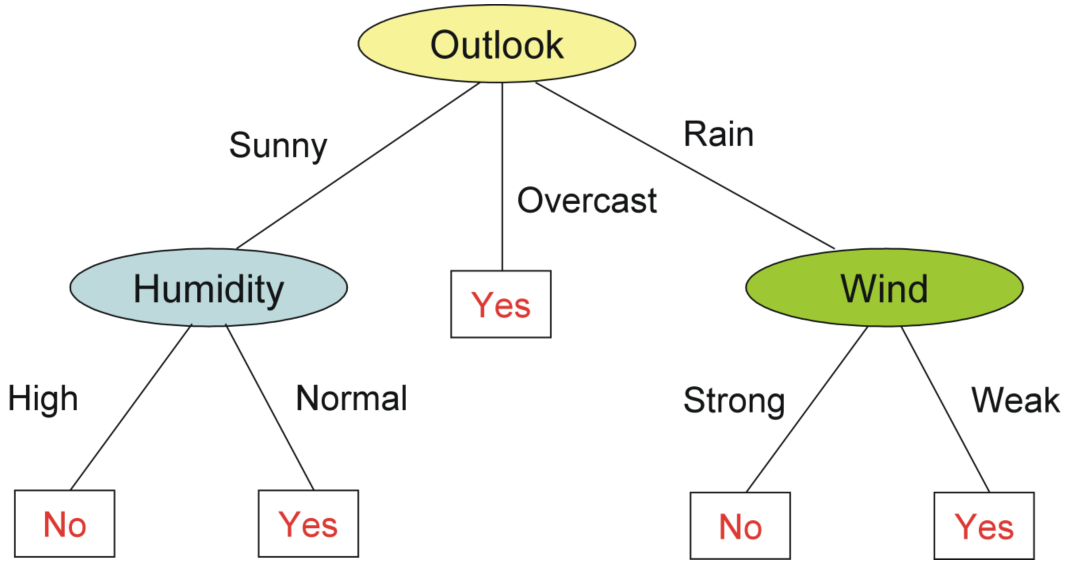
ÁRBOLES DE DECISIÓN

- Árbol de decisión:
  - Es una representación de los procesos de decisión involucrados en las tareas de clasificación
  - Elementos:
    - Hojas: describen la etiqueta asociada a una clasificación
    - Nodos: describen la pregunta acerca de un cierto atributo
  - Ramas de un nodo:
    - Representan los diferentes valores que puede tomar el atributo respecto del que se pregunta en el nodo

Árboles de Decisión

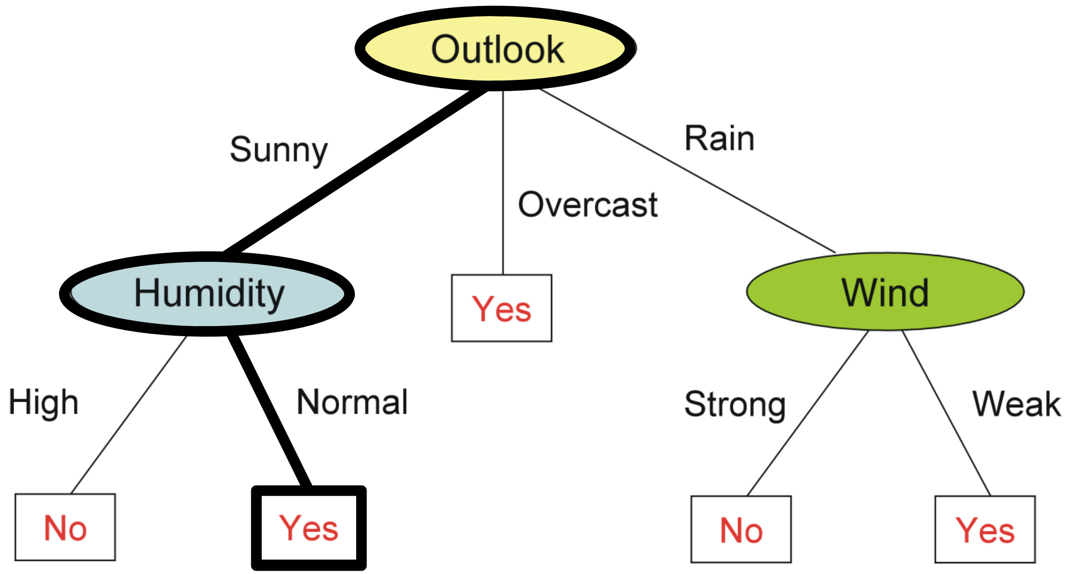
- Ejemplo:
  - Árbol para clasificar si un día es adecuado para jugar al tenis:



- Atributos: Outlook, Humidity, Wind
- Clases: Yes / No

Árboles de Decisión

- Ejemplo:
  - Árbol para clasificar si un día es adecuado para jugar al tenis
    - Si aparece una nueva instancia, se recorre el árbol
      - Ejemplo: (*Outlook=sunny*, *Temperature=Hot*, *Wind=Strong*, *Humidity=Normal*) → se clasifica como *Yes*



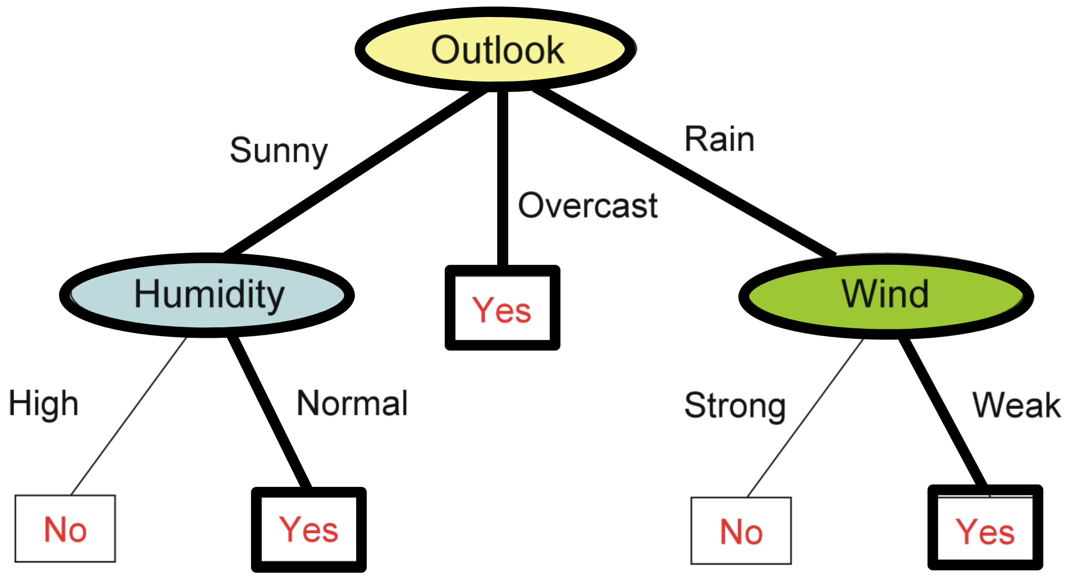
Árboles de Decisión

- En el caso de clasificar en dos clases, el árbol representa una disyunción de conjunciones de restricciones sobre los valores de los atributos de las instancias
  - Un camino = conjunción de tests de atributos
  - Todo el árbol = disyunción de esas conjunciones
  - Ejemplo: árbol anterior:
    - (*Outlook*=*Sunny* ∧ *Humidity*=*Normal*) ∨ (*Outlook*=*Overcast*) ∨ (*Outlook*=*Rain* ∧ *Wind*=*Weak*)

ÁRBOLES DE DECISIÓN

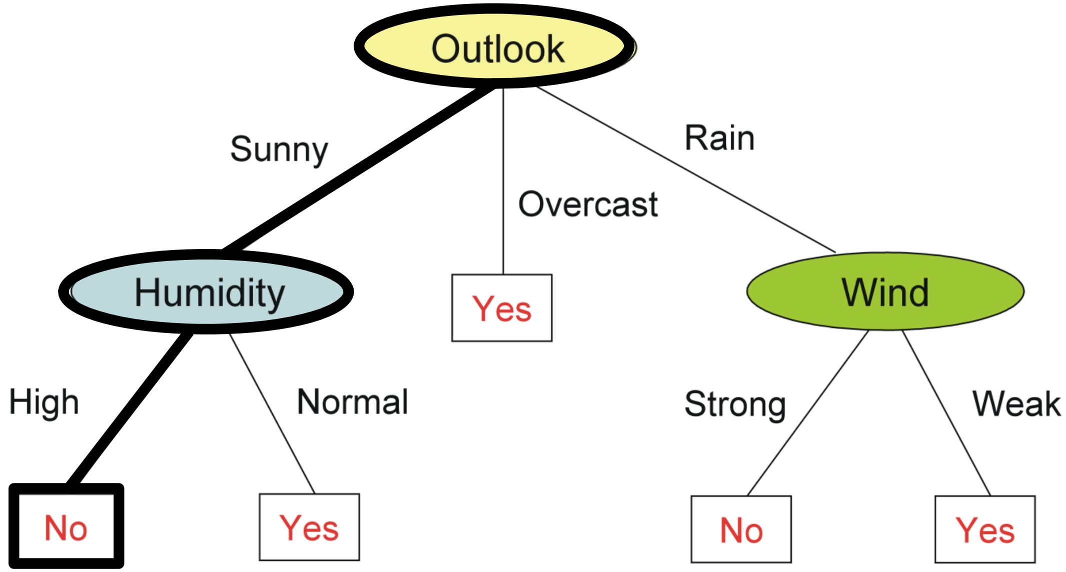
- Disyunción de conjunciones:
  - Ejemplo: árbol anterior:

