) + \frac{4}{14} H(\text{Clima, Cub}) + \frac{5}{14} H(\text{Clima, Lluv}) \right)$$

$$G(\text{Clima}) = 0.94 - \left(\frac{5}{14} \times 0.971 + \frac{4}{14} \times 0 + \frac{5}{14} \times 0.971 \right) = 0.2467$$

Ganancia del atributo Clima

FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- CUARTO: Seleccionamos como nodo raíz aquel atributo que tenga la mayor ganancia de información.

Clima	CLASE
Soleado	NO
Soleado	NO
Cubierto	SI
Lluvia	SI
Lluvia	SI
Lluvia	NO
Cubierto	SI
Soleado	NO
Soleado	SI
Lluvia	SI
Soleado	SI
Cubierto	SI
Cubierto	SI
Lluvia	NO

Temperatura	CLASE
Calor	NO
Calor	NO
Calor	SI
Media	SI
Fria	SI
Fria	NO
Fria	SI
Media	NO
Fria	SI
Media	SI
Media	SI
Media	SI
Calor	SI
Media	NO

Humedad	CLASE
Alta	NO
Alta	NO
Alta	SI
Alta	SI
Normal	SI
Normal	NO
Normal	SI
Alta	NO
Normal	SI
Normal	SI
Normal	SI
Alta	SI
Normal	SI
Alta	NO

Viento	CLASE
Débil	NO
Fuerte	NO
Débil	SI
Débil	SI
Débil	SI
Fuerte	NO
Fuerte	SI
Débil	NO
Débil	SI
Débil	SI
Fuerte	SI
Fuerte	SI
Débil	SI
Fuerte	NO

CLIMA	SI	NO
Soleado	2	3
Cubierto	4	0
Lluvioso	3	2

TEMP	SI	NO
Calor	2	2
Media	4	2
Fria	3	1

HUMED	SI	NO
Alta	3	4
Normal	6	1

VIENTO	SI	NO
Débil	6	2
Fuerte	3	3

FACULTAD DE INGENIERÍA Y CIENCIAS

UAI UNIVERSIDAD ADOLFO IBÁÑEZ

Supervisados

- kNN
- Mahalanobis
- LDA
- **Árboles de decisión**
- Bayesiano
- Redes Neuronales

- CUARTO:
  Seleccionamos como nodo raíz aquel atributo que tenga la mayor ganancia de información.

CLIMA	SI	NO
Soleado	2	3
Cubierto	4	0
Lluvioso	3	2

$$H(\text{Clima}) = -\frac{9}{14}\times\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\times\log_2\left(\frac{5}{14}\right) = 0.94$$

TEMP	SI	NO
Calor	2	2
Media	4	2
Fria	3	1

$$H(\text{Temperatura}) = -\frac{9}{14}\times\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\times\log_2\left(\frac{5}{14}\right) = 0.94$$

HUMED	SI	NO
Alta	3	4
Normal	6	1

$$H(\text{Humedad}) = -\frac{9}{14}\times\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\times\log_2\left(\frac{5}{14}\right) = 0.94$$

VIENTO	SI	NO
Débil	6	2
Fuerte	3	3

$$H(\text{Viento}) = -\frac{9}{14}\times\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\times\log_2\left(\frac{5}{14}\right) = 0.94$$

Supervisados

- kNN
- Mahalanobis
- LDA
- **Árboles de decisión**
- Bayesiano
- Redes Neuronales

CUARTO:

Seleccionamos como nodo raíz aquel atributo que tenga la mayor ganancia de información.

CLIMA	SI	NO
Soleado	2	3
Cubierto	4	0
Lluvioso	3	2

$$H(\text{Clima, Soleado}) = -\frac{2}{5}\times\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\times\log_2\left(\frac{3}{5}\right) = 0.971$$

$$H(\text{Clima, Nublado}) = -\frac{4}{4}\times\log_2\left(\frac{4}{4}\right) - 0\times\log_2\left(\frac{0}{4}\right) = 0$$

$$H(\text{Clima, Lluvioso}) = -\frac{3}{5}\times\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\times\log_2\left(\frac{2}{5}\right) = 0.971$$

TEMP	SI	NO
Calor	2	2
Media	4	2
Fria	3	1

$$H(\text{Temperatura, Calor}) = -\frac{2}{4}\times\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\times\log_2\left(\frac{2}{4}\right) = 1$$

$$H(\text{Temperatura, Media}) = -\frac{4}{6}\times\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\times\log_2\left(\frac{2}{6}\right) = 0.918$$

$$H(\text{Temperatura, Fria}) = -\frac{3}{4}\times\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\times\log_2\left(\frac{1}{4}\right) = 0.811$$

HUMED	SI	NO
Alta	3	4
Normal	6	1

$$H(\text{Humedad, Alta}) = -\frac{3}{7}\times\log_2\left(\frac{3}{7}\right) - \frac{4}{7}\times\log_2\left(\frac{4}{7}\right) = 0.985$$

$$H(\text{Humedad, Normal}) = -\frac{6}{7}\times\log_2\left(\frac{6}{7}\right) - \frac{1}{7}\times\log_2\left(\frac{1}{7}\right) = 0.591$$

VIENTO	SI	NO
Débil	6	2
Fuerte	3	3

$$H(\text{Viento, Débil}) = -\frac{6}{8}\times\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\times\log_2\left(\frac{2}{8}\right) = 0.811$$

$$H(\text{Viento, Fuerte}) = -\frac{3}{6}\times\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\times\log_2\left(\frac{3}{6}\right) = 1$$

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- CUARTO:
  Seleccionamos como nodo raíz aquel atributo que tenga la mayor ganancia de información.

$$H(\text{Clima}) = 0.94$$

	registros
$H(\text{Clima, Soleado}) = 0.971$	5/14
$H(\text{Clima, Cubierto}) = 0$	4/14
$H(\text{Clima, Lluvioso}) = 0.971$	5/14

$$H(\text{Humedad}) = 0.94$$

	registros
$H(\text{Humedad, Alta}) = 0.985$	7/14
$H(\text{Humedad, Normal}) = 0.591$	7/14

$$H(\text{Temperatura}) = 0.94$$

	registros
$H(\text{Temperatura, Calor}) = 1$	4/14
$H(\text{Temperatura, Media}) = 0.918$	6/14
$H(\text{Temperatura, Fria}) = 0.811$	4/14

$$H(\text{Viento}) = 0.94$$

	registros
$H(\text{Viento, Débil}) = 0.811$	8/14
$H(\text{Viento, Fuerte}) = 1$	6/14

- Ganancia de Información

$$G(\text{Clima}) = 0.94 - \left(\frac{5}{14}\times 0.971 + \frac{4}{14}\times 0 + \frac{5}{14}\times 0.971\right) = \boxed{0.2467}$$

$$G(\text{Temperatura}) = 0.94 - \left(\frac{4}{14}\times 1 + \frac{6}{14}\times 0.918 + \frac{4}{14}\times 0.811\right) = 0.029$$

$$G(\text{Humedad}) = 0.94 - \left(\frac{7}{14}\times 0.985 + \frac{7}{14}\times 0.591\right) = 0.15$$

$$G(\text{Viento}) = 0.94 - \left(\frac{8}{14}\times 0.811 + \frac{6}{14}\times 1\right) = 0.0485$$

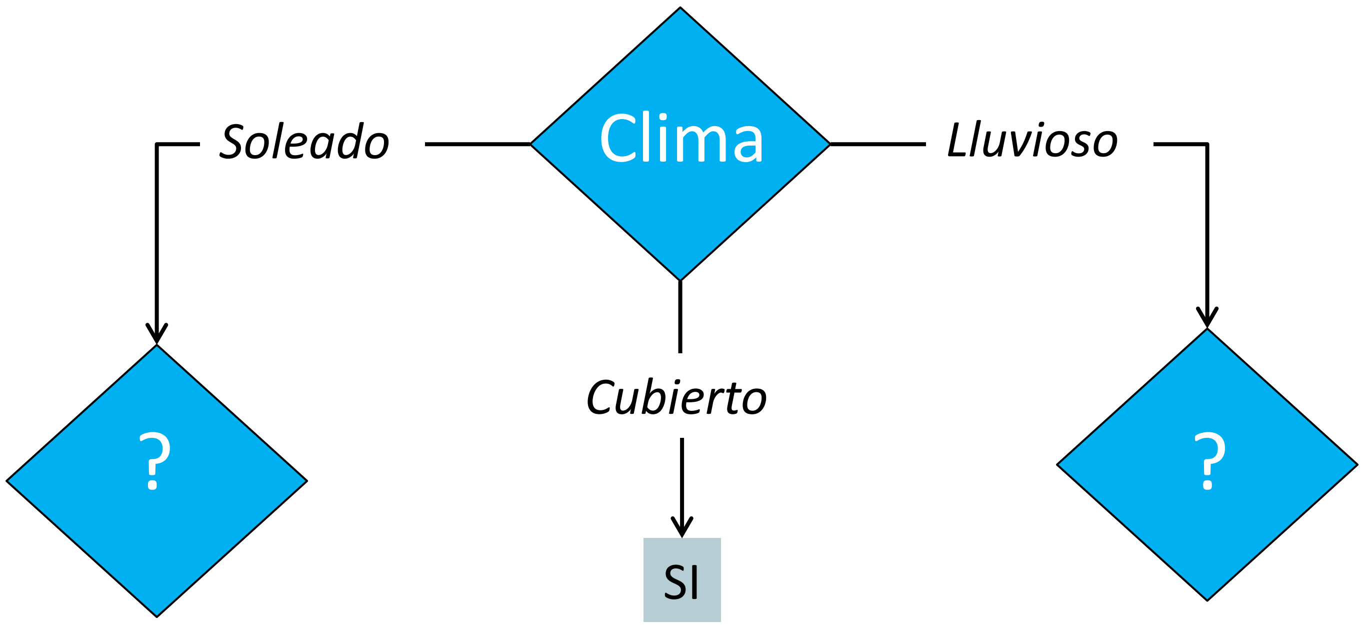
La mayor Ganancia de información corresponde al atributo *Clima*. Por lo tanto este será el nodo raíz.

FACULTAD DE INGENIERÍA Y CIENCIAS
UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- **Primer Nivel:**
  Como resultado del proceso anterior, calculamos el primer nodo raíz.



- Analicemos los el nodo izquierdo: Clima (y) Soleado

Clima	Temperatura	Humedad	Viento	CLASE
Soleado	Calor	Alta	Débil	NO
Soleado	Calor	Alta	Fuerte	NO
Soleado	Media	Alta	Débil	NO
Soleado	Fria	Normal	Débil	SI
Soleado	Media	Normal	Fuerte	SI

FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- **Segundo Nivel:**
  Como resultado del proceso anterior, calculamos el primer nodo raíz. El proceso comienza nuevamente con el primer paso del algoritmo

Clima	Temperatura	Humedad	Viento	CLASE
Soleado	Calor	Alta	Débil	NO
Soleado	Calor	Alta	Fuerte	NO
Soleado	Media	Alta	Débil	NO
Soleado	Fria	Normal	Débil	SI
Soleado	Media	Normal	Fuerte	SI

Temperatura	CLASE
Calor	NO
Calor	NO
Media	NO
Fria	SI
Media	SI

Viento	CLASE
Débil	NO
Fuerte	NO
Débil	NO
Débil	SI
Fuerte	SI

Humedad	CLASE
Alta	NO
Alta	NO
Alta	NO
Normal	SI
Normal	SI

TEMP	CLASE SI	CLASE NO
Calor	0	2
Media	1	1
Fria	1	0

VIENTO	CLASE SI	CLASE NO
Débil	2	1
Fuerte	1	1

HUMED	CLASE SI	CLASE NO
Alta	0	3
Normal	2	0

FACULTAD DE INGENIERÍA Y CIENCIAS
UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- **Segundo Nivel:**
  En este paso analizamos en detalle la Entropía de cada clase (*Segundo paso*)

TEMP	CLASE SI	CLASE NO
Calor	0	2
Media	1	1
Fria	1	0

$$H(\text{Temp}) = -\frac{2}{5}\times\log_2\left(\frac{2}{5}\right)-\frac{3}{5}\times\log_2\left(\frac{3}{5}\right)=0.971$$

VIENTO	CLASE SI	CLASE NO
Débil	2	1
Fuerte	1	1

$$H(\text{Viento}) = -\frac{3}{5}\times\log_2\left(\frac{3}{5}\right)-\frac{2}{5}\times\log_2\left(\frac{2}{5}\right)=0.971$$

HUMED	CLASE SI	CLASE NO
Alta	0	3
Normal	2	0

$$H(\text{Humedad}) = -\frac{2}{5}\times\log_2\left(\frac{2}{5}\right)-\frac{3}{5}\times\log_2\left(\frac{3}{5}\right)=0.971$$

- Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- **Segundo Nivel:**
  En este paso analizamos en detalle la Entropía de cada clase (*Segundo paso*)

TEMP	CLASE SI	CLASE NO
Calor	0	2
Media	1	1
Fria	1	0

$$H(\text{Temperatura, Calor}) = -\frac{0}{2}\times\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\times\log_2\left(\frac{2}{2}\right) = 0$$

$$H(\text{Temperatura, Media}) = -\frac{1}{2}\times\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\times\log_2\left(\frac{1}{2}\right) = 1$$

$$H(\text{Temperatura, Fria}) = -\frac{1}{1}\times\log_2\left(\frac{1}{1}\right) - \frac{0}{1}\times\log_2\left(\frac{0}{1}\right) = 0$$

VIENTO	CLASE SI	CLASE NO
Débil	2	1
Fuerte	1	1

$$H(\text{Viento, Débil}) = -\frac{2}{3}\times\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\times\log_2\left(\frac{1}{3}\right) = 0.918$$

$$H(\text{Viento, Fuerte}) = -\frac{1}{2}\times\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\times\log_2\left(\frac{1}{2}\right) = 1$$

HUMED	CLASE SI	CLASE NO
Alta	0	3
Normal	2	0

$$H(\text{Humedad, Alta}) = -\frac{0}{3}\times\log_2\left(\frac{0}{3}\right) - \frac{3}{3}\times\log_2\left(\frac{3}{3}\right) = 0$$

$$H(\text{Humedad, Normal}) = -\frac{2}{2}\times\log_2\left(\frac{2}{2}\right) - \frac{0}{2}\times\log_2\left(\frac{0}{2}\right) = 0$$

FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- **Segundo Nivel:**
  Seleccionamos como nodo raíz aquel atributo que tenga la mayor ganancia de información (*Tercer paso*)

$$H(\text{Temperatura}) = 0.971$$

registros

$H(\text{Temperatura, Calor}) = 0 \rightarrow 2/5$

$H(\text{Temperatura, Media}) = 1 \rightarrow 2/5$

$H(\text{Temperatura, Fria}) = 0 \rightarrow 1/5$

$$H(\text{Humedad}) = 0.971$$

$H(\text{Humedad, Alta}) = 0 \rightarrow 3/5$

$H(\text{Humedad, Normal}) = 0 \rightarrow 2/5$

$$H(\text{Viento}) = 0.971$$

registros

$H(\text{Viento, Débil}) = 0.981 \rightarrow 3/5$

$H(\text{Viento, Fuerte}) = 1 \rightarrow 2/5$

- Ganancia de Información

$$G(\text{Temperatura}) = 0.971 - \left(\frac{2}{5}\times 0 + \frac{2}{5}\times 1 + \frac{1}{5}\times 0\right) = 0.571$$

$$G(\text{Viento}) = 0.971 - \left(\frac{3}{5}\times 0.981 + \frac{2}{5}\times 1\right) = -0.0176$$