    ($Outlook=Sunny \wedge Humidity=Normal$) ∨
    ($Outlook=Overcast$) ∨
    ($Outlook=Rain \wedge Wind=Weak$)



Árboles de Decisión

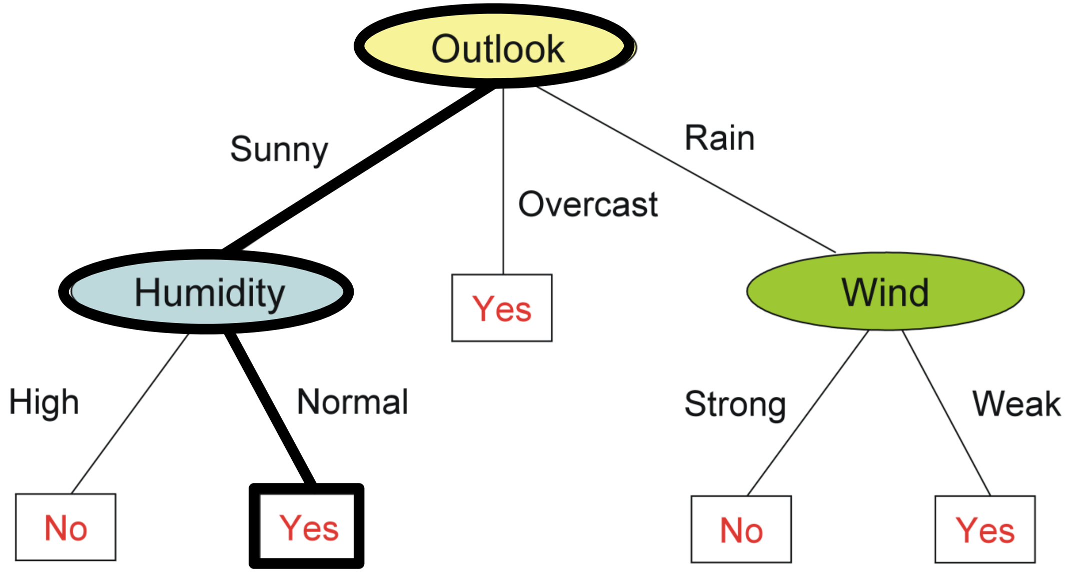
- Además, del árbol se pueden generar una serie de reglas:



  - R1: If (*Outlook=Sunny*) AND (*Humidity=High*) then *PlayTennis=No*

ÁRBOLES DE DECISIÓN

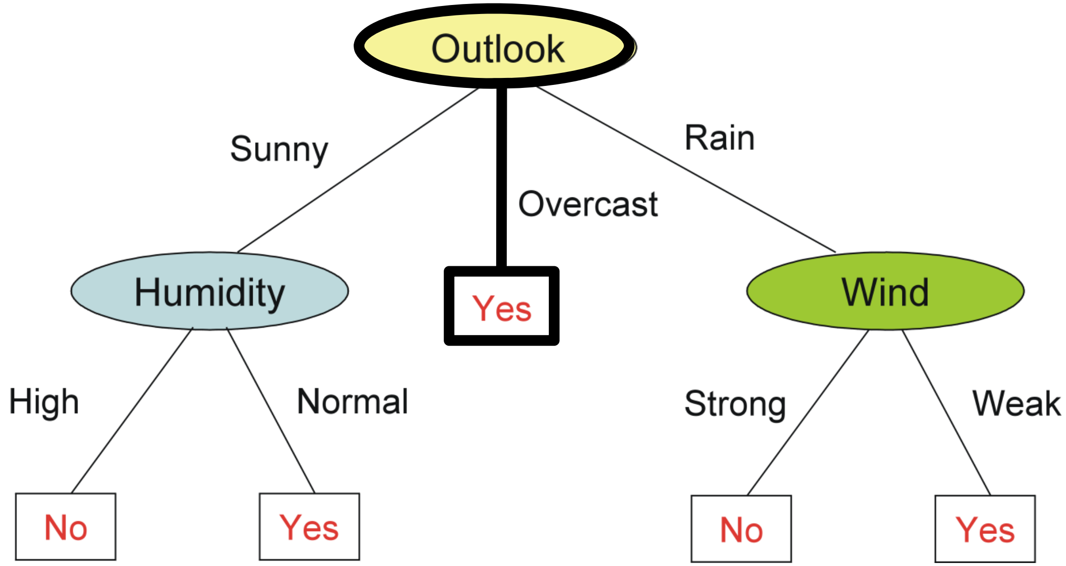
- Además, del árbol se pueden generar una serie de reglas:



  - R1: If (*Outlook=Sunny*) AND (*Humidity=High*) then *PlayTennis=No*
  - R2: If (*Outlook=Sunny*) AND (*Humidity=Normal*) then *PlayTennis=Yes*

Árboles de Decisión

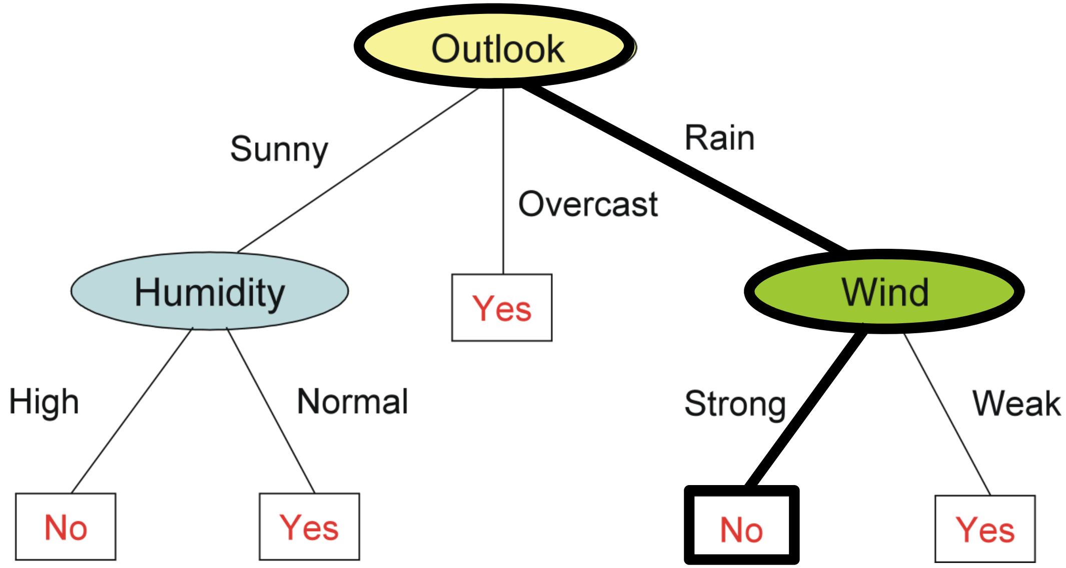
- Además, del árbol se pueden generar una serie de reglas:



  - R1: If (*Outlook=Sunny*) AND (*Humidity=High*) then *PlayTennis=No*
  - R2: If (*Outlook=Sunny*) AND (*Humidity=Normal*) then *PlayTennis=Yes*
  - R3: If (*Outlook=Overcast*) then *PlayTennis=Yes*