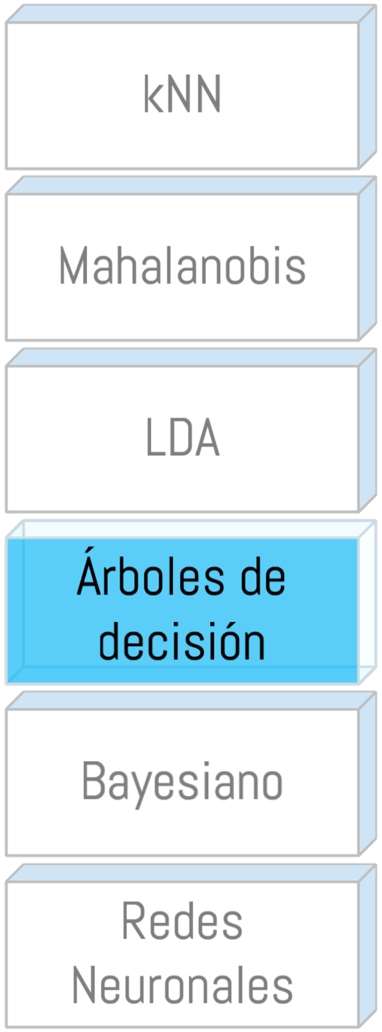
$$G(\text{Humedad}) = 0.971 - \left(\frac{3}{5}\times 0 + \frac{2}{5}\times 0\right) = 0.971$$

La mayor Ganancia de información corresponde al atributo Humedad. Por lo tanto este será el nodo raíz del nodo izquierdo

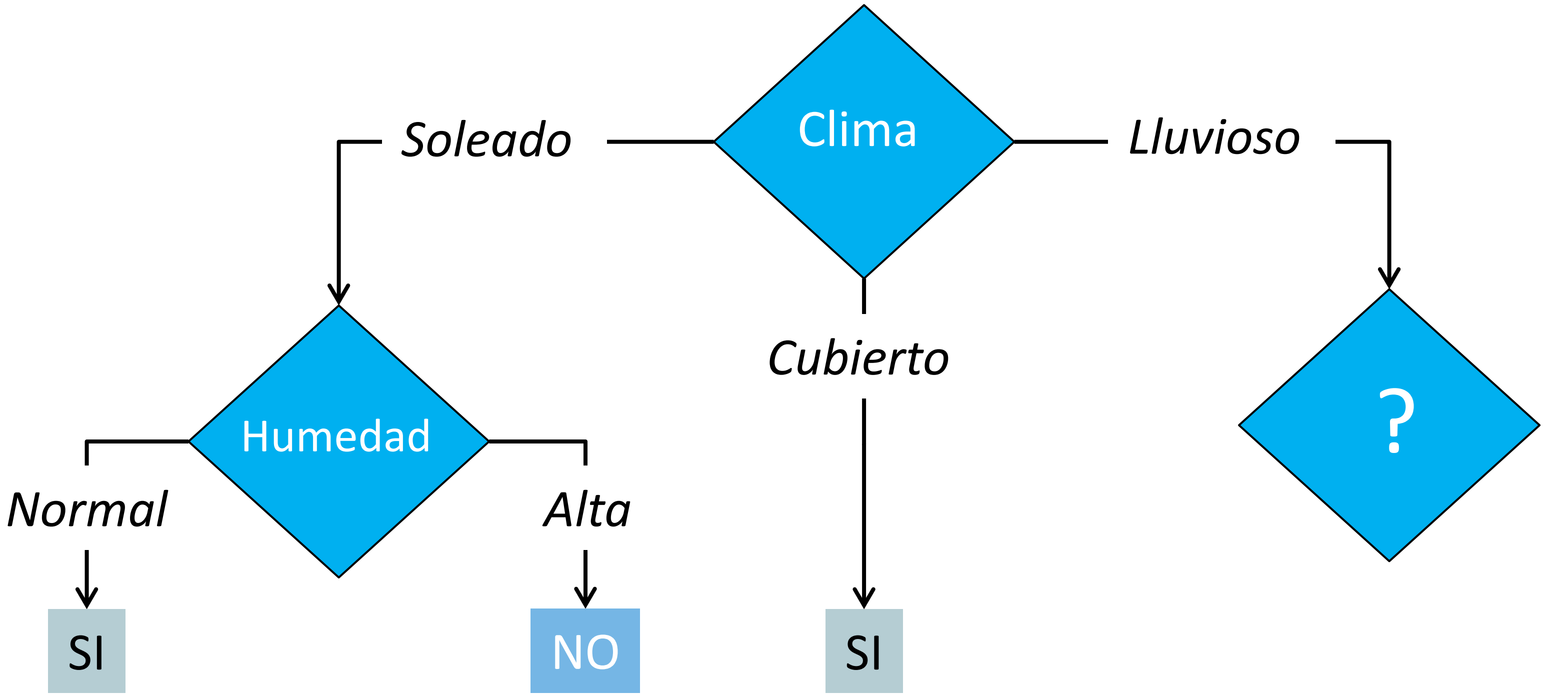
FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados



- **Segundo Nivel:**
Ahora vamos por el nodo izquierdo cuando Clima y Lluvioso.



Clima	Temperatura	Viento	CLASE
Lluvia	Media	Débil	SI
Lluvia	Fria	Débil	SI
Lluvia	Fria	Fuerte	NO
Lluvia	Media	Débil	SI
Lluvia	Media	Fuerte	NO

FACULTAD DE INGENIERÍA Y CIENCIAS

UAI UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

kNN

Mahalanobis

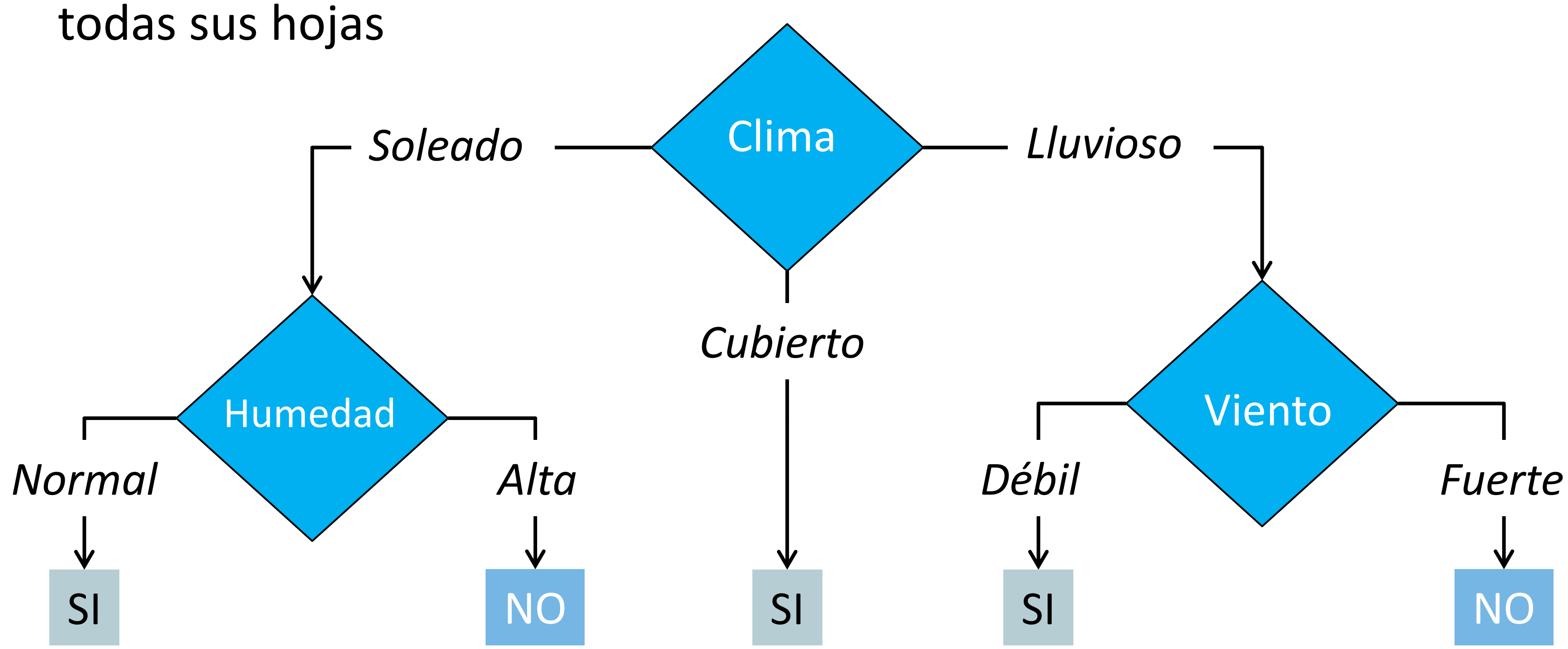
LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- Árbol de Decisión:
  El proceso de construcción finaliza cuando hemos completado el árbol, y todas sus hojas

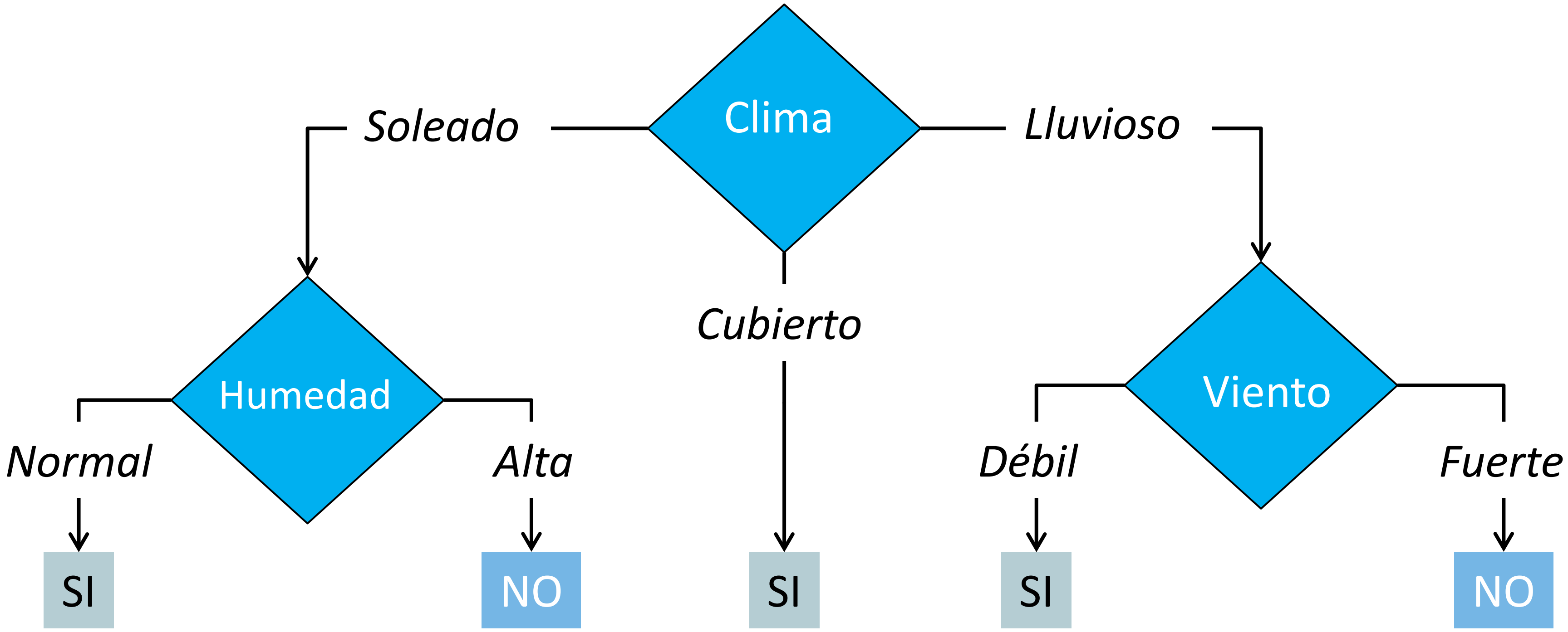


Finalmente obtenemos un árbol de clasificación. De esta forma si tenemos un nuevo registro, podemos navegar en él hasta obtener una clasificación

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- **Árbol de Decisión:**
Veamos el proceso de Clasificación. Simplemente consiste en recorrer el árbol hasta llegar a las hojas



Clima	Temperatura	Humedad	Viento	Clasificación
Lluvia	Fria	Normal	Fuerte	NO
Soleado	Media	Normal	Débil	SI
Cubierto	Calor	Alta	Fuerte	SI
Lluvia	Calor	Normal	Fuerte	NO
Lluvia	Media	Alta	Débil	SI

- Supervisados

- **Árbol de Decisión:**
  Algunos puntos importantes sobre los árboles de decisión se discuten a continuación:

  - Los árboles de decisión pueden trabajar con valores discretos o continuos. En este último caso el algoritmo debe separar por rangos los valores y luego aplicar

  - Una vez que se ha seleccionado un atributo, la búsqueda no tiene vuelta atrás (Backtracking). Una forma de evitar esto es construir un árbol completo y luego podarlo según alguna métrica (ver Anexo).

  - Siempre es preferible construir árboles cortos en vez de específicos. (principio Occam's razor: De un conjunto de hipótesis, debemos siempre preferir aquella que sea la más simple).

Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- **Árbol de Decisión:**
  Algunos puntos importantes sobre los árboles de decisión se discuten a continuación:

  - Si un atributo tiene muchos valores distintos, causará la generación de nuevas ramas en el árbol, lo cual afecta su capacidad de generalización. Para evitar esto, Quinlan presentó una mejora de su algoritmo a través dela *Razón de Ganancia* (Gain Ratio)

Ganancia

$$ratio(S,X) = \frac{G(S,X)}{\text{info}(S,X)}$$

Split Information

$$\text{info}(S,X) = -\sum_{i=1}^{c} \frac{|S_i|}{S} \log_2\left(\frac{|S_i|}{S}\right)$$

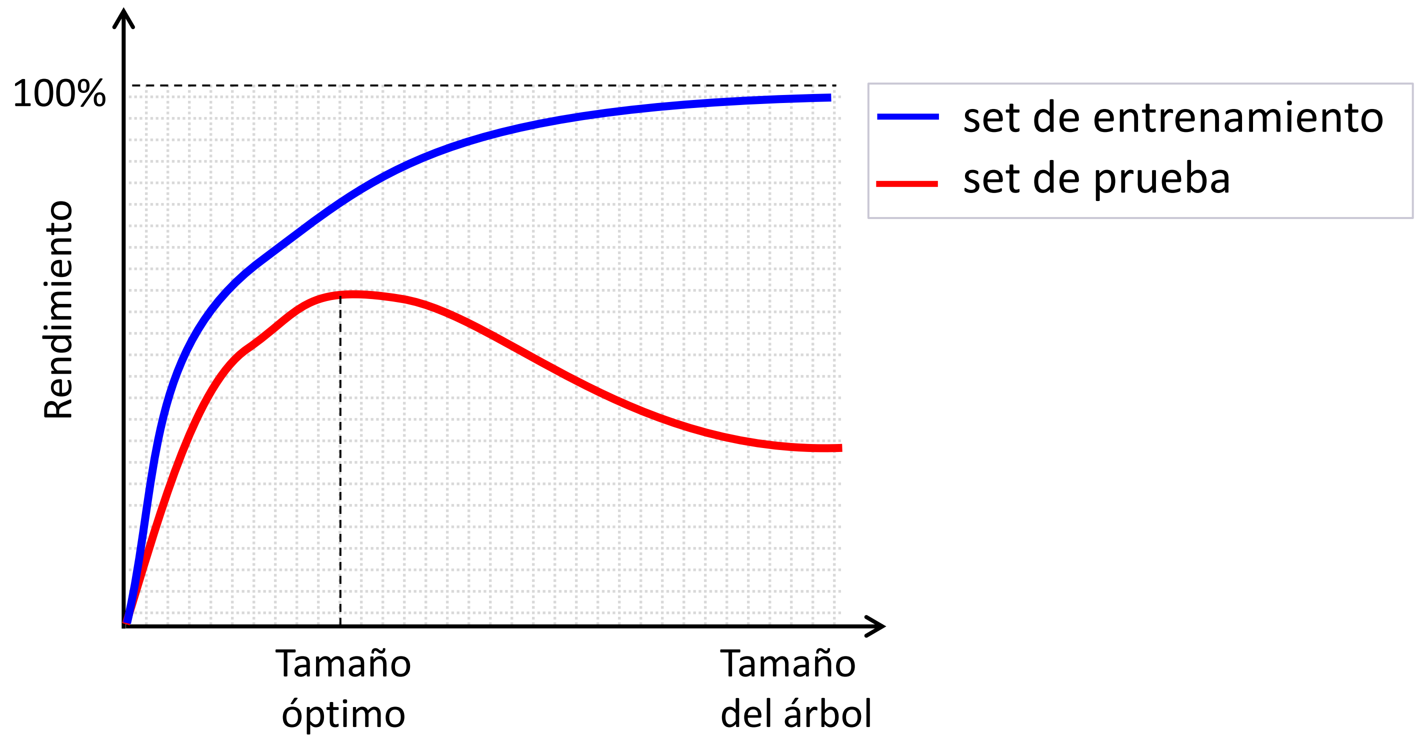
  - El algoritmo ID3 utiliza la *Ganancia de información* como método separar seleccionar un atributo nodo. En cambio el algoritmo C4.5 utiliza la *razón de ganancia.*

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- **Árbol de Decisión:**
  El mayor problema asociado a la construcción del árbol de decisión se conoce como *overfitting*.

- El *overfitting* sucede cuando el árbol comienza a ajustarse a los datos del set de entrenamiento. Esto implica que el árbol construya todas las ramas para el set de entrenamiento. El resultado de este sobreajuste implica la *pérdida de capacidad de generalización*, lo cual genera un bajo rendimiento.

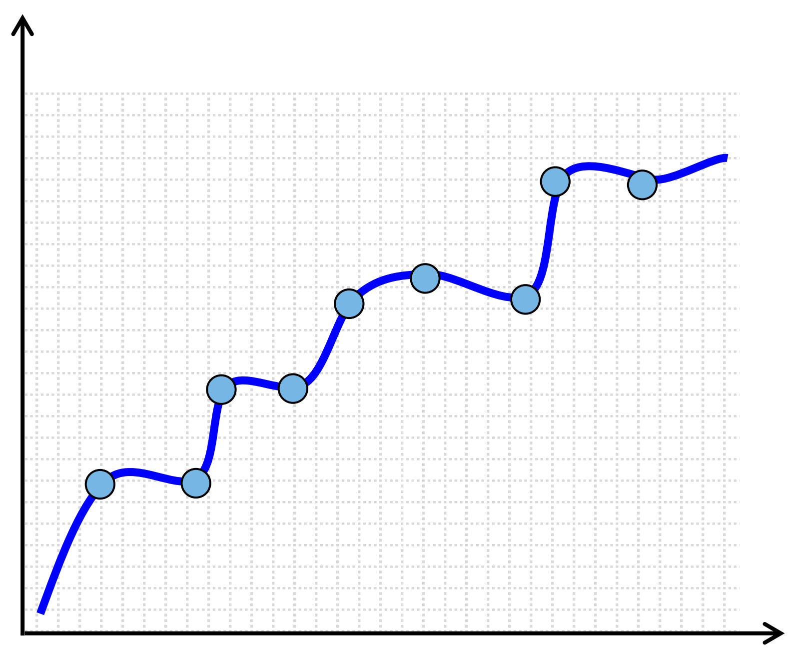


- Supervisados

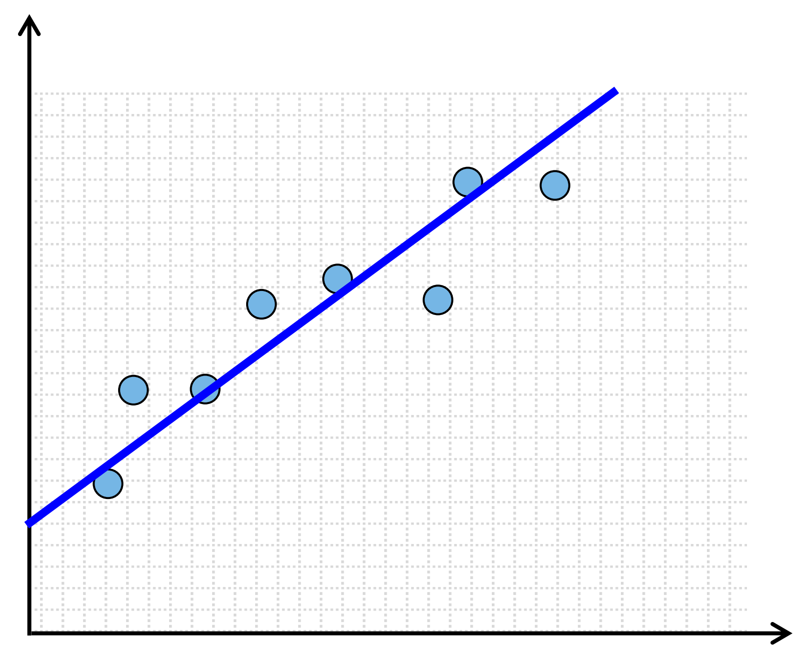
kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- Árbol de Decisión:
  El mayor problema asociado a la construcción del árbol de decisión se conoce como *overfitting*.

- El *overfitting* sucede cuando el árbol comienza a ajustarse a los datos del set de entrenamiento. Esto implica que el árbol construya todas las ramas para el set de entrenamiento. El resultado de este sobreajuste implica la *pérdida de capacidad de generalización*, lo cual genera un bajo rendimiento.



Unir los puntos



Interpolación lineal

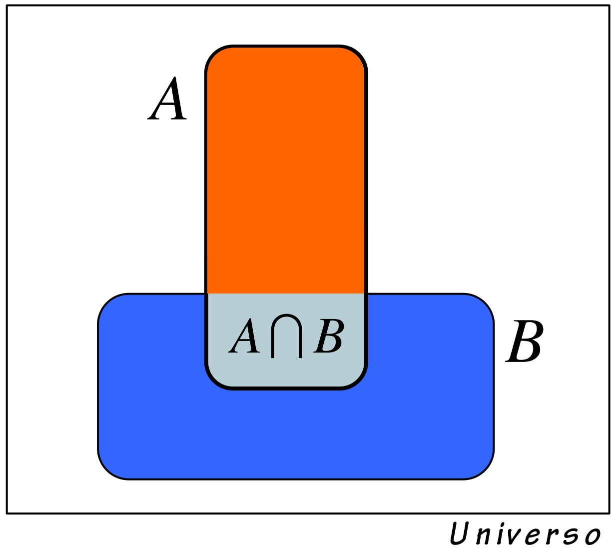
- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- **Clasificador Inexperto de Bayes:**
  El clasificador Inexperto de Bayes es uno de los algoritmos de clasificación más empleados y exitosos. Su principal característica es asumir que las características o atributos son independientes entre sí.

  Asumir independencia no siempre es correcto, sin embargo, este tipo de clasificador ha tenido un muy buen rendimiento en problemas reales, tales como clasificación de texto, diagnóstico de enfermedades, análisis de acciones, entre otros.

- **Preliminares**
  - Suponga que de un conjunto de personas, $P(A)$ es la probalidad que seas fanático de Pink Floyd y $P(B)$ es la probabilidad que seas fanático de la *música clásica*.
  - La intersección $P(A \cap B)$ es el conjunto de personas que le gusta la *música clásica* y *Pink Floyd*



Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- **Bayesiano**
- Redes Neuronales

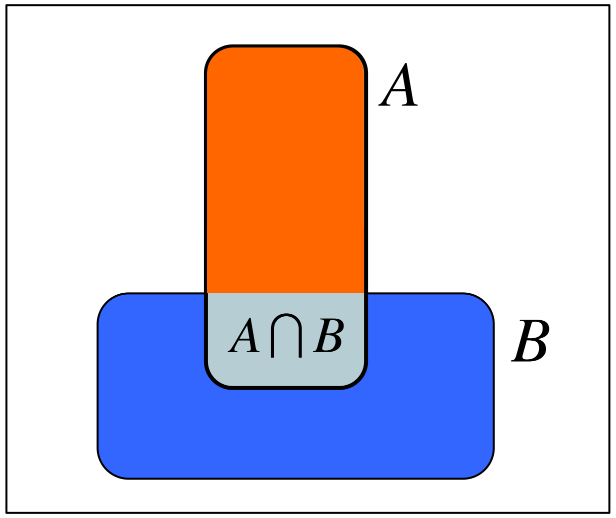
Probabilidad condicionada

- En concepto de probabilidad condicionada definido como :

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- Corresponde a la proporción entre (1) todas las personas les gusta la *música clásica* y *Pink Floyd* $P(A \cap B)$ y, (2) todas las personas que les gusta *la música clásica* $P(B)$

Universo



Regla de la cadena

$$P(A_1, A_2, \ldots, A_n) = P(A_1 \mid A_2, \ldots, A_n) \times \boxed{P(A_2, \ldots, A_n)}$$

$$\downarrow$$

$$\boxed{P(A_2 \mid A_3, \ldots, A_n) \times P(A_3, \ldots, A_n)}$$

$$\downarrow$$

$$\boxed{P(A_{n-1} \mid A_n) \times P(A_n)}$$

En conclusión

$$\boxed{P(A_1, A_2, \ldots, A_n) = P(A_1 \mid A_2, \ldots, A_n) \times \cdots \times P(A_{n-1} \mid A_n) \times P(A_n)}$$

- Despejando la ecuación anterior, podemos decir que:

$$P(A \cap B) = P(A, B) = P(A \mid B) \times P(B)$$

- También podríamos decir que:

$$P(B \mid A) = \frac{P(B, A)}{P(A)}$$

Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- **Bayesiano**
- Redes Neuronales

Probabilidad condicionada

- Reemplazando (1) en (2) y asumiendo que $P(A,B) = P(B,A)$ obtenemos el teorema de Bayes

(1) $P(A,B) = P(A \mid B) \times P(B)$

(2) $P(B \mid A) = \dfrac{P(B,A)}{P(A)}$

Teorema de Bayes

$$P(B \mid A) = \frac{P(A \mid B) \times P(B)}{P(A)}$$

Clasificador Inexperto de Bayes

$$P(clase_1 \mid a_1, a_2, \ldots, a_n) = ?$$

- Cuál es la probabilidad la instancia sea de la clase$_1$ dado el conjunto de n características o atributos $\{a_1, a_2, \ldots, a_n\}$

$$P(clase_1 \mid a_1, a_2, \ldots, a_n) = \frac{P(a_1, a_2, \ldots, a_n \mid clase_1) \times P(clase_1)}{P(a_1, a_2, \ldots, a_n)}$$

Simplemente aplicamos la fórmula de Bayes

Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

Clasificador Inexperto de Bayes

- En nuestro problemas tenemos más de una clase. Suponiendo que tenemos m clases entonces tendríamos que calcular:

$$P\left(clase_1 \mid a_1, a_2, \ldots, a_n\right) = \frac{P\left(a_1, a_2, \ldots, a_n \mid clase_1\right) \times P\left(clase_1\right)}{P\left(a_1, a_2, \ldots, a_n\right)}$$

$$\vdots$$

$$P\left(clase_j \mid a_1, a_2, \ldots, a_n\right) = \frac{P\left(a_1, a_2, \ldots, a_n \mid clase_j\right) \times P\left(clase_j\right)}{P\left(a_1, a_2, \ldots, a_n\right)}$$

$$\vdots$$

$$P\left(clase_m \mid a_1, a_2, \ldots, a_n\right) = \frac{P\left(a_1, a_2, \ldots, a_n \mid clase_m\right) \times P\left(clase_m\right)}{P\left(a_1, a_2, \ldots, a_n\right)}$$

De todas ellas debemos escoger la que nos entregue el mayor resultado. Sin embargo, si buscamos la que nos entregue el mayor valor, el denominador no está aportando información, ya que en todos los casos siempre tiene el mismo valor.

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- Clasificador Inexperto de Bayes
  - Por ello para facilitar el problema lo descartamos y ocupamos sólo la proporción $(\propto)$

$$P(clase_1 \mid a_1, a_2, \ldots, a_n) \propto P(a_1, a_2, \ldots, a_n \mid clase_1) \times P(clase_1)$$

$$P(clase_j \mid a_1, a_2, \ldots, a_n) \propto P(a_1, a_2, \ldots, a_n \mid clase_j) \times P(clase_j)$$

$$P(clase_m \mid a_1, a_2, \ldots, a_n) \propto P(a_1, a_2, \ldots, a_n \mid clase_m) \times P(clase_m)$$

  - Matemáticamente estamos buscando aquella clase que tenga la mayor probabilidad a posteriori. Lo anterior se expresa como:

$$clase_{MAP} = \underset{clase_k \in clases}{\operatorname{argmax}} P(a_1, a_2, \ldots, a_n \mid clase_k) \times P(clase_k)$$

Probabilidad de tener los atributos $a_1,\ldots,a_n$ dada la clase k

Probabilidad de tener la clase k. Esta probablidad es simplemente una frecuencia

Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

Clasificador Inexperto de Bayes

- Problema: Estimar la probabilidad $P(a_1,a_2,\ldots,a_n \mid clase_k)$ es extremadamente complejo o incluso imposible.

El supuesto del clasificador Inexperto de Bayes es suponer que los atributos son independientes. Esto significa que $P(a_1,a_2,\ldots,a_n \mid clase_m)$ se transforma en

1. $$P(a_1,a_2,\ldots,a_n \mid clase_k) \Rightarrow \prod_{i=1}^{n} P(a_i \mid clase_k)$$

2. $$clase_{MAP} = \underset{clase_k \in clases}{\operatorname{argmax}} P(a_1,a_2,\ldots,a_n \mid clase_k) \times P(clase_k)$$

- Finalmente remplazando (1) en (2) el clasificador inexperto de Bayes resulta

$$clase_{MAP} = \underset{clase_k \in clases}{\operatorname{argmax}} \prod_{i=1}^{n} P(a_i \mid clase_k) \times P(clase_k)$$

posterior = likelihood × prior

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- Clasificador Inexperto de Bayes
  - La función de puntaje es el *likelihood*, donde los parámetros de la función se define según la distribución de probabilidades las clases.