Árboles de Decisión

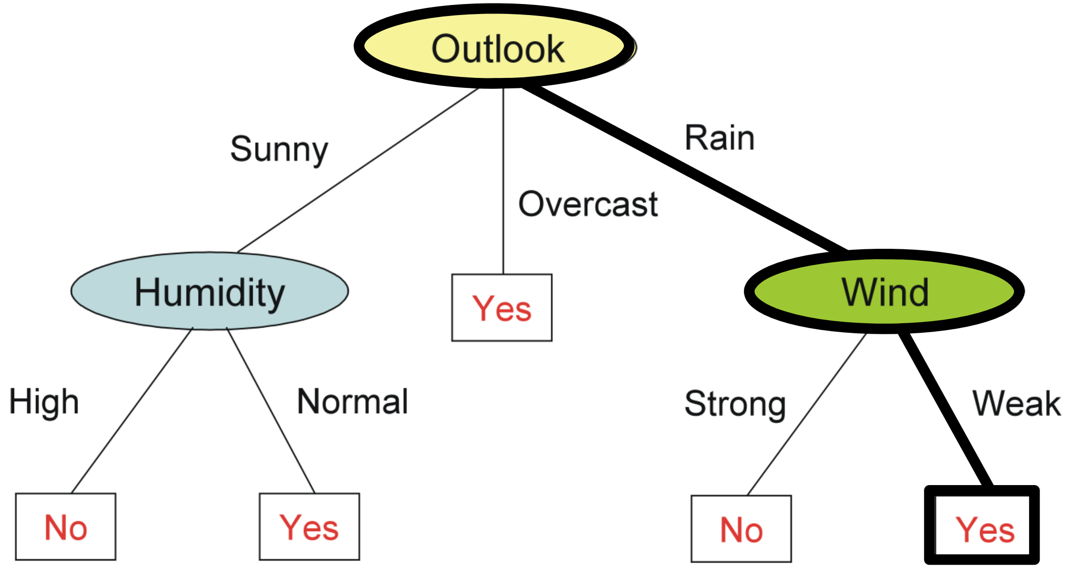
- Además, del árbol se pueden generar una serie de reglas:



  - R1: If (*Outlook*=*Sunny*) AND (*Humidity*=*High*) then *PlayTennis*=*No*
  - R2: If (*Outlook*=*Sunny*) AND (*Humidity*=*Normal*) then *PlayTennis*=*Yes*
  - R3: If (*Outlook*=*Overcast*) then *PlayTennis*=*Yes*
  - R4: If (*Outlook*=*Rain*) AND (*Wind*=*Strong*) then *PlayTennis*=*No*

Árboles de Decisión

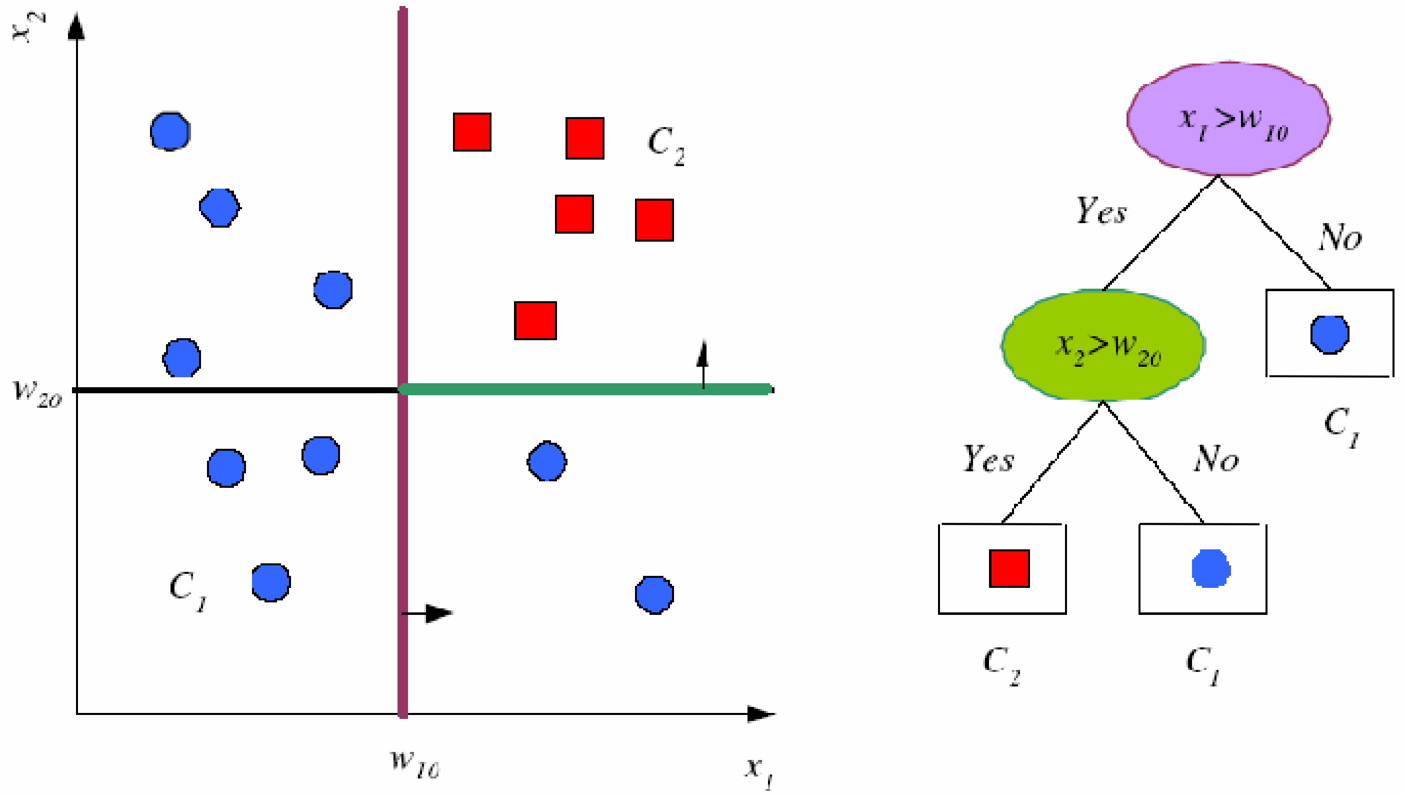
- Además, del árbol se pueden generar una serie de reglas:



  - R1: If (*Outlook=Sunny*) AND (*Humidity=High*) then *PlayTennis=No*
  - R2: If (*Outlook=Sunny*) AND (*Humidity=Normal*) then *PlayTennis=Yes*
  - R3: If (*Outlook=Overcast*) then *PlayTennis=Yes*
  - R4: If (*Outlook=Rain*) AND (*Wind=Strong*) then *PlayTennis=No*
  - R5: If (*Outlook=Rain*) AND (*Wind=Weak*) then *PlayTennis=Yes*

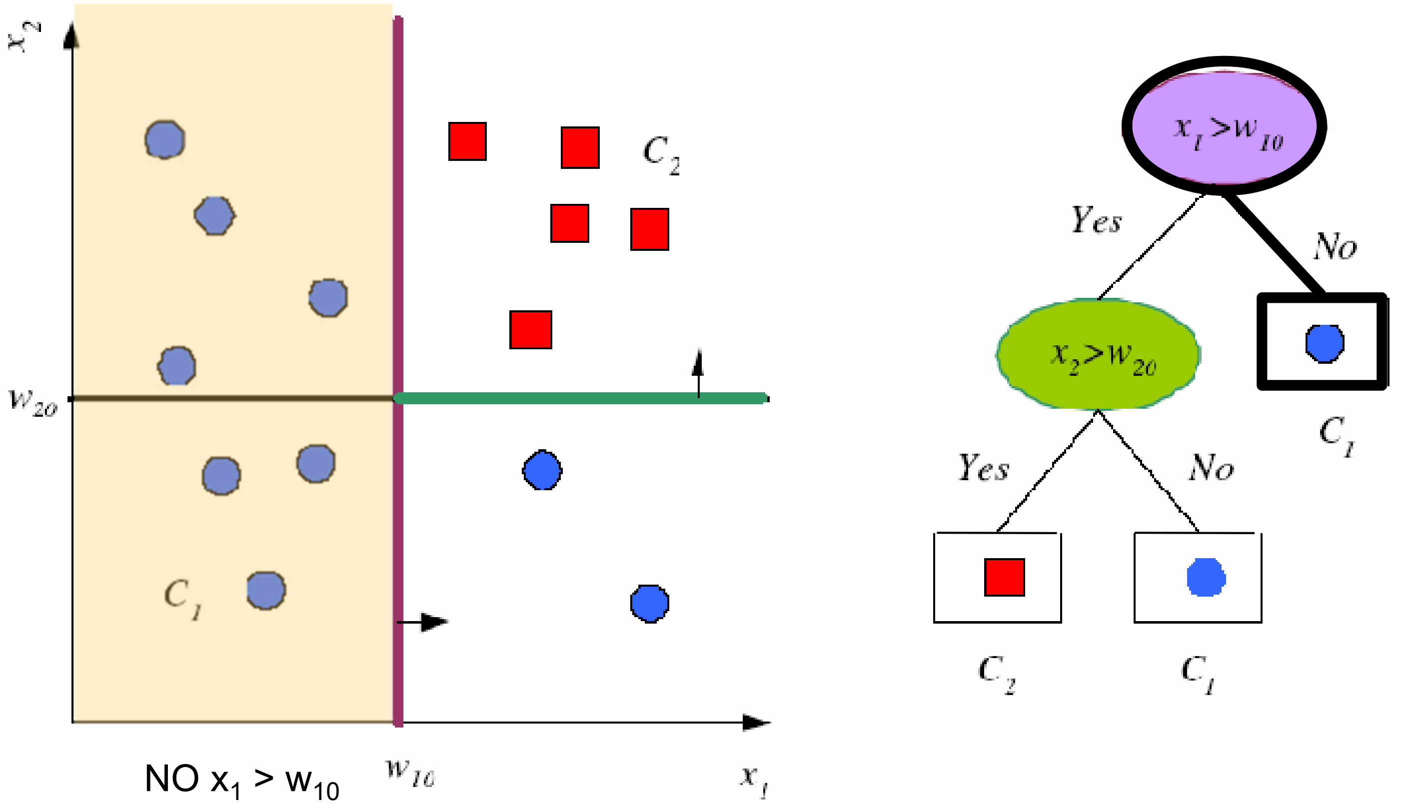
ÁRBOLES DE DECISIÓN

- División de los patrones:
  - Se divide el espacio en regiones etiquetadas con una sola clase y son hiperrectángulos



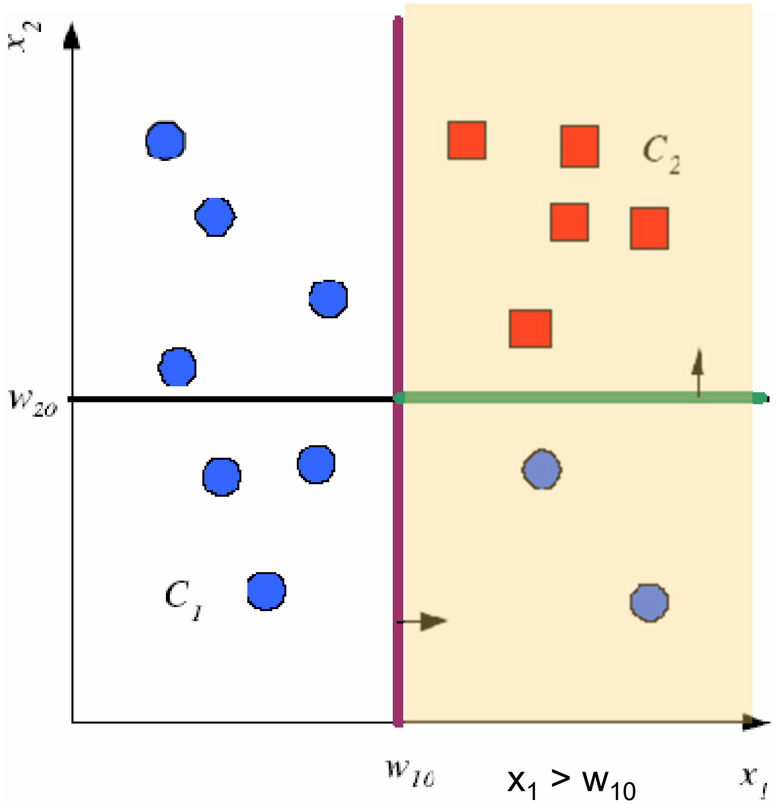
Árboles de Decisión

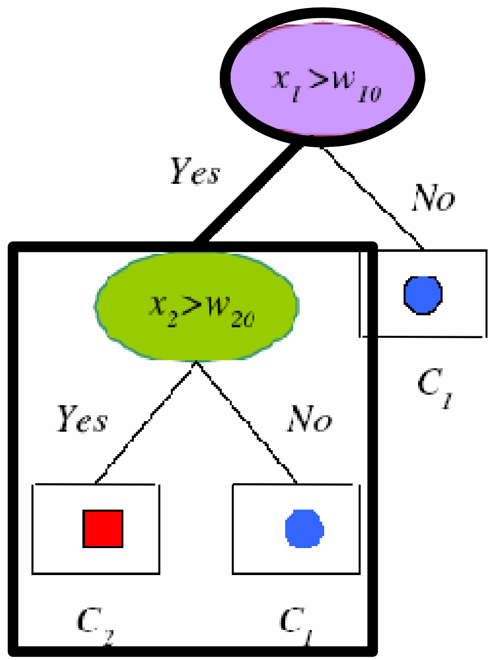
- División de los patrones:
  - Se divide el espacio en regiones etiquetadas con una sola clase y son hiperrectángulos



ÁRBOLES DE DECISIÓN

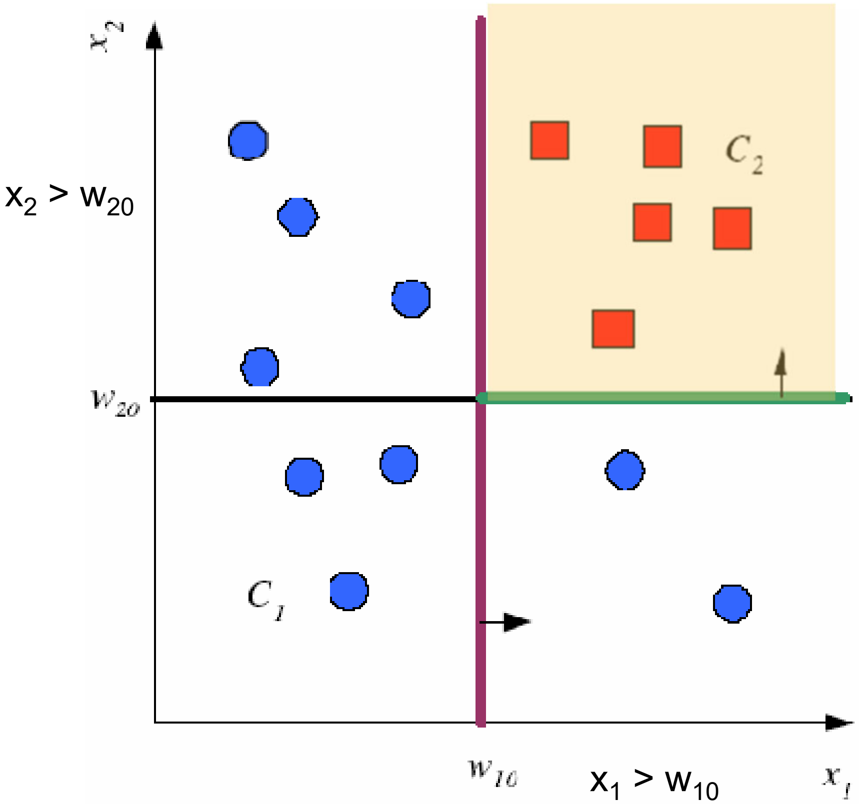
- División de los patrones:
  - Se divide el espacio en regiones etiquetadas con una sola clase y son hiperrectángulos