1 Las probabilidades *a priori* se calculan como:

$$P(C=i)=\frac{|C_i|}{n}$$

donde $|C_i|$ es el número de elementos que pertenecen a la clase i

2 Las probabilidades para valores discretos se calculan como:

$$P(X_j=k|C=i)=\frac{|X_{jki}|}{|C_i|}$$

donde $|X_{jki}|$ es el número de elementos del atributo *j* con valor k y clase i

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- Clasificador Inexperto de Bayes
  Analicemos el clasificador inexperto de Bayes sobre el siguiente conjunto de datos

Características o Atributos | Clase

Día	Clima	Temperatura	Humedad	Viento	¿Juego Tenis?
1	Soleado	Calor	Alta	Débil	NO
2	Soleado	Calor	Alta	Fuerte	NO
3	Cubierto	Calor	Alta	Débil	SI
4	Lluvia	Media	Alta	Débil	SI
5	Lluvia	Fria	Normal	Débil	SI
6	Lluvia	Fria	Normal	Fuerte	NO
7	Cubierto	Fria	Normal	Fuerte	SI
8	Soleado	Media	Alta	Débil	NO
9	Soleado	Fria	Normal	Débil	SI
10	Lluvia	Media	Normal	Débil	SI
11	Soleado	Media	Normal	Fuerte	SI
12	Cubierto	Media	Alta	Fuerte	SI
13	Cubierto	Calor	Normal	Débil	SI
14	Lluvia	Media	Alta	Fuerte	NO

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- PRIMERO: Estimemos las probabilidades condicionales y totales de cada evento.

Día	¿Juego Tenis?
1	NO
2	NO
3	SI
4	SI
5	SI
6	NO
7	SI
8	NO
9	SI
10	SI
11	SI
12	SI
13	SI
14	NO

Todas las probabilidades que estimadas a continuación están calculadas sobre el conjunto de entrenamiento:

$$P(JugarTenis = SI) = \frac{9}{14}$$

$$P(JugarTenis = NO) = \frac{5}{14}$$

Esta proporción se calcula según al frecuencia de NO sobre el total de instancias

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- PRIMERO: Estimemos las probabilidades condicionales y totales de cada evento.

Día	Clima	¿Juego Tenis?
1	Soleado	NO
2	Soleado	NO
3	Cubierto	SI
4	Lluvia	SI
5	Lluvia	SI
6	Lluvia	NO
7	Cubierto	SI
8	Soleado	NO
9	Soleado	SI
10	Lluvia	SI
11	Soleado	SI
12	Cubierto	SI
13	Cubierto	SI
14	Lluvia	NO

Todas las probabilidades que estimadas a continuación están calculadas sobre el conjunto de entrenamiento:

Es la probabilidad de Clima soleado y clase SI versus el total de clases SI

$$P\left(\text{Clima} = soleado \mid Clase = SI\right) = \frac{2}{9}$$

$$P\left(\text{Clima} = soleado \mid Clase = NO\right) = \frac{3}{5}$$

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- PRIMERO: Estimemos las probabilidades condicionales y totales de cada evento.

Día	Temperatura	¿Juego Tenis?
1	Calor	NO
2	Calor	NO
3	Calor	SI
4	Media	SI
5	Fria	SI
6	Fria	NO
7	Fria	SI
8	Media	NO
9	Fria	SI
10	Media	SI
11	Media	SI
12	Media	SI
13	Calor	SI
14	Media	NO

Todas las probabilidades que estimadas a continuación están calculadas sobre el conjunto de entrenamiento:

Es la probabilidad de Temp. media y Clases SI versus el total de clases SI

$$P\left(\text{Temp} = media \mid Clase = SI\right) = \frac{4}{9}$$

$$P\left(\text{Temp} = media \mid Clase = NO\right) = \frac{2}{5}$$

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- PRIMERO: Estimemos las probabilidades condicionales y totales de cada evento.

Día	Humedad	¿Juego Tenis?
1	Alta	NO
2	Alta	NO
3	Alta	SI
4	Alta	SI
5	Normal	SI
6	Normal	NO
7	Normal	SI
8	Alta	NO
9	Normal	SI
10	Normal	SI
11	Normal	SI
12	Alta	SI
13	Normal	SI
14	Alta	NO

Todas las probabilidades que estimadas a continuación están calculadas sobre el conjunto de entrenamiento:

Es la probabilidad de Humedad alta y clases SI versus el total de clases SI

$$P\left(\text{Humedad} = alta \mid Clase = SI\right) = \frac{3}{9}$$

$$P\left(\text{Humedad} = alta \mid Clase = NO\right) = \frac{4}{5}$$

FACULTAD DE INGENIERÍA Y CIENCIAS

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- PRIMERO: Estimemos las probabilidades condicionales y totales de cada evento.

Día	Viento	¿Juego Tenis?
1	Débil	NO
2	Fuerte	NO
3	Débil	SI
4	Débil	SI
5	Débil	SI
6	Fuerte	NO
7	Fuerte	SI
8	Débil	NO
9	Débil	SI
10	Débil	SI
11	Fuerte	SI
12	Fuerte	SI
13	Débil	SI
14	Fuerte	NO

Todas las probabilidades que estimadas a continuación están calculadas sobre el conjunto de entrenamiento:

Es la probabilidad de Viento débil y clase SI versus el total de clases SI

$$P\left(\text{Viento} = \textit{débil} \mid Clase = SI\right) = \frac{6}{9}$$

$$P\left(\text{Viento} = \textit{débil} \mid Clase = NO\right) = \frac{2}{5}$$

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- Clasificador Inexperto de Bayes

$$P(JugarTenis = SI) = \frac{9}{14} = 0.642 \qquad P(JugarTenis = NO) = \frac{5}{14} = 0.357$$

$$P(\text{Clima} = soleado \mid Clase = SI) = \frac{2}{9} \qquad P(\text{Clima} = soleado \mid Clase = NO) = \frac{3}{5}$$

$$P(\text{Temp} = media \mid Clase = SI) = \frac{4}{9} \qquad P(\text{Temp} = media \mid Clase = NO) = \frac{2}{5}$$

$$P(\text{Humedad} = alta \mid Clase = SI) = \frac{3}{9} \qquad P(\text{Humedad} = alta \mid Clase = NO) = \frac{4}{5}$$

$$P(\text{Viento} = débil \mid Clase = SI) = \frac{6}{9} \qquad P(\text{Viento} = débil \mid Clase = NO) = \frac{2}{5}$$

- ¿Cómo clasificamos la siguiente instancia?

$$\{\text{Clima=Soleado, Temperatura=Media, Humedad=Alta, Viento= débil}\}$$

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- Clasificador Inexperto de Bayes

$$clase_{MAP} = \underset{clase_k \in clases}{\operatorname{argmax}} \prod_{i=1}^{n} P(a_i \mid clase_k) \times P(clase_k)$$

Clima, Temperatura, Humedad, Viento

$$P(Soleado \mid clase_k) \times P(Media \mid clase_k) \times P(Alta \mid clase_k) \times P(Débil \mid clase_k) \times P(clase_k)$$

- Reemplazando los valores tenemos:

CLASE SI :

$$P(Soleado \mid SI) \times P(Media \mid SI) \times P(Alta \mid SI) \times P(Débil \mid SI) \times P(SI) = \frac{2}{9} \times \frac{4}{9} \times \frac{3}{9} \times \frac{6}{9} \times \frac{9}{14}$$

CLASE NO :

$$P(Soleado \mid NO) \times P(Media \mid NO) \times P(Alta \mid NO) \times P(Débil \mid NO) \times P(NO) = \frac{3}{5} \times \frac{2}{5} \times \frac{4}{5} \times \frac{2}{5} \times \frac{5}{14}$$

FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- Clasificador Inexperto de Bayes

$$clase_{MAP} = \underset{clase_k \in clases}{\operatorname{argmax}} \prod_{i=1}^{n} P(a_i \mid clase_k) \times P(clase_k)$$

Clima, Temperatura, Humedad, Viento

$$P(Soleado \mid clase_k) \times P(Media \mid clase_k) \times P(Alta \mid clase_k) \times P(Débil \mid clase_k) \times P(clase_k)$$

- Reemplazando los valores tenemos:

CLASE SI :

$$P(Soleado \mid SI) \times P(Media \mid SI) \times P(Alta \mid SI) \times P(Débil \mid SI) \times P(SI) = 0.0141093474$$

CLASE NO :

$$P(Soleado \mid NO) \times P(Media \mid NO) \times P(Alta \mid NO) \times P(Débil \mid NO) \times P(NO) = 0.0274285714$$

FACULTAD DE INGENIERÍA Y CIENCIAS

UAI UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- Clasificador Inexperto de Bayes

$$clase_{MAP} = \underset{clase_k \in clases}{\operatorname{argmax}} \prod_{i=1}^{n} P(a_i \mid clase_k) \times P(clase_k)$$

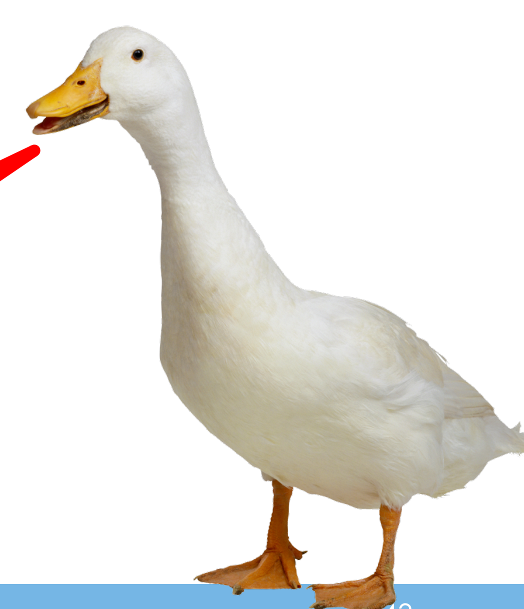
Clima, Temperatura, Humedad, Viento

$$P(Soleado \mid clase_k) \times P(Media \mid clase_k) \times P(Alta \mid clase_k) \times P(Débil \mid clase_k) \times P(clase_k)$$

- Asignamos la clase correspondiente al mayor valor.

$$clase_{MAP} = \operatorname{argmax}\{0.014109 \ , \ 0.02742\}$$

CLASE NO



- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- Probabilidades cero

Existen algunos casos donde la probabilidad de algún escenario sea cero, en esta tabla, ¿en qué caso sería?

Día	Clima	Temperatura	Humedad	Viento	¿Juego Tenis?
	Características o Atributos				*Clase*
1	Soleado	Calor	Alta	Débil	NO
2	Soleado	Calor	Alta	Fuerte	NO
3	Cubierto	Calor	Alta	Débil	NO
4	Lluvia	Media	Alta	Débil	SI
5	Lluvia	Fria	Normal	Débil	SI
6	Lluvia	Fria	Normal	Fuerte	NO
7	Cubierto	Fria	Normal	Fuerte	SI
8	Soleado	Media	Alta	Débil	NO
9	Soleado	Fria	Normal	Débil	NO
10	Lluvia	Media	Normal	Débil	SI
11	Soleado	Media	Normal	Fuerte	NO
12	Cubierto	Media	Alta	Fuerte	SI
13	Cubierto	Calor	Normal	Débil	SI
14	Lluvia	Media	Alta	Fuerte	NO

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- Estimemos las probabilidades condicionales y totales de cada evento.

Día	Clima	¿Juego Tenis?
1	Soleado	NO
2	Soleado	NO
3	Cubierto	NO
4	Lluvia	SI
5	Lluvia	SI
6	Lluvia	NO
7	Cubierto	SI
8	Soleado	NO
9	Soleado	NO
10	Lluvia	SI
11	Soleado	NO
12	Cubierto	SI
13	Cubierto	SI
14	Lluvia	NO

Todas las probabilidades que estimadas a continuación están calculadas sobre el conjunto de entrenamiento:

Es la probabilidad de Clima soleado y clase SI versus el total de clases SI

$$P(\textsc{Clima} = Soleado | Clase = \mathrm{SI}) = \frac{0}{6}$$

$$P(\textsc{Clima} = Soleado | Clase = \mathrm{NO}) = \frac{5}{8}$$

Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- **Bayesiano**
- Redes Neuronales

- Para evitar las probabilidades con valor cero, suavizamos la estimación de probabilidades agregamos valores en el numerador y el denominador

Día	Clima	¿Juego Tenis?
1	Soleado	NO
2	Soleado	NO
3	Cubierto	NO
4	Lluvia	SI
5	Lluvia	SI
6	Lluvia	NO
7	Cubierto	SI
8	Soleado	NO
9	Soleado	NO
10	Lluvia	SI
11	Soleado	NO
12	Cubierto	SI
13	Cubierto	SI
14	Lluvia	NO

La corrección de Laplace agrega 1 en el numerador y el número $|C|$ de posibles valores en el denominador

$$P(X_j = k | C = i) = \frac{|X_{jki}| + 1}{|C_i| + |C|}$$

sin Laplace

C	SI	NO
Soleado	0/6	5/8
Cubierto	3/6	1/8
Lluvia	3/6	2/8
∑	1	1

con Laplace

C	SI	NO
Soleado	1/9	6/11
Cubierto	4/9	2/11
Lluvia	4/9	3/11
∑	1	1

$|C|$==3

- Normalmente empleamos variables categóricas para construir las tablas de frecuencias. Pero existe otra opción a través del cálculo de Funciones de Densidad de Probabilidad (PDF)

PASO 1: Calcular la media y desviación estándar para cada una de las clases.

Día	Humedad	¿Juego Tenis?
1	85	NO
2	90	NO
3	86	SI
4	96	SI
5	80	SI
6	65	SI
7	70	SI
8	70	NO
9	95	NO
10	80	SI
11	91	NO
12	70	SI
13	90	SI
14	75	SI

SI

Día	Humedad
3	86
4	96
5	80
6	65
7	70
10	80
12	70
13	90
14	75
Mean	79.1
Stdev	10.2

NO

Día	Humedad
1	85
2	90
8	70
9	95
11	91
Mean	86.2
Stdev	9.7

- Normalmente empleamos variables categóricas para construir las tablas de frecuencias. Pero existe otra opción a través del cálculo de Funciones de Densidad de Probabilidad (PDF)

PASO 2: Calcular la Función de Densidad de Probabilidad (PDF) para cada una de las clases. Asumimos normalidad en los datos

distribución normal

$$f(x) = \frac{1}{\sqrt{(2\pi)}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

	YES	NO
Mean	79.1	86.2
Stdev	10.2	9.7

Estimación PDF

$$P(humedad = 74 | play = \text{YES}) = \frac{1}{\sqrt{(2\pi)}(10.2)} \cdot e^{-\frac{(74-79.1)^2}{2*(10.2)^2}} = 0.0345$$

$$P(humedad = 74 | play = \text{NO}) = \frac{1}{\sqrt{(2\pi)}(9.7)} \cdot e^{-\frac{(74-86.2)^2}{2*(9.7)^2}} = 0.01865$$

Naive Bayes

Características

Fortalezas

- Fácil de implementar
- A menudo tiene un buen rendimiento a pesar que las variables pueden no ser independientes
- Puede aprender en forma incremental
- Valores faltantes son ignorados en el proceso de aprendizaje
- Modelo robusto frente a datos atípicos y/o irrelevantes

Debilidades

- Asimir clases condicionadas produce probabilidades sesgadas
- Dependencias entre las variables no pueden ser modeladas

Ejercicios

FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

Naive Bayes

El siguiente dataset muestra los siguientes atributos de autos: color tipo, origen, y la probabilidad de ser robado (baja, medio, alta). Aplique el clasificador Naive Bayes sobre la última columna:

1. Calcule las distribuciones a priori
2. Calcule las probabilidades condicionales.
3. Como clasifica el siguiente caso: Auto Rojo, SUV, Doméstico, de ser robado. Calcule todas las probabilidades
4. Cómo puede mejorar la predicción del punto 3. Explique y recalcule.