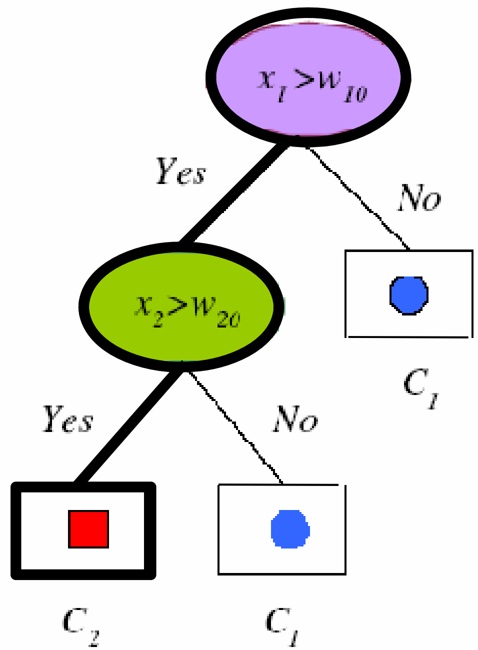




ÁRBOLES DE DECISIÓN

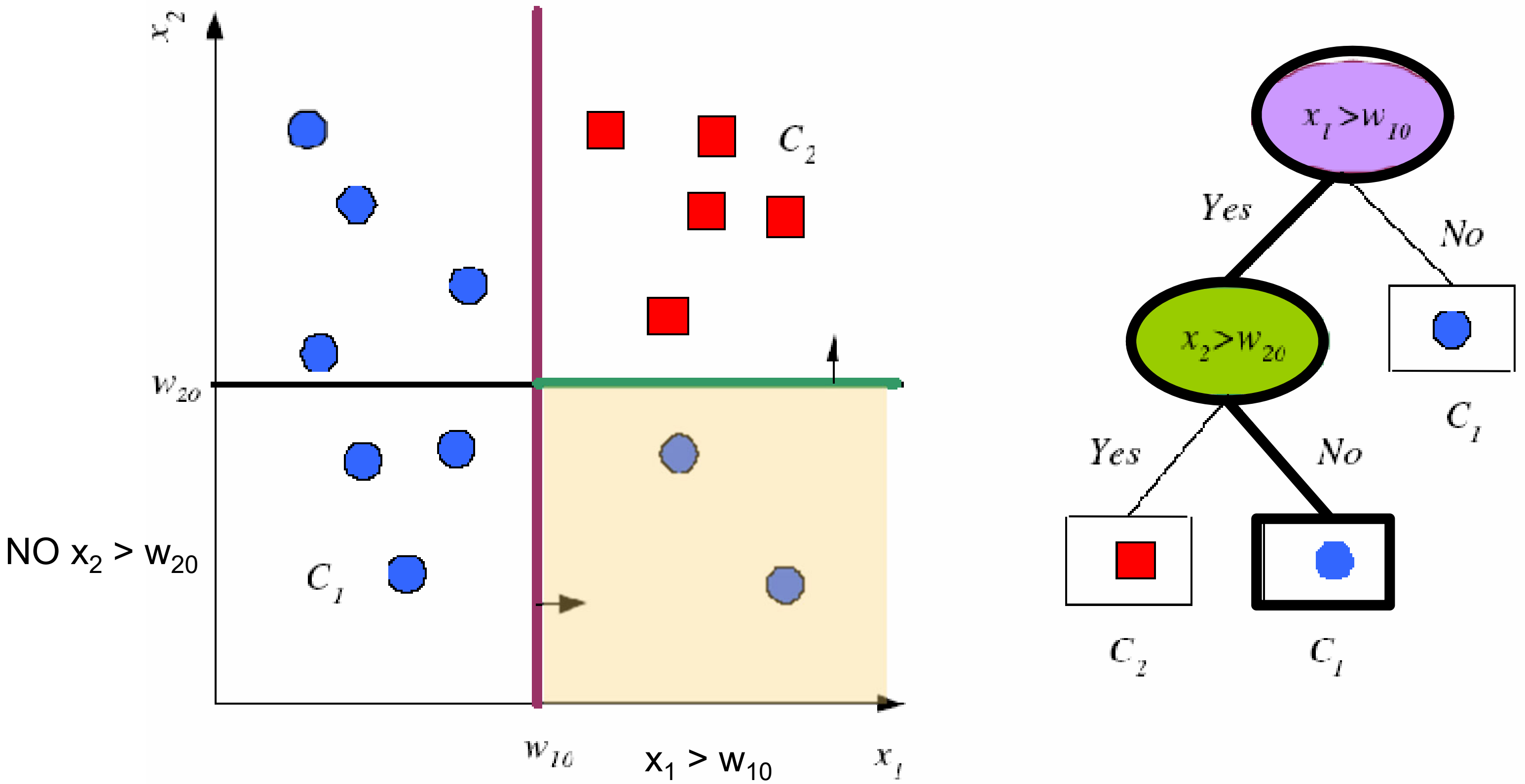
- División de los patrones:
  - Se divide el espacio en regiones etiquetadas con una sola clase y son hiperrectángulos





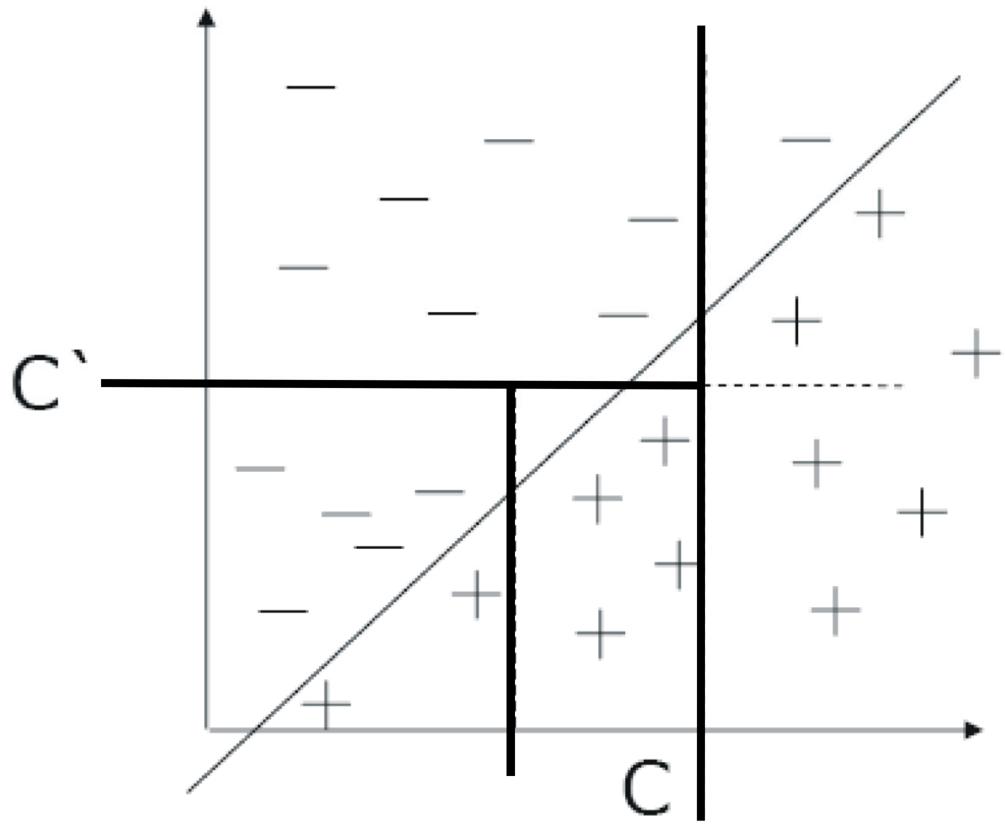
Árboles de Decisión

- División de los patrones:
  - Se divide el espacio en regiones etiquetadas con una sola clase y son hiperrectángulos



ÁRBOLES DE DECISIÓN

- División de los patrones:
  - Se divide el espacio en regiones etiquetadas con una sola clase y son hiperrectángulos



Árboles de Decisión

- Generación del árbol:
  - Varios algoritmos disponibles
  - Objetivos:
    - Dado un conjunto de ejemplos clasificados (aprendizaje supervisado), generar el árbol de decisión óptimo
      - Aquel que permita describir la clasificación con el menor número de cuestiones posible
  - Principales algoritmos:
    - ID3
    - C4.5
    - CART
    - MARS

Árboles de Decisión

- ID3 (*Iterative Dichotomiser 3*) (Quinlan, 1986)
  - Basado en el algoritmo CLS (*Concept Learning Systems*) que usaba sólo atributos binarios
  - Construir el árbol de arriba abajo:
    - ¿Qué atributo seleccionar como nodo raíz?
      - Se evalúa cada atributo para determinar cuán bien clasifica los ejemplos él solo
    - Se selecciona el mejor atributo como nodo y
      - Se abre el árbol para cada posible valor del atributo
      - Los ejemplos se clasifican y colocan en las ramas
        - Según los valores de ese atributo seleccionado como nodo
        - "Partir" la base de datos en subconjuntos, uno para cada rama

Árboles de Decisión

- ID3 (*Iterative Dichotomiser 3*) (Quinlan, 1986)
  - Basado en el algoritmo CLS (*Concept Learning Systems*) que usaba sólo atributos binarios
  - Construir el árbol de arriba abajo:
    - Repetir este proceso de seleccionar atributo / abrir árbol usando los ejemplos asociados con la rama en el que estemos
      - Siempre hacia delante, buscando entre los atributos no usados en este camino
    - Parar cuando el árbol clasifica correctamente los ejemplos o cuando se han usado todos los atributos
    - Etiquetar el nodo hoja con la clase de los ejemplos

Árboles de Decisión

- Algoritmo ID3:
  - ID3 (instancias)
    - SI todas las instancias son de la misma clase C ENTONCES devolver Hoja(C)
    - SINO SI el conjunto de instancias está vacío ENTONCES devolver Hoja(Clase_por_defecto)
    - SINO SI el conjunto de instancias no contiene ningún atributo ENTONCES devolver Hoja(Clase_mayoritaria)
    - SINO
      - Elegir atributo A con mayor ganancia de información
      - Crear nodo con el atributo seleccionado
      - Para cada valor V del atributo A
        - Crear una rama con el valor V
        - Seleccionar las instancias con el valor V del atributo A
        - Eliminar el atributo A de este conjunto de instancias Cv
        - Asignar a la rama el árbol devuelto por la llamada recursiva ID3(Cv)
      - Devolver nodo

Árboles de Decisión

- ID3 (*Iterative Dichotomiser 3*) (Quinlan, 1986)
  - Paso clave: cómo seleccionar el atributo a testear en cada nodo del árbol
    - Seleccionar el atributo más útil
      - El que separe mejor los ejemplos
    - Se utiliza la cantidad de información mutua como medida de la valía de cada atributo
      - Se escoge el atributo que tenga mayor ganancia de información
      - Se espera reducir la entropía (incertidumbre), causada al particionar los ejemplos de acuerdo a este atributo

Árboles de Decisión

- ID3: Definiciones:
  - C: conjunto de clases
  - n: número de patrones
  - n_c: número de patrones de la clase $c \in C$

ÁRBOLES DE DECISIÓN

- ID3: Definiciones:
  - Ejemplo: base de datos de células:
    - C = {Normal, Cancerígena}
      - n = 6

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

ÁRBOLES DE DECISIÓN

- ID3: Definiciones:
  - Ejemplo: base de datos de células:
    - C = {Normal, Cancerígena}
      - $n = 6$
      - $n_1 = 3 \rightarrow 3$ patrones de la clase 1