Color	Type	Origin	stolen odds
Red	Sports	Imported	low
Yellow	Sports	Imported	medium
Yellow	Sports	Domestic	medium
Yellow	Sports	Domestic	high
Yellow	Sports	Imported	high
Red	SUV	Imported	low
Yellow	SUV	Imported	low
Yellow	Sports	Domestic	medium
Red	SUV	Imported	medium
Red	Sports	Imported	high

- Naive Bayes

El siguiente dataset muestra los siguientes atributos de autos: color tipo, origen, y la probabilidad de ser robado (baja, medio, alta). Aplique el clasificador Naive Bayes sobre la última columna:

Color	Type	Origin	stolen odds
Red	Sports	Imported	low
Yellow	Sports	Imported	medium
Yellow	Sports	Domestic	medium
Yellow	Sports	Domestic	high
Yellow	Sports	Imported	high
Red	SUV	Imported	low
Yellow	SUV	Imported	low
Yellow	Sports	Domestic	medium
Red	SUV	Imported	medium
Red	Sports	Imported	high

1. Calcule las distribuciones a priori

Prob.	Low	Medium	High
Prior			

2. Calcule las probabilidades condicionales.

Color	Low	Medium	High
Red			
Yellow			

Origin	Low	Medium	High
Domestic			
Imported			

Type	Low	Medium	High
Sports			
SUV			

Naive Bayes

El siguiente dataset muestra los siguientes atributos de autos: color tipo, origen, y la probabilidad de ser robado (baja, medio, alta). Aplique el clasificador Naive Bayes sobre la última columna:

Color	Type	Origin	stolen odds
Red	Sports	Imported	low
Yellow	Sports	Imported	medium
Yellow	Sports	Domestic	medium
Yellow	Sports	Domestic	high
Yellow	Sports	Imported	high
Red	SUV	Imported	low
Yellow	SUV	Imported	low
Yellow	Sports	Domestic	medium
Red	SUV	Imported	medium
Red	Sports	Imported	high

1. Calcule las distribuciones a priori

Prob.	Low	Medium	High
Prior	3/10	4/10	3/10

2. Calcule las probabilidades condicionales.

Color	Low	Medium	High
Red	2/3	1/4	1/3
Yellow	1/3	3/4	2/3

Origin	Low	Medium	High
Domestic	0/3	2/4	1/3
Imported	3/3	2/4	2/3

Type	Low	Medium	High
Sports	1/3	3/4	3/3
SUV	2/3	1/4	0/3

- Naive Bayes

3. Como clasifica el siguiente caso: Auto Rojo, SUV, Doméstico, de ser robado. Calcule todas las probabilidades

$$P(low|red, SUV, Domestic) = 2/3 \cdot 2/3 \cdot 0/3 \cdot 3/10 = 0$$
$$P(medium|red, SUV, Domestic) = 1/4 \cdot 1/4 \cdot 2/4 \cdot 4/10 = 1/80$$
$$P(high|red, SUV, Domestic) = 1/3 \cdot 0/3 \cdot 1/3 \cdot 3/10 = 0$$

R. Existe una probabilidad media de ser robado ya que es aquella que maximiza la verosimilitud

Color	Low	Medium	High
Red	2/3	1/4	1/3
Yellow	1/3	3/4	2/3

Origin	Low	Medium	High
Domestic	0/3	2/4	1/3
Imported	3/3	2/4	2/3

Type	Low	Medium	High
Sports	1/3	3/4	3/3
SUV	2/3	1/4	0/3

Prob.	Low	Medium	High
Prior	3/10	4/10	3/10

Ejercicios

- Naive Bayes

4. Cómo puede mejorar la predicción del punto 3. Explique y recalcule.

   R. Vamos a emplear la corrección de Laplace y recalcular

Prob.	Low	Medium	High
Prior	3/10	4/10	3/10

sin Laplace

Color	Low	Medium	High
Red	2/3	1/4	1/3
Yellow	1/3	3/4	2/3

Origin	Low	Medium	High
Domestic	0/3	2/4	1/3
Imported	3/3	2/4	2/3

Type	Low	Medium	High
Sports	1/3	3/4	3/3
SUV	2/3	1/4	0/3

con Laplace

Color	Low	Medium	High
Red	2/3	1/4	1/3
Yellow	1/3	3/4	2/3

Origin	Low	Medium	High
Domestic	1/5	3/6	2/5
Imported	4/5	3/6	3/5

Type	Low	Medium	High
Sports	2/5	4/6	4/5
SUV	3/5	2/6	1/5

- Naive Bayes

4. Cómo puede mejorar la predicción del punto 3. Explique y recalcule.

R. Vamos a emplear la corrección de Laplace y recalcular

Como clasifica el siguiente caso: Auto Rojo, SUV, Doméstico, de ser robado.

Calcule todas las probabilidades

$$P(low|red, SUV, Domestic) = 2/3 \cdot 3/5 \cdot 1/5 \cdot 3/10 = 0.0240$$
$$P(medium|red, SUV, Domestic) = 1/4 \cdot 2/6 \cdot 3/6 \cdot 4/10 = 0.0167$$
$$P(high|red, SUV, Domestic) = 1/3 \cdot 1/5 \cdot 2/5 \cdot 3/10 = 0.0080$$

con Laplace

Color	Low	Medium	High
Red	2/3	1/4	1/3
Yellow	1/3	3/4	2/3

Origin	Low	Medium	High
Domestic	1/5	3/6	2/5
Imported	4/5	3/6	3/5

Type	Low	Medium	High
Sports	2/5	4/6	4/5
SUV	3/5	2/6	1/5

Priors

Prob.	Low	Medium	High
Prior	3/10	4/10	3/10

R. Existe una baja probabilidad de ser robado ya que es aquella que maximiza la verosimilitud (likelihood)

Anexo

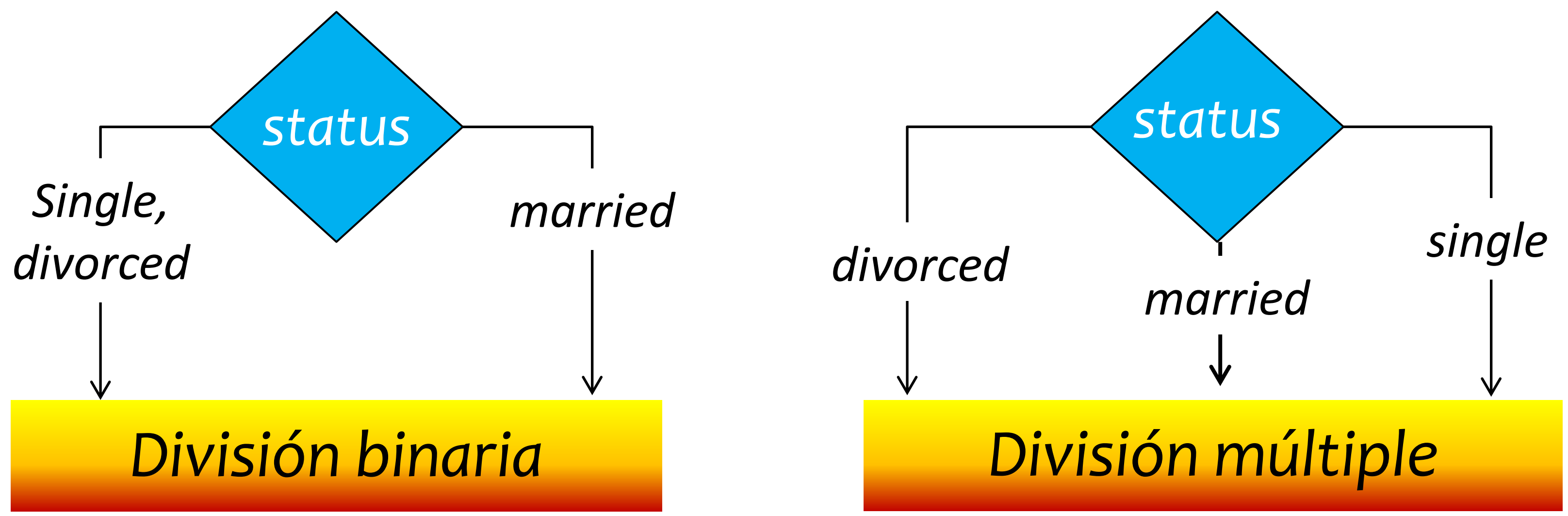
Árboles de Decisión (más información)

Temario

- Árboles de decisión
  - Modelos
  - Aprendizaje

- **Árbol de Decisión:**
  Algunos puntos importantes sobre los árboles de decisión se discuten a continuación:

  - Existen dos formas de dividir un nodo



  - Tipos de datos que podemos emplear

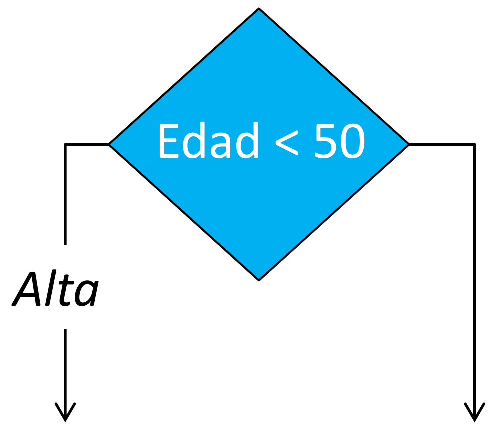
Nominal	Ordinal	Contínuos
Solo nombres, etiquetas, sin orden	Ordenado, diferencias sin sentido entre unidades	Corresponde al tipo intervalo o ratio.

debemos discretizarlo

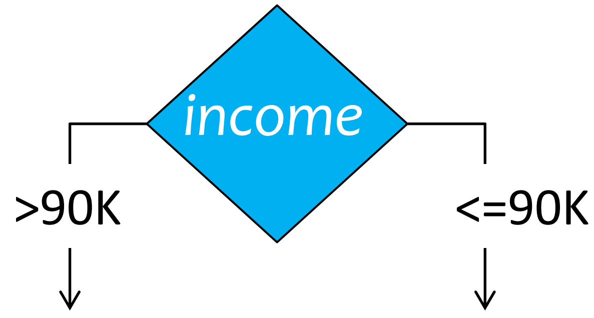
FACULTAD DE INGENIERÍA Y CIENCIAS
UNIVERSIDAD ADOLFO IBÁÑEZ

- Árbol de Decisión:
  Algunos puntos importantes sobre los árboles de decisión se discuten a continuación:

  - Decisión binaria: El regla de decisión divide con una condición lógica. Puede ser complejo de calcular computacionalmente ya que debe considerar todas las posibles divisiones para encontrar la óptima



  - Decisión de discretización: Convierte los datos en ordinales categóricos. Se puede realizar en dos modos:
    - Estática: discretiza al comienzo los rangos (edad <=20, 20<edad<30, 30<=edad
    - Dinámica: los rangos pueden ser de intervalos iguales, misma frecuencia de grupos (percentiles), o basado en clustering



FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

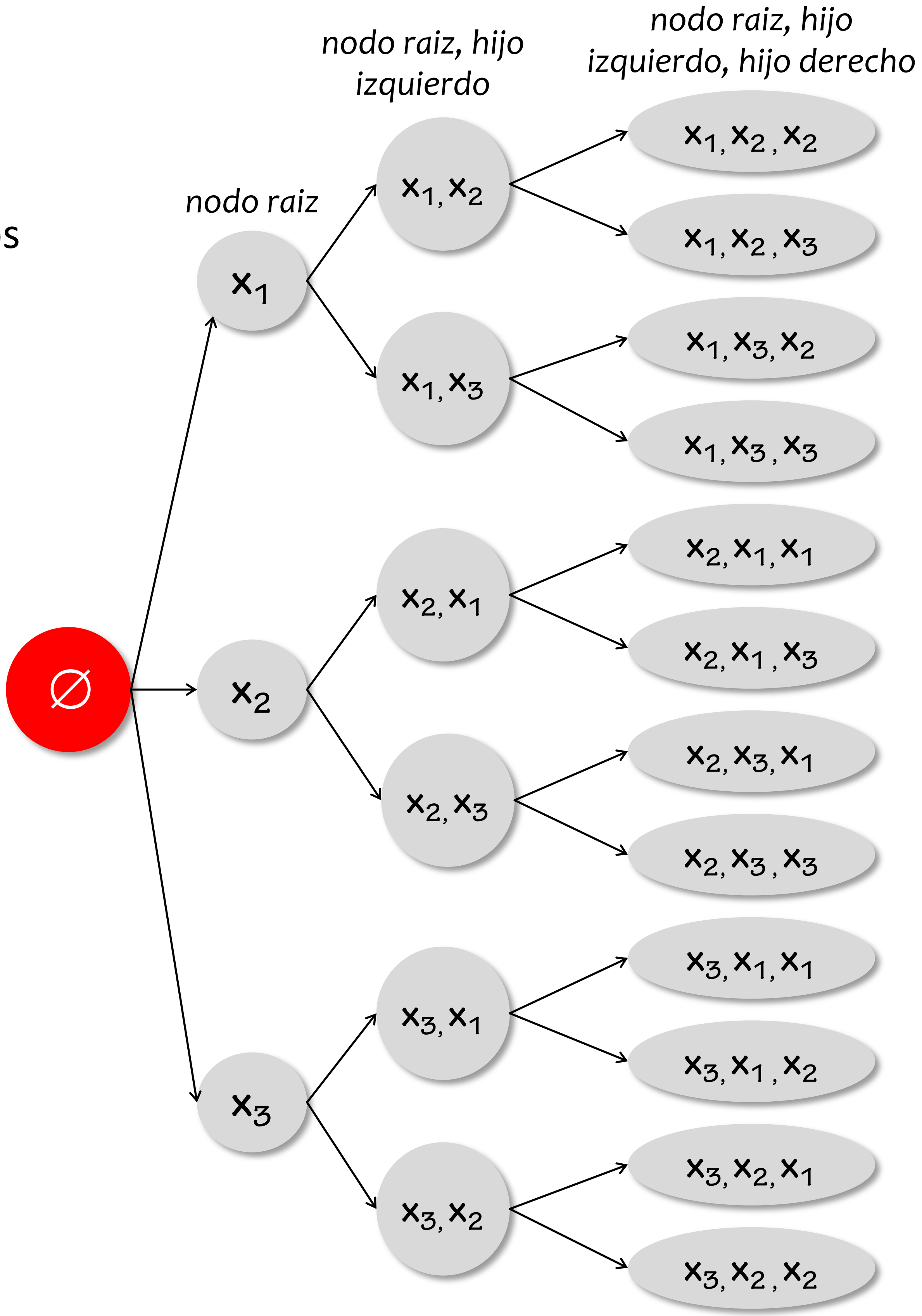
- **Árbol de Decisión:**
  Considere un espacio de M posibles modelos de una estructura de árbol

$$M = \{M_1, M_2, \ldots, M_k\}$$

¿Cuál de todos los modelos minimiza el error del conjunto de entrenamiento?