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

ÁRBOLES DE DECISIÓN

- ID3: Definiciones:
  - Ejemplo: base de datos de células:
    - C = {Normal, Cancerígena}
      - n = 6
      - $n_2 = 3 \rightarrow 3$ patrones de la clase 2

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

ÁRBOLES DE DECISIÓN

- ID3: Definiciones:
  - A: conjunto de atributos
  - A_i: atributo i del conjunto de atributos A
  - V_i: conjunto de valores de A_i

Árboles de Decisión

- ID3: Definiciones:
  - Ejemplo: base de datos de células:
    - A = {Antenas, Colas, Núcleos, Cuerpo}
      - A_1 = Antenas A_2 = Colas A_3 = Núcleos A_4 = Cuerpo
      - $V_1 = \{0,1,2\}$ $V_2 = \{0,1,2\}$ $V_3 = \{0,1,2\}$ V_4 = {Blanco, Rayado}

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Definiciones:
  - n_{ij}: número de patrones que en el atributo A_i toma el valor $j \in V_i$

ÁRBOLES DE DECISIÓN

- ID3: Definiciones:
  - Ejemplo: base de datos de células:
    - A = {Antenas, Colas, Núcleos, Cuerpo}
      - A_4 = Cuerpo, V_4 = {Blanco, Rayado}
        - $n_{41} = 1$

Instancias que en el atributo 4 'Cuerpo' toman el valor 1 'Blanco'

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

ÁRBOLES DE DECISIÓN

- ID3: Definiciones:
  - Ejemplo: base de datos de células:
    - A = {Antenas, Colas, Núcleos, Cuerpo}
      - A_4 = Cuerpo, V_4 = {Blanco, Rayado}
        - $n_{42} = 5$

Instancias que en el atributo 4 'Cuerpo' toman el valor 2 'Rayado'

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Definiciones:
  - n_{ijc}: número de patrones que en el atributo A_i toma el valor j y pertenecen a la clase c

ÁRBOLES DE DECISIÓN

- ID3: Definiciones:
  - Ejemplo: base de datos de células:
    - C = {Normal, Cancerígena}
    - A = {Antenas, Colas, Núcleos, Cuerpo}
      - A_4 = Cuerpo, V_4 = {Blanco, Rayado}
        - $n_{421} = 3$

Instancias que en el atributo 4 ‘Cuerpo’ toman el valor 2 ‘Rayado’ y pertenecen a la clase 1 ‘Normal’

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

ÁRBOLES DE DECISIÓN

- ID3: Definiciones:
  - Ejemplo: base de datos de células:
    - C = {Normal, Cancerígena}
    - A = {Antenas, Colas, Núcleos, Cuerpo}
      - A_4 = Cuerpo, V_4 = {Blanco, Rayado}
        - $n_{422} = 2$

Instancias que en el atributo 4 'Cuerpo' toman el valor 2 'Rayado' y pertenecen a la clase 2 'Cancerígena'

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Definiciones:
  - Al escoger un atributo como nodo ("Cuerpo"):

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

Se divide la base de datos:

Cuerpo

Blanco

Ejemplo	Antenas	Colas	Núcleos	Clase
2	1	0	1	Cancerígena

Rayado

Ejemplo	Antenas	Colas	Núcleos	Clase
1	1	0	2	Normal
3	1	2	0	Normal
4	0	2	1	Normal
5	1	1	1	Cancerígena
6	2	2	1	Cancerígena

ÁRBOLES DE DECISIÓN

- ID3: Definiciones:
  - Al escoger un atributo como nodo ("Cuerpo"):
    - Para cada rama:
      - Rama "Blanco": 1 instancia

Ejemplo	Antenas	Colas	Núcleos	Clase
2	1	0	1	Cancerígena

        - Todas las instancias son de la misma clase: certidumbre total
          - **Entropía 0**
      - Rama "Rayado": 5 instancias
        - 3 instancias clase "Normal" y 2 instancias "Cancerígena"
          - No hay certidumbre
            - **Entropía > 0** (entropía = 0.971)

Ejemplo	Antenas	Colas	Núcleos	Clase
1	1	0	2	Normal
3	1	2	0	Normal
4	0	2	1	Normal
5	1	1	1	Cancerígena
6	2	2	1	Cancerígena

Árboles de Decisión

- ID3: Definiciones:
  - Al escoger un atributo como nodo ("Cuerpo"):
    - Por tanto, la entropía como resultado de esta división con el atributo "Cuerpo" es
      - Suma ponderada de las entropías de las ramas

$$\frac{1 \cdot 0 + 5 \cdot 0.971}{1 + 5} = 0.81$$

      - da una valoración de la incertidumbre en las ramas al dividir con ese atributo

Árboles de Decisión

- ID3: Definiciones:
  - Al escoger un atributo como nodo ("Colas"):

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

Colas

Ejemplo	Antenas	Núcleos	Cuerpo	Clase
5	1	1	Rayado	Cancerígena

Ejemplo	Antenas	Núcleos	Cuerpo	Clase
1	1	2	Rayado	Normal
2	1	1	Blanco	Cancerígena

Ejemplo	Antenas	Núcleos	Cuerpo	Clase
3	1	0	Rayado	Normal
4	0	1	Rayado	Normal
6	2	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Definiciones:
  - Al escoger un atributo como nodo ("Colas"):
    - Para cada rama:
      - Rama "0": 2 instancias
        - Igual número de instancias de cada clase: incertidumbre total
          - **Entropía 1**
      - Rama "1": 1 instancias
        - Todas las instancias de la misma clase: certidumbre total
          - **Entropía 0**
      - Rama "2": 3 instancias
          - 2 instancias de la clase "Normal", 1 "Cancerígena"
            - **Entropía > 0**

              (entropía = 0.9183)

ÁRBOLES DE DECISIÓN

- ID3: Definiciones:
  - Al escoger un atributo como nodo ("Colas"):
    - Por tanto, la entropía como resultado de esta división con el atributo "Colas" es
      - Suma ponderada de las entropías de las ramas

$$\frac{2 \cdot 1 + 1 \cdot 0 + 3 \cdot 0.918}{2 + 1 + 3} = 0.793$$

      - da una valoración de la incertidumbre en las ramas al dividir con ese atributo

Árboles de Decisión

- ID3: Definiciones:
  - Al escoger un atributo como nodo:
    - Comparación de entropías:
      - Dividir con "Cuerpo": Entropía = 0.81
      - Dividir con "Colas": Entropía = 0.79
    - El atributo "Colas" genera una partición con menor entropía (mayor certidumbre) en las ramas
    - Se calcularía la entropía con el resto de atributos
      - Es escoge el de menor entropía
      - Se aplica este proceso de forma recursiva sobre cada subárbol

ÁRBOLES DE DECISIÓN

- ID3: Definiciones:
  - Entropía de información del conjunto de patrones
    - $I = - \sum_{c} (n_c/n) \log_2(n_c/n)$

Árboles de Decisión

- ID3: Definiciones:
  - Entropía de información del conjunto de patrones
    - $I = - \sum_{c} (n_c/n) \log_2(n_c/n)$
    - Ejemplo:
      - $I = - (3/6)\log_2(3/6) - (3/6)\log_2(3/6)$

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Definiciones:
  - Entropía de información del conjunto de patrones
    - $I = - \sum_{c} (n_c/n) \log_2(n_c/n)$
    - Ejemplos:
      - 1 sola clase: $n_c = n$
        - $I = - (n_c/n) \log_2 (n_c/n) = 0$
      - 2 clases, con igual número de patrones en cada una: $n_c = n/2$
        - $I = - 2 * (n_c/n) \log_2 (n_c/n) = - 2 * (1/2) \log_2 (1/2) = 1$
      - 5 clases, con igual número de patrones en cada una: $n_c = n/5$
        - $I = - 5 * (n_c/n) \log_2 (n_c/n) = - 5 * (1/5) \log_2 (1/5) = 2.32$
    - Es decir, cuanto más repartidos estén los patrones en distintas clases, mayor es la entropía
      - Mayor incertidumbre, menos información sobre los patrones

Árboles de Decisión

- ID3: Definiciones:
  - Entropía de información del conjunto de patrones
    - $I = - \sum_{c} (n_c/n) \log_2(n_c/n)$
  - Entropía de información del valor j del atributo A_i
    - $I_{ij} = - \sum_{c}(n_{ijc}/n_{ij}) \log_2(n_{ijc}/n_{ij})$
    - Se realiza la misma operación, pero sólo entre los patrones con valor V_j en el atributo A_i
      - Subtabla
        - Resultado de dividir los datos con el atributo A_i
      - Se mide la entropía en esa subtabla

Árboles de Decisión

- ID3: Definiciones:
  - Entropía de información del conjunto de patrones
    - $I = - \sum_{c} (n_c/n) \log_2(n_c/n)$
  - Entropía de información del valor j del atributo A_i
    - $I_{ij} = - \sum_{c} (n_{ijc}/n_{ij}) \log_2(n_{ijc}/n_{ij})$
    - Ejemplo: I_{42}?
      - Entropía de los patrones que en "Cuerpo" tienen "Rayado"

A = {Antenas, Colas, Núcleos, Cuerpo}
A_4 = Cuerpo, V_4 = {Blanco, Rayado}

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

ÁRBOLES DE DECISIÓN

- ID3: Definiciones:
  - Entropía de información del conjunto de patrones
    - $I = - \sum_{c} (n_c/n) \log_2(n_c/n)$
  - Entropía de información del valor j del atributo A_i
    - $I_{ij} = - \sum_{c} (n_{ijc}/n_{ij}) \log_2(n_{ijc}/n_{ij})$
    - Ejemplo: I_{42}?