- **Búsqueda exhaustiva:**
  Consiste en generar todos los posibles modelos y seleccionar el mejor de todos ellos.

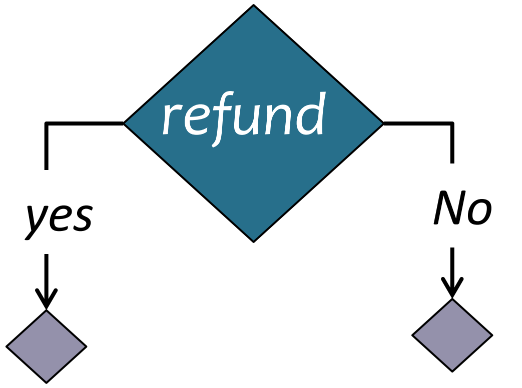
Lamentablemente hay un número exponencial de modelos en el espacio discreto, lo cual significa que sea impracticable o intratable. Sin embargo, si pudiéramos generarlo, garantizaríamos encontrar el mejor modelo entre todos



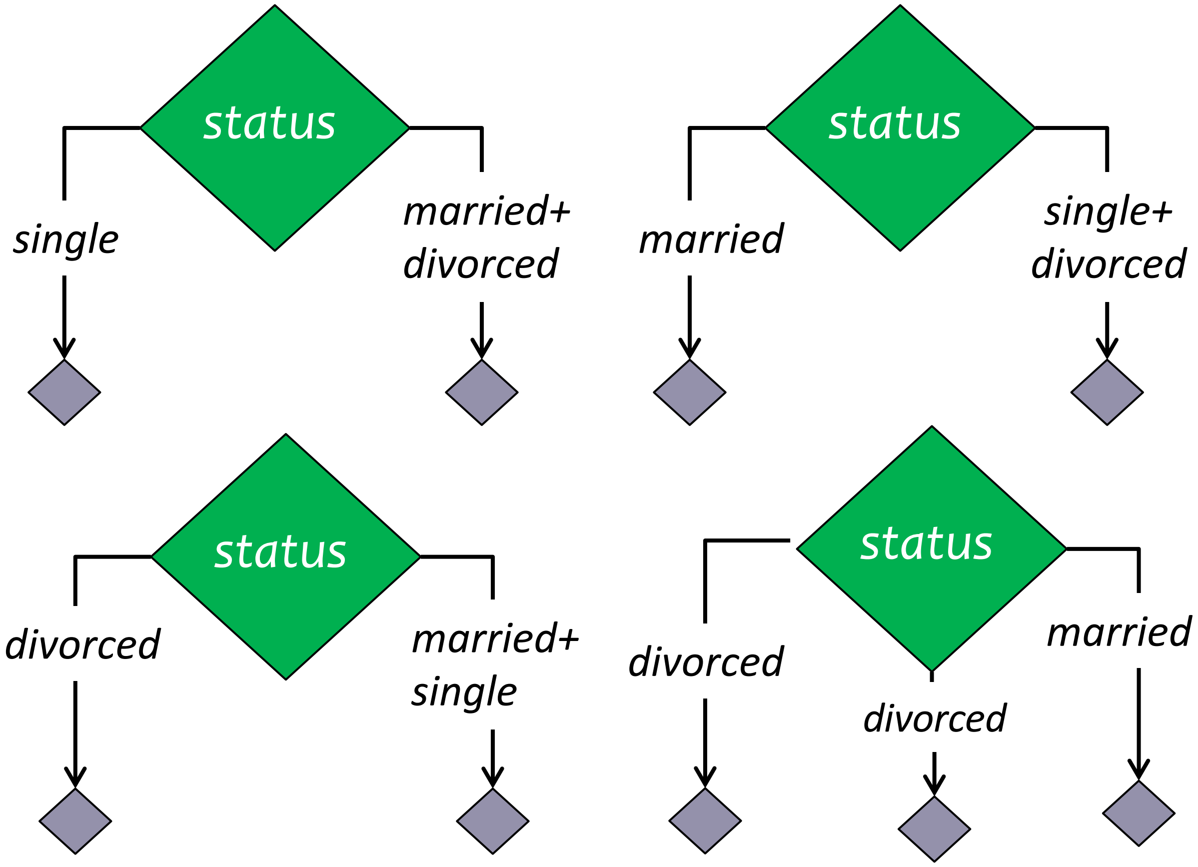
- Configuraciones

  Calculemos las siguientes opciones de configuración para la siguiente tabla:

  - refund: 1 separación binaria
  - status: posee 4 posibles separaciones



refund	status	income	affair
yes	single	125K	no
no	married	100K	no
no	single	70K	no
yes	married	120K	no
no	divorced	95K	yes
no	married	60K	no
yes	divorced	220K	no
no	single	85K	yes
no	married	75K	no
no	single	90K	yes



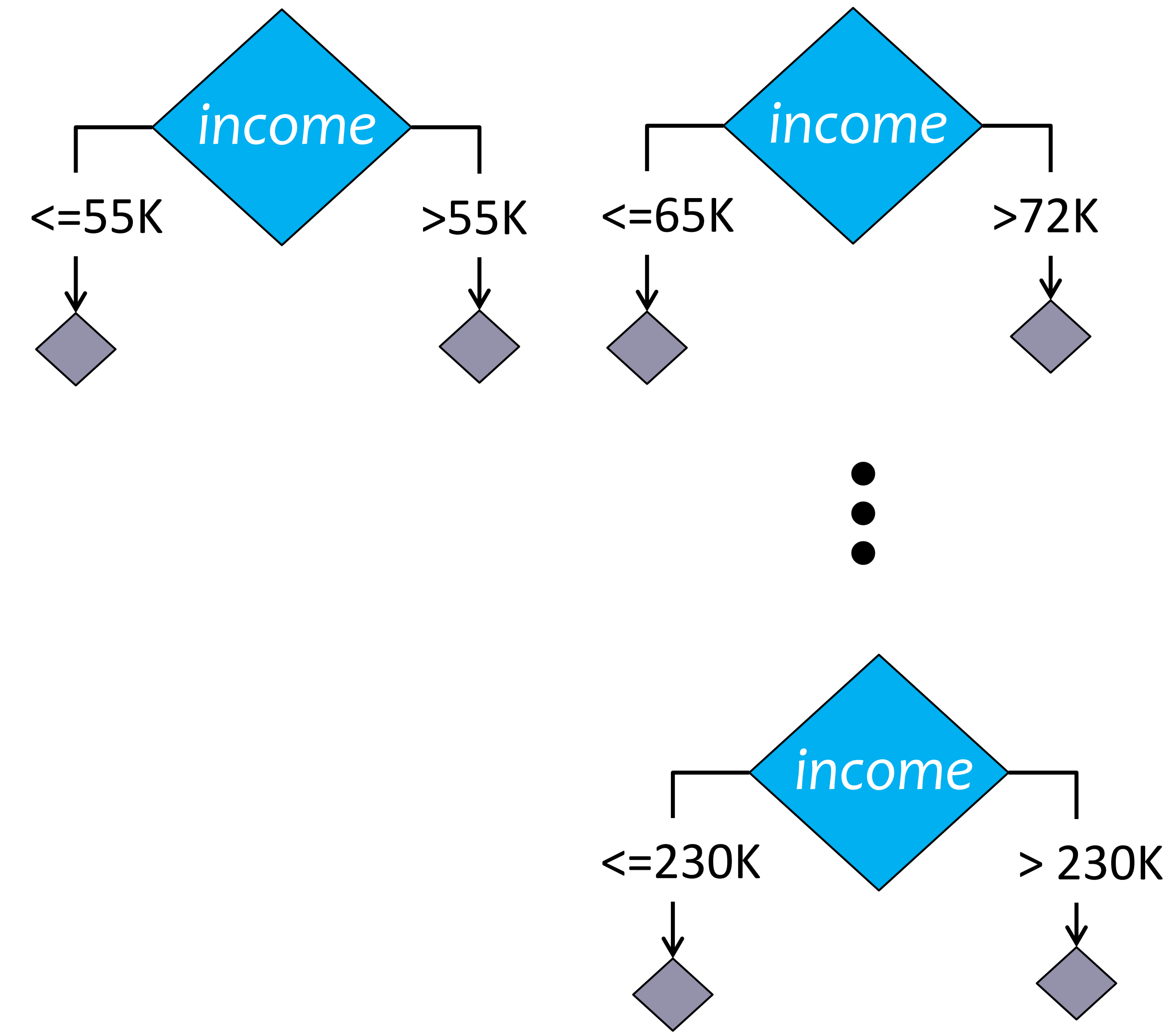
- **Configuraciones**
  Calculemos las siguientes opciones de configuración para la siguiente tabla:

  - income: posee 11 posibles separaciones

income
125K
100K
70K
120K
95K
60K
220K
85K
75K
90K

ordenamos intervalos
60K
70K
75K
85K
90K
95K
100K
120K
125K
220K

nuevos rangos	
55K	1
65K	2
72K	3
80K	4
87K	5
92K	6
97K	7
110K	8
122K	9
172K	10
230	11



Clasificadores
Generación de árboles

- **Peor caso**

  En el peor escenario, tenemos 44 posibles árboles SOLO para el siguiente orden de las variables (refund > status > income)

- refund: 1 separación binaria
- status: posee 4 posibles separaciones
- income: posee 11 posibles separaciones

$$1 * 4 * 11 = 44$$

refund	status	income	affair
yes	single	125K	no
no	married	100K	no
no	single	70K	no
yes	married	120K	no
no	divorced	95K	yes
no	married	60K	no
yes	divorced	220K	no
no	single	85K	yes
no	married	75K	no
no	single	90K	yes

Dado que tenemos tres variables, tenemos en total 3! configuraciones distintas, es decir, 6 posibes configuraciones

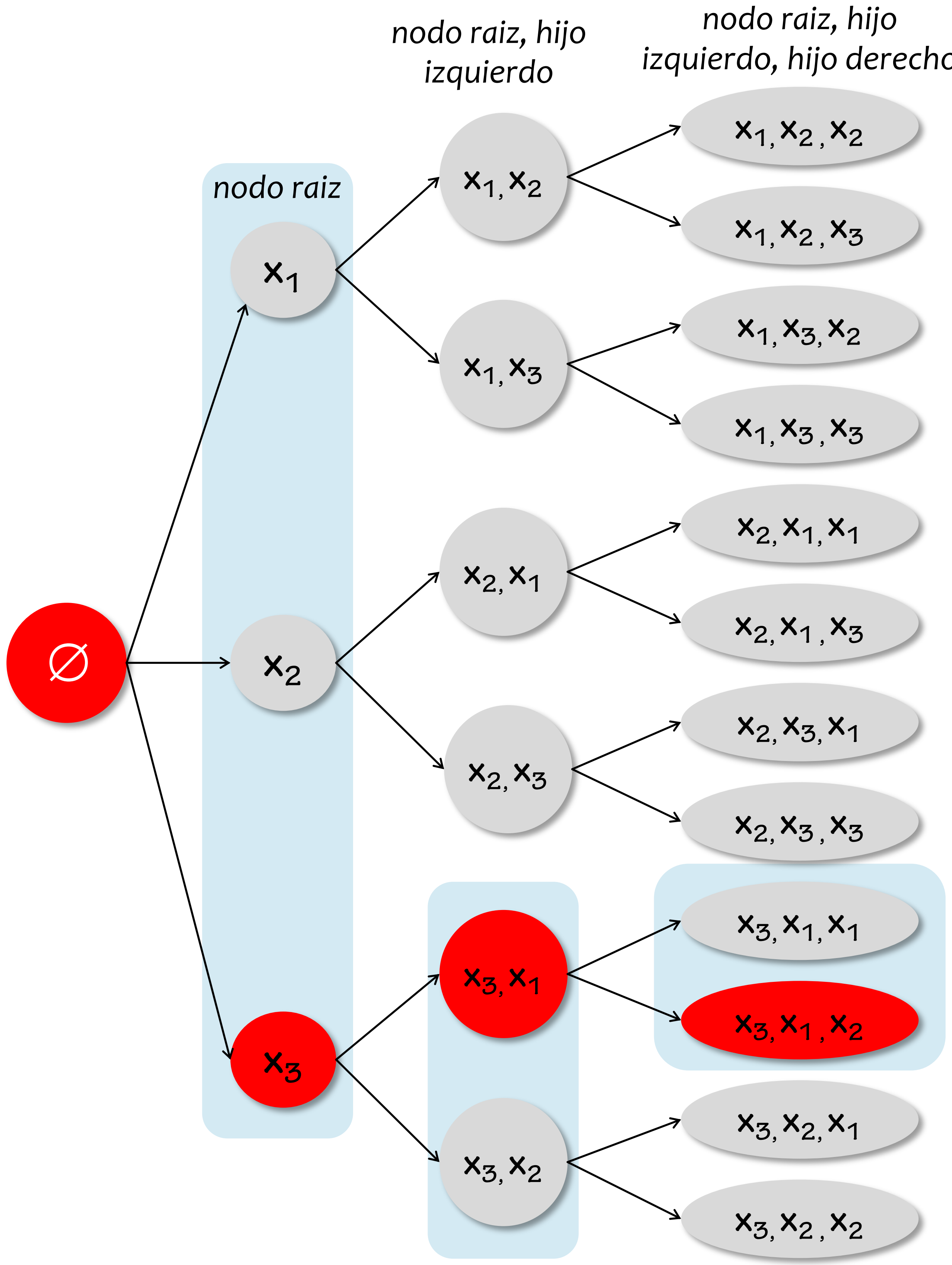
$$6 * 44 = 264$$

posibles árboles

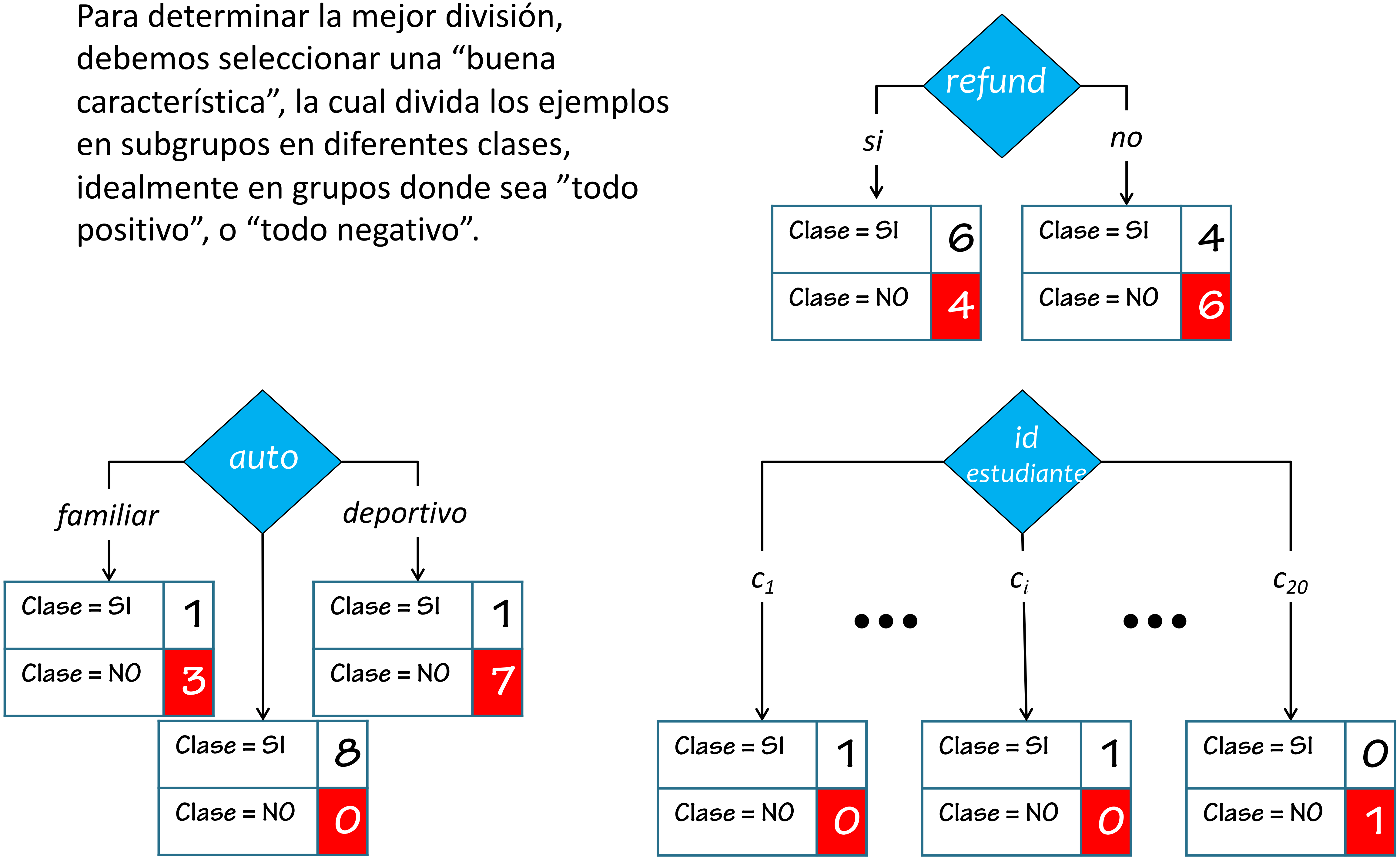
- **Búsqueda basada en heurísticas:**
  En cada paso de división del árbol evalúa las alternativas directas basada en la información disponible y toma una decisión.

La búsqueda basada en heurísticas NO realiza una búsqueda exhaustiva del espacio de solución

- El modelo seleccionado es un óptimo local.



- **Puntaje**
  Para determinar la mejor división, debemos seleccionar una "buena característica", la cual divida los ejemplos en subgrupos en diferentes clases, idealmente en grupos donde sea "todo positivo", o "todo negativo".



Clasificadores
Puntaje

FACULTAD DE INGENIERÍA Y CIENCIAS

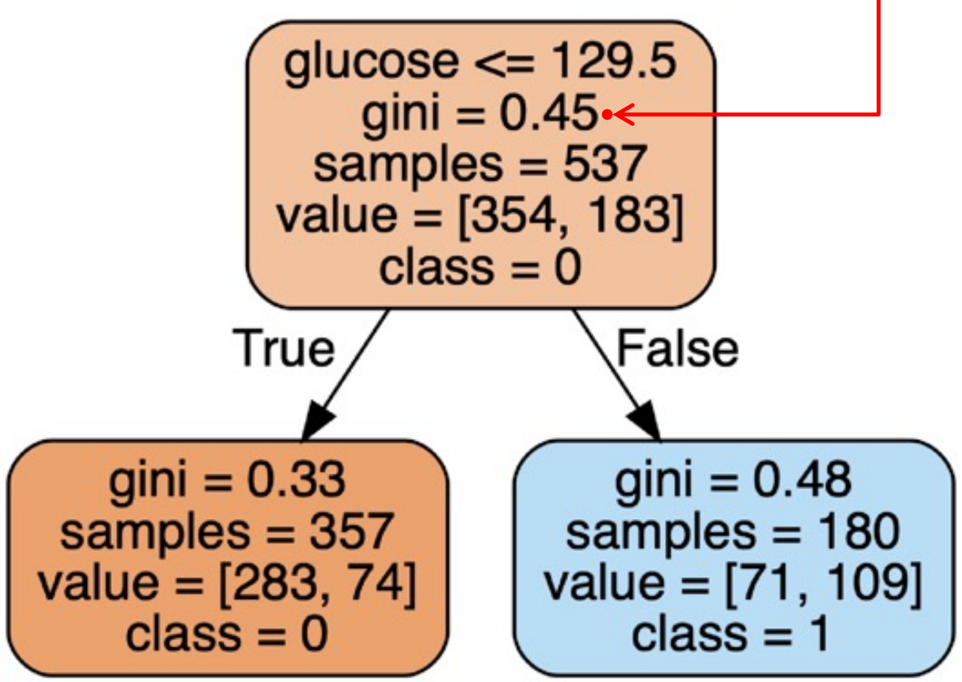
UNIVERSIDAD ADOLFO IBÁÑEZ

Gini

Gini impurity

$$Gini(X) = 1 - \sum_x p(x)^2$$

$$1 - \left(\frac{354}{537}\right) - \left(\frac{183}{537}\right) = 0.45$$

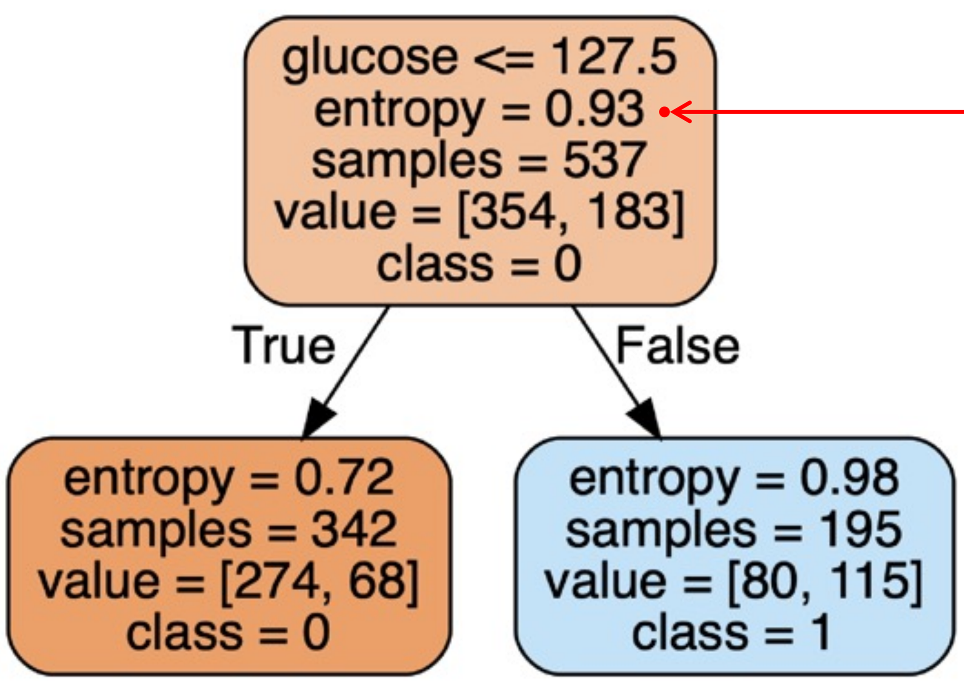


Entropía

Ganancia de Información

$$H(X) = -\sum_x p(x) \log_2 p(x)$$

$$-\left(\frac{354}{537}\right) \log_2 \left(\frac{354}{537}\right) - \left(\frac{183}{537}\right) log_2 \left(\frac{183}{537}\right) = 0.93$$



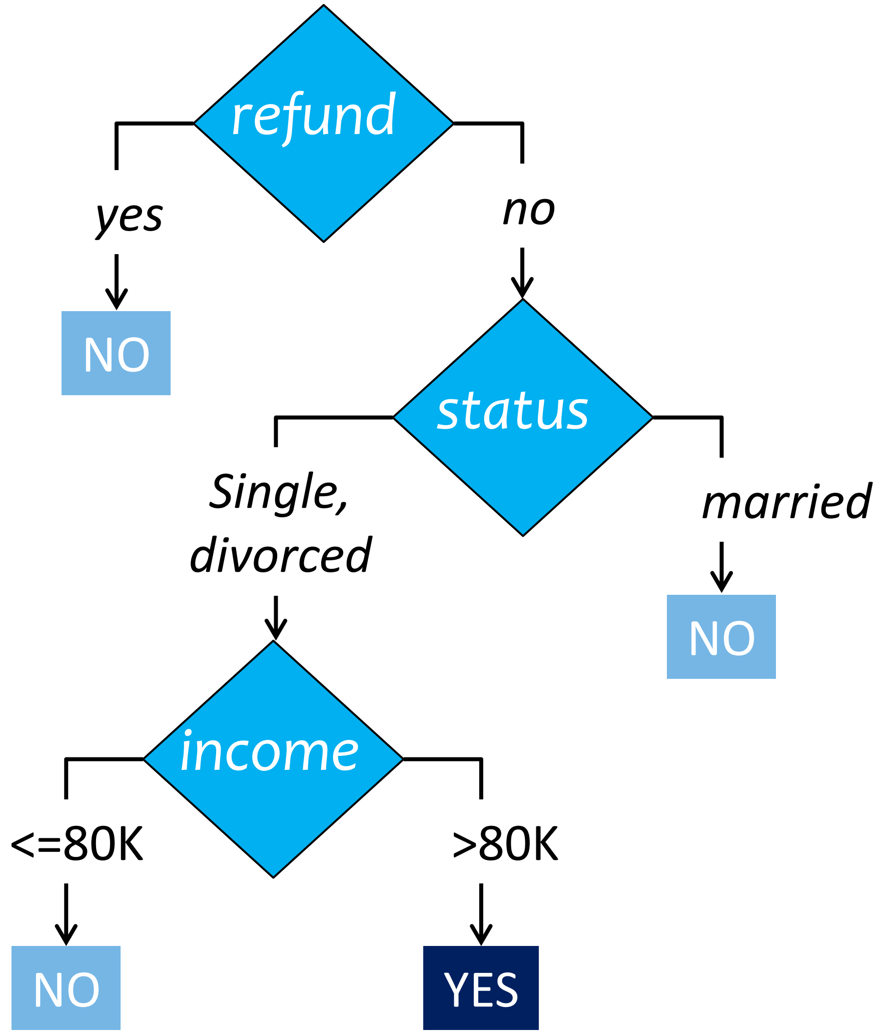
Temario

- Árboles de decisión
  - Modelos
  - Aprendizaje

- Árbol de Decisión:

  El proceso de aprendizaje del árbol de decisión es del tipo top-down recursivo de la forma divide para conquistar

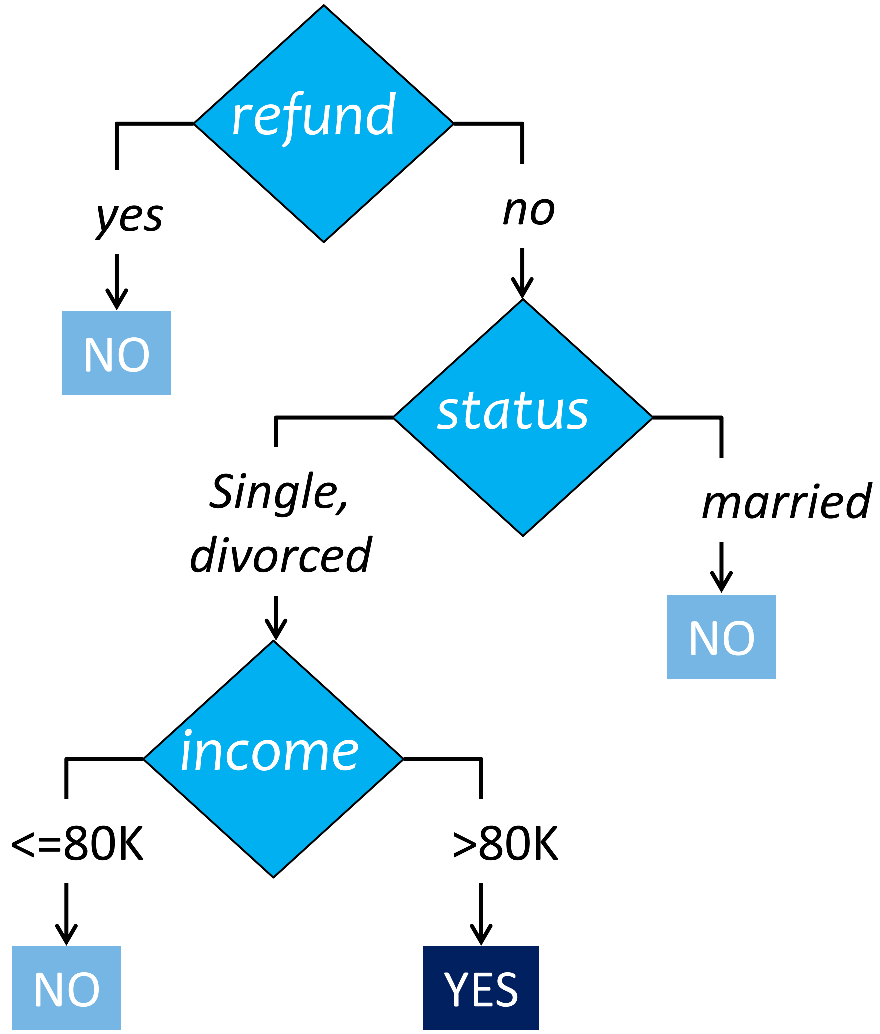
  > Comienza por la raíz
  > Selecciona el mejor atributo o característica
  > Divide los ejemplos según el atributo seleccionado
  > Recursión y repetición