      A = {Antenas, Colas, Núcleos, Cuerpo}
      A_4 = Cuerpo, V_4 = {Blanco, Rayado}

      - Subtabla con los patrones que en "Cuerpo" tienen "Rayado":

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

ÁRBOLES DE DECISIÓN

- ID3: Definiciones:
  - Entropía de información del conjunto de patrones
    - $I = - \sum_{c} (n_c/n) \log_2(n_c/n)$
  - Entropía de información del valor j del atributo A_i
    - $I_{ij} = - \sum_{c}(n_{ijc}/n_{ij}) \log_2(n_{ijc}/n_{ij})$
    - Ejemplo: I_{42}?

A = {Antenas, Colas, Núcleos, Cuerpo}
A_4 = Cuerpo, V_4 = {Blanco, Rayado}

      - Subtabla con los patrones que en "Cuerpo" tienen "Rayado":

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

      - $I_{42} = - (3/5)\log_2(3/5) - (2/5)\log_2(2/5) = 0.971$

Árboles de Decisión

- ID3: Definiciones:
  - Entropía de información del conjunto de patrones
    - $I = - \sum_{c} (n_c/n) \log_2(n_c/n)$
  - Entropía de información del valor j del atributo A_i
    - $I_{ij} = - \sum_{c}(n_{ijc}/n_{ij}) \log_2(n_{ijc}/n_{ij})$
    - Ejemplo: I_{41}?
      - Entropía de los patrones que en "Cuerpo" tienen "Blanco"

A = {Antenas, Colas, Núcleos, Cuerpo}
A_4 = Cuerpo, V_4 = {Blanco, Rayado}

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Definiciones:
  - Entropía de información del conjunto de patrones
    - $I = - \sum_c (n_c/n) \log_2(n_c/n)$
  - Entropía de información del valor j del atributo A_i
    - $I_{ij} = - \sum_c (n_{ijc}/n_{ij}) \log_2(n_{ijc}/n_{ij})$
    - Ejemplo: I_{41}?

A = {Antenas, Colas, Núcleos, Cuerpo}
A_4 = Cuerpo, V_4 = {Blanco, Rayado}

- Subtabla de los patrones que en "Cuerpo" tienen "Blanco":

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
2	1	0	1	Blanco	Cancerígena

- $I_{41} = 0$
  - Certidumbre total

ÁRBOLES DE DECISIÓN

- ID3: Definiciones:
  - Entropía de información del conjunto de patrones
    - $I = - \sum_c (n_c/n) \log_2(n_c/n)$
  - Entropía de información del valor j del atributo A_i
    - $I_{ij} = - \sum_c (n_{ijc}/n_{ij}) \log_2(n_{ijc}/n_{ij})$
    - Se realiza la misma operación, pero sólo entre los patrones con valor V_j en el atributo A_i
      - Subtabla
      - Se mide la entropía en esa subtabla
    - Igual que antes, cuanto más repartidos estén los patrones con valor V_j en el atributo A_i entre las distintas clases, mayor será la entropía
      - Más incertidumbre

Árboles de Decisión

- ID3: Definiciones:
  - Entropía de información del conjunto de patrones
    - $I = - \sum_c (n_c/n) \log_2(n_c/n)$
  - Entropía de información del valor j del atributo A_i
    - $I_{ij} = - \sum_c (n_{ijc}/n_{ij}) \log_2(n_{ijc}/n_{ij})$
  - Entropía de información del atributo A_i
    - $I(A_i) = \sum_j (n_{ij}/n)\, I_{ij}$
    - La media ponderada de las entropías para cada uno de los posibles valores en el atributo A_i
      - La media ponderada de las entropías de las subtablas para cada valor de A_i

ÁRBOLES DE DECISIÓN

- ID3: Definiciones:
  - Entropía de información del conjunto de patrones
    - $I = - \sum_{c} (n_c/n) \log_2(n_c/n)$
  - Entropía de información del valor j del atributo A_i
    - $I_{ij} = - \sum_{c}(n_{ijc}/n_{ij}) \log_2(n_{ijc}/n_{ij})$
  - Entropía de información del atributo A_i
    - $I(A_i) = \sum_{j}(n_{ij}/n) I_{ij}$
    - Ejemplo: $I(A_4)$
      - Entropía del atributo "Cuerpo"
      - $I(A_4) = (1/6)*0 + (5/6)*1.971 = 0.8091$
        - Media ponderada de la entropía de ese atributo con cada uno de los valores posibles

A = {Antenas, Colas, Núcleos, Cuerpo}
A_4 = Cuerpo, V_4 = {Blanco, Rayado}

Árboles de Decisión

- ID3: Definiciones:
  - Entropía de información del conjunto de patrones
    - $I = - \sum_c (n_c/n) \log_2(n_c/n)$
  - Entropía de información del valor j del atributo A_i
    - $I_{ij} = - \sum_c (n_{ijc}/n_{ij}) \log_2(n_{ijc}/n_{ij})$
  - Entropía de información del atributo A_i
    - $I(A_i) = \sum_j (n_{ij}/n)\, I_{ij}$
    - Dá una medida de la **incertidumbre asociada al atributo**
      - La incertidumbre que quedaría al dividir los patrones con todos los valores ese atributo
        - Al crear las subtablas, una para cada valor del atributo

Árboles de Decisión

- ID3: Definiciones:
  - Entropía de información del conjunto de patrones
    - $I = - \sum_c (n_c/n) \log_2(n_c/n)$
  - Entropía de información del valor j del atributo A_i
    - $I_{ij} = - \sum_c (n_{ijc}/n_{ij}) \log_2(n_{ijc}/n_{ij})$
  - Entropía de información del atributo A_i
    - $I(A_i) = \sum_j (n_{ij}/n)\, I_{ij}$
  - Ganancia de información del atributo A_i
    - $G(A_i) = I - I(A_i)$
      - Disminución de la entropía en un conjunto de datos tras dividirlo usando ese atributo

Árboles de Decisión

- ID3: Ejemplo:
  - Se desea generar un árbol de decisión que clasifique entre células normales y células cancerígenas según los datos de la siguiente tabla:

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del atributo A_1 (Antenas):
      - Dividir la tabla según los valores de ese atributo

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
4	0	2	1	Rayado	Normal

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
5	1	1	1	Rayado	Cancerígena

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
6	2	2	1	Rayado	Cancerígena

      - Calcular las entropías de esas subtablas
      - Suma de esas entropías ponderadas por el núm. de patrones

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 0 del atributo A_1 (Antenas)

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	**0**	**2**	**1**	**Rayado**	**Normal**
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 0 del atributo A_1 (Antenas)
      - $I_{10} = - (1/1) \log_2(1/1) – (0/1) \log_2 (0/1) = 0$

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
4	0	2	1	Rayado	Normal

ÁRBOLES DE DECISIÓN

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 0 del atributo A_1 (Antenas)
      - $I_{10} = - (1/1)\ \log_2(1/1) - (0/1)\ \log_2(0/1) = 0$

Número de antenas de valor 0, normales

Número de antenas de valor 0

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
4	0	2	1	Rayado	Normal

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 0 del atributo A_1 (Antenas)
      - $I_{10} = -(1/1)\log_2(1/1) - (0/1)\log_2(0/1) = 0$

Número de antenas de valor 0, normales

Número de antenas de valor 0

Número de antenas de valor 0, cancerígenas

Número de antenas de valor 0

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
4	0	2	1	Rayado	Normal

ÁRBOLES DE DECISIÓN

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 0 del atributo A_1 (Antenas)
      - $I_{10} = - (1/1) \log_2(1/1) – (0/1) \log_2 (0/1) = 0$
      - Entropía = 0:
        - Total certidumbre sobre la clase en la que se clasifican los patrones

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
4	0	2	1	Rayado	Normal

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 1 del atributo A_1 (Antenas)

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	**1**	**0**	**2**	**Rayado**	**Normal**
2	**1**	**0**	**1**	**Blanco**	**Cancerígena**
3	**1**	**2**	**0**	**Rayado**	**Normal**
4	0	2	1	Rayado	Normal
5	**1**	**1**	**1**	**Rayado**	**Cancerígena**
6	2	2	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 1 del atributo A_1 (Antenas)
      - $I_{11} = -(2/4)\log_2(2/4) - (2/4)\log_2(2/4) = 1$