- Problemas

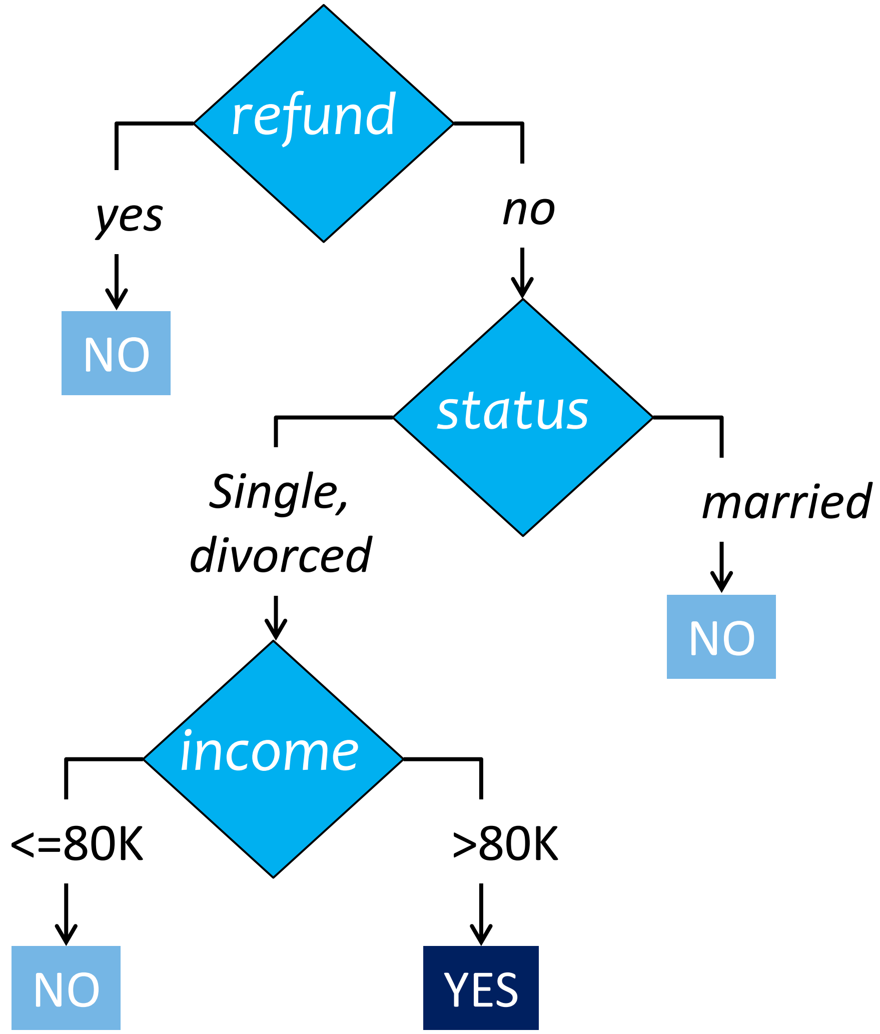
  > Cuando parar de crecer
  > Cómo cortar las partes irrelevantes del árbol

- Métodos de crecimiento completo
  - Todas las instancias de un nodo pertenecen a la misma clase
  - No quedan atributos para más divisiones
  - No hay más instancias para evaluar



- ¿Qué impacto tiene lo anterior en la calidad del árbol?
  - El árbol sobreentrena los dato, lo cual implica que el rendimiento decrece. El modelo aprende los datos de entrenamiento, pero no generaliza con nuevos datos
  - Para evitar esto, empleamos poda (Pruning)

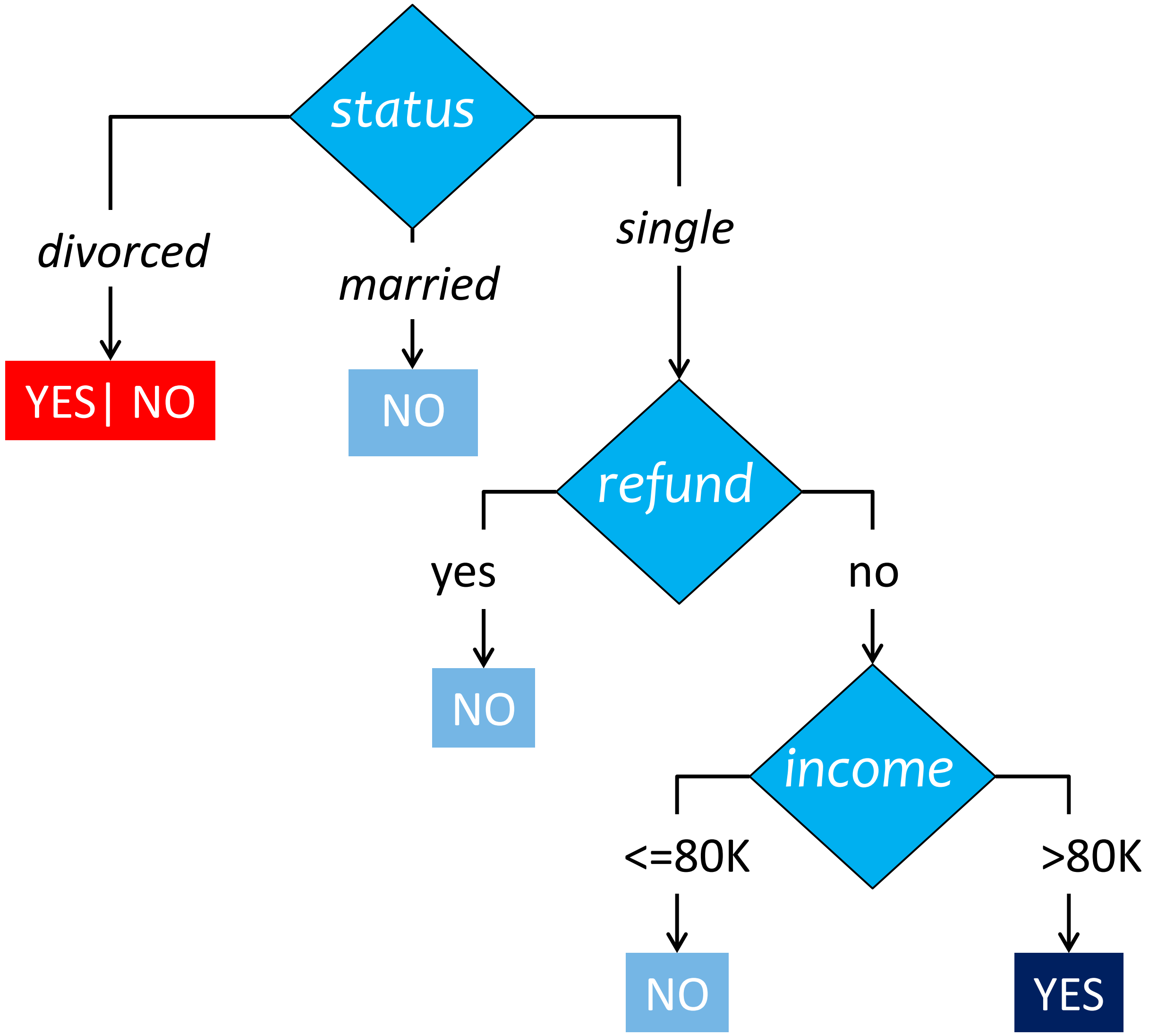
- Prepruning
  - Aplicar un test estadístico para decidir si expandir un nodo
  - Emplear una medida de complejidad para penalizar árboles largo (*Minimum Description Lenght*)



- Postpruning
  - Evaluar un conjunto de ejemplos para evaluar la utilidad de podar nodos del árbol (después que el árbol se encuentre completo)

- Descripción
  Detener el algoritmo antes que se convierta en un árbol completo
  - Parar si el número de instancias de un nodo es menor a un umbral especificado por el usuario (i.e., un porcentaje específico de los datos)

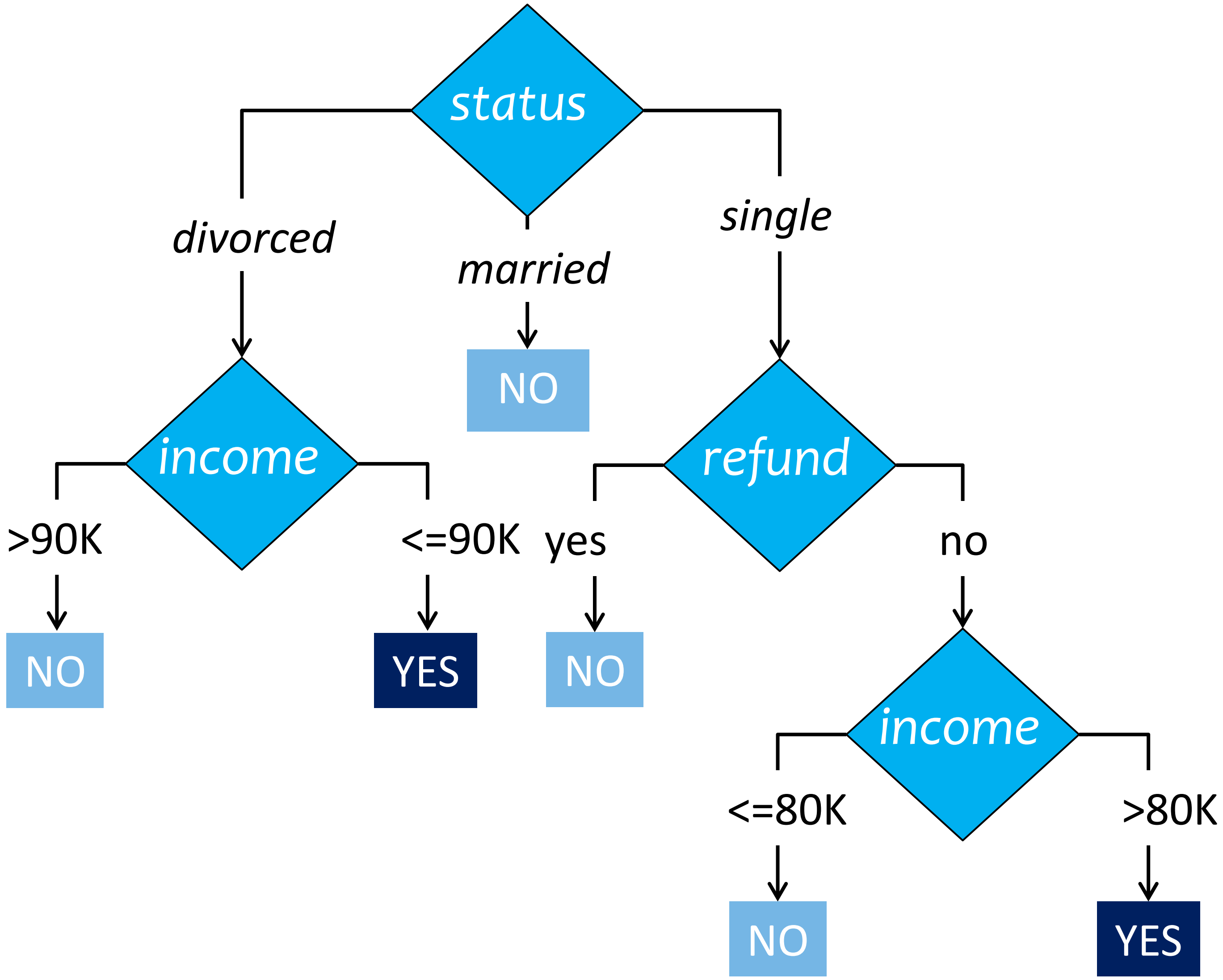
refund	status	income	affair
yes	single	125K	no
no	married	100K	no
no	single	70K	no
yes	married	120K	no
no	divorced	95K	yes
no	married	60K	no
yes	divorced	220K	no
no	single	85K	yes
no	married	75K	no
no	single	90K	yes



- **Descripción**
  Detener el algoritmo antes que se convierta en un árbol completo

  > Parar si expandir el nodo no mejorar una medida de impureza, por ejemplo, entropía.

  > Parar si la distribución de clases de las instancias son independientes de las características disponibles, ejemplo, test Chi-cuadrado, p> 0.05



- Descripción
  Una vez que el árbol esté completo, la poda comienza por los nodos del árbol de decisión en un modo *bottom-up* (abajo hacia arriba)

1. Antes de entrenar, separe un conjunto de instancias X_{prune} para evaluar la utilidad de podar los nodos del árbol

2. Evalue el set X_{prune} empleando el árbol de clasificación

3. Pode un nodo y reemplaze el subárbol por la hoja (la clase de la hoja es determinada por la mayoría de clases de las instancias del subárbol)

4. Re-evalúa el set X_{prune}. Si el rendimiento mejora, acepte la poda

- **Descripción**
  Una vez que el árbol esté completo, la poda comienza por los nodos del árbol de decisión en un modo *bottom-up* (abajo hacia arriba)

- **Métricas**
  Sea X un dataset con N puntos. El error sobre los datos es

$$e(X, M) = \frac{1}{N}\sum_{i=1}^{N} I[f(x(i); M), y(i)] \qquad \text{where } I(a,b) = \begin{cases} 1 & a \neq b \\ 0 & \text{otherwise} \end{cases}$$

> Aproximación optimista: $e(X_{prune}, M) = e(X_{train}, M)$

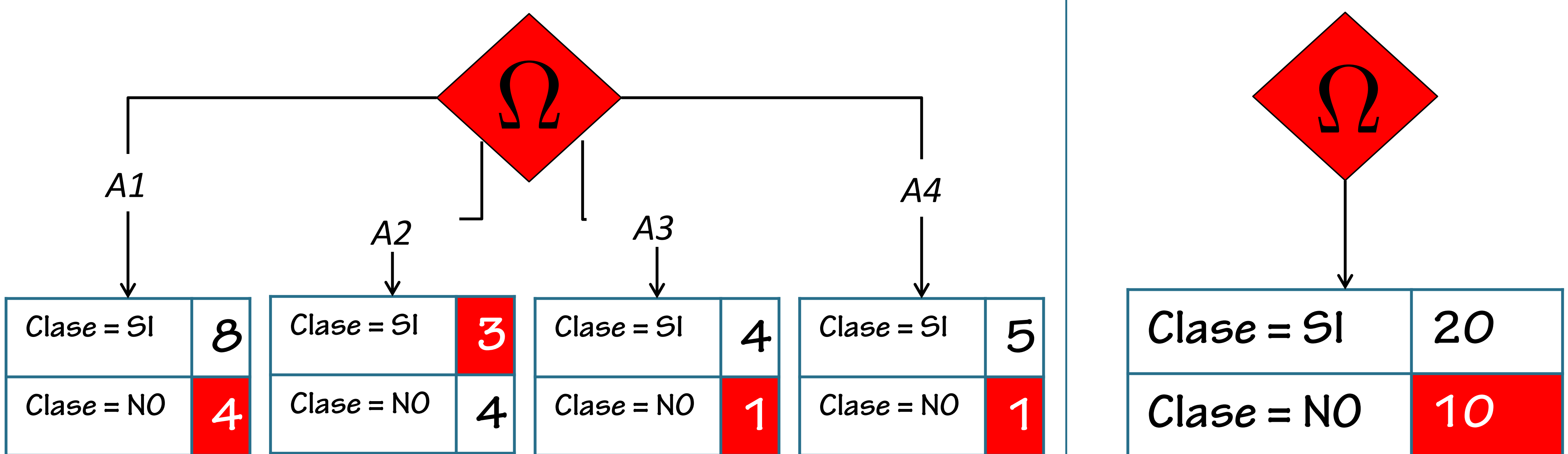
> Aproximación pesimista: $e(X_{prune}, M) = e(X_{train}, M) + \frac{(0.5 \cdot N_{leaf})}{N}$

> Reduced Error Prunning (REP): $e(X_{prune}, M)$

es un conjunto distinto al set de entrenamiento

Clasificadores
Post pruning

- **Ejemplo**
  Suponga dos escenarios sobre un mismo atributo



	Modelo 1	Modelo 2
Aproximación optimista:	$e(X_{pruned}, M) = \dfrac{(4+3+1+1)}{30} = 0.3$	$e(X_{pruned}, M) = \dfrac{10}{30} = 0.33$
Aproximación pesimista:	$e(X_{pruned}, M) = \dfrac{(4+3+1+1)}{30} + \dfrac{0.5 \cdot 4}{30} = 0.37$	$e(X_{pruned}, M) = \dfrac{10}{30} + \dfrac{0.5}{30} = 0.35$
	$N_{leaf} = 4$	$N_{leaf} = 1$