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
5	1	1	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 1 del atributo A_1 (Antenas)
      - $I_{11} = - (2/4) \log_2(2/4) - (2/4) \log_2(2/4) = 1$

Número de antenas de valor 1, normales

Número de antenas de valor 1

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
5	1	1	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 1 del atributo A_1 (Antenas)
      - $I_{11} = -(2/4)\log_2(2/4) - (2/4)\log_2(2/4) = 1$

Número de antenas de valor 1, normales

Número de antenas de valor 1

Número de antenas de valor 1, cancerígenas

Número de antenas de valor 1

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
5	1	1	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 1 del atributo A_1 (Antenas)
      - $I_{11} = - (2/4) \log_2(2/4) - (2/4) \log_2(2/4) = 1$
      - Entropía = 1:
        - Mayor entropía (incertidumbre) que se puede tener con 2 clases
          - Repartidas las dos clases en un 50% de los patrones cada una
          - Si hubiese más patrones de una clase que de otra, habría más certidumbre, menos entropía

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
5	1	1	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 2 del atributo A_1 (Antenas)

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	**2**	**2**	**1**	**Rayado**	**Cancerígena**

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 2 del atributo A_1 (Antenas)
      - $I_{12} = -(0/1)\log_2(0/1) - (1/1)\log_2(1/1) = 0$

Número de antenas de valor 2, normales

Número de antenas de valor 2

Número de antenas de valor 2, cancerígenas

Número de antenas de valor 2

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
6	2	2	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 0 del atributo A_1 (Antenas)
      - $I_{10} = - (1/1) \log_2(1/1) – (0/1) \log_2 (0/1) = 0$
    - Entropía de información del valor 1 del atributo A_1 (Antenas)
      - $I_{11} = - (2/4) \log_2(2/4) – (2/4) \log_2(2/4) = 1$
    - Entropía de información del valor 2 del atributo A_1 (Antenas)
      - $I_{12} = - (0/1) \log_2(0/1) – (1/1) \log_2(1/1) = 0$

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del atributo A_1 (Antenas)
      - $I(A_1) = \Sigma\ (n_{ij}/n)\ I_{ij} = (1/6) \cdot I_{10} + (4/6) \cdot I_{11} + (1/6) \cdot I_{12} = \mathbf{0.66}$

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

ÁRBOLES DE DECISIÓN

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del atributo A_1 (Antenas)
      - $I(A_1) = \Sigma\,(n_{ij}/n)\,I_{ij} = (1/6) \cdot I_{10} + (4/6) \cdot I_{11} + (1/6) \cdot I_{12} = \mathbf{0.66}$

Número de antenas de valor 0

Número de antenas de valor 2

Número de antenas de valor 1

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del atributo A_1 (Antenas)
      - $I(A_1) = \Sigma\,(n_{ij}/n)\, I_{ij} = (1/6) \cdot I_{10} + (4/6) \cdot I_{11} + (1/6) \cdot I_{12} = \mathbf{0.66}$

Número de antenas de valor 0

Número de antenas de valor 2

Número de antenas de valor 1

Número de antenas (patrones)

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 0 del atributo A_2 (Colas)
      - $I_{20} = -(1/2)\log_2(1/2) - (1/2)\log_2(1/2) = 1$
    - Entropía de información del valor 1 del atributo A_2 (Colas)
      - $I_{21} = -(0/1)\log_2(0/1) - (1/1)\log_2(1/1) = 0$
    - Entropía de información del valor 2 del atributo A_2 (Colas)
      - $I_{22} = -(2/3)\log_2(2/3) - (1/3)\log_2(1/3) = 0.9183$
    - Entropía de información del atributo A_2 (Colas)
      - $I(A_2) = \Sigma\,(n_{ij}/n)\,I_{ij} = (2/6) \cdot I_{20} + (1/6) \cdot I_{21} + (3/6) \cdot I_{22} = \mathbf{0.79}$

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 0 del atributo A_3 (Núcleos)
      - $I_{30} = -(1/1)\log_2(1/1) - (0/1)\log_2(0/1) = 0$
    - Entropía de información del valor 1 del atributo A_3 (Núcleos)
      - $I_{31} = -(1/4)\log_2(1/4) - (3/4)\log_2(3/4) = 0.8113$
    - Entropía de información del valor 2 del atributo A_3 (Núcleos)
      - $I_{32} = -(1/1)\log_2(1/1) - (0/1)\log_2(0/1) = 0$
    - Entropía de información del atributo A_3 (Núcleos)
      - $I(A_3) = \Sigma\,(n_{ij}/n)\,I_{ij} = (1/6)\cdot I_{30} + (4/6)\cdot I_{31} + (1/6)\cdot I_{32} = \mathbf{0.54}$

ÁRBOLES DE DECISIÓN

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor blanco del atributo A_4 (Cuerpo)
      - $I_{40} = -(0/1)\log_2(0/1) - (1/1)\log_2(1/1) = 0$
    - Entropía de información del valor rayado del atributo A_4 (Cuerpo)
      - $I_{41} = -(2/5)\log_2(2/5) - (3/5)\log_2(3/5) = 0.9710$
    - Entropía de información del atributo A_4 (Cuerpo)
      - $I(A_4) = \Sigma\,(n_{ij}/n)\, I_{ij} = (1/6) \cdot I_{40} + (5/6) \cdot I_{41} = \mathbf{0.81}$

ÁRBOLES DE DECISIÓN

- ID3: Ejemplo:
  - Se escoge el atributo con menor entropía:
    - La entropía de cada atributo determina el nivel de incertidumbre que se tiene al separar la base de datos con todos los valores de ese atributo
    - El atributo con menor entropía indica que al separar los patrones con ese atributo, cada uno de los conjuntos resultantes tendrá más certidumbre en los datos
      - Es decir, los patrones de las distintas clases se aunarán más en cada uno de esos conjuntos (cada uno de los valores de ese atributo)
      - Habrá menos disparidad de patrones de distintas clases en cada conjunto
      - Es decir, explica mejor los conjuntos
    - Atributo con menor entropía: Núcleos

Árboles de Decisión

- ID3: Ejemplo:
  - Se escoge el atributo con menor entropía: Núcleos
  - Se dividen los patrones:
    - Núcleos = 0 (entropía: 0)

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
3	1	2	0	Rayado	Normal

- Núcleos = 1 (entropía: 0.8113)

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
2	1	0	1	Blanco	Cancerígena
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

- Núcleos = 2 (entropía: 0)

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal

Árboles de Decisión

- ID3: Ejemplo:
  - Si se escogiese como atributo para particionar Antenas:

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
4	0	2	1	Rayado	Normal

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena
3	1	2	0	Rayado	Normal
5	1	1	1	Rayado	Cancerígena

- La entropía de este conjunto con Antenas=1 (entropía: 1) hace que la entropía de este atributo sea mayor que la de Núcleos
  - Entropía de Núcleos con valor 1: 0.8113
  - Mayor disparidad de clases con Antenas=1 que con Núcleos=1

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
6	2	2	1	Rayado	Cancerígena

ÁRBOLES DE DECISIÓN

- ID3: Ejemplo:
  - Si se escogiese como atributo para particionar Colas:

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
2	1	0	1	Blanco	Cancerígena

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
5	1	1	1	Rayado	Cancerígena

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
6	2	2	1	Rayado	Cancerígena

- En Colas=0 y Colas=2 conjuntamente se tiene mayor disparidad que con Núcleos=1

ÁRBOLES DE DECISIÓN

- ID3: Ejemplo:
  - Si se escogiese como atributo para particionar Cuerpo:

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
1	1	0	2	Rayado	Normal
3	1	2	0	Rayado	Normal
4	0	2	1	Rayado	Normal
5	1	1	1	Rayado	Cancerígena
6	2	2	1	Rayado	Cancerígena

- Entropía muy alta (0.971) en un número muy alto de patrones

Ejemplo	Antenas	Colas	Núcleos	Cuerpo	Clase
2	1	0	1	Blanco	Cancerígena

Árboles de Decisión

- ID3: Ejemplo:
  - Se escoge el atributo con menor entropía: Núcleos
  - Se dividen los patrones y se elimina ese atributo:
    - Núcleos = 0

Ejemplo	Antenas	Colas	Cuerpo	Clase
3	1	2	Rayado	Normal

- Núcleos = 1

Ejemplo	Antenas	Colas	Cuerpo	Clase
2	1	0	Blanco	Cancerígena
4	0	2	Rayado	Normal
5	1	1	Rayado	Cancerígena
6	2	2	Rayado	Cancerígena

- Núcleos = 2

Ejemplo	Antenas	Colas	Cuerpo	Clase
1	1	0	Rayado	Normal

ÁRBOLES DE DECISIÓN

- ID3: Ejemplo:

Núcleos

0

Ejemplo	Antenas	Colas	Cuerpo	Clase
3	1	2	Rayado	Normal

2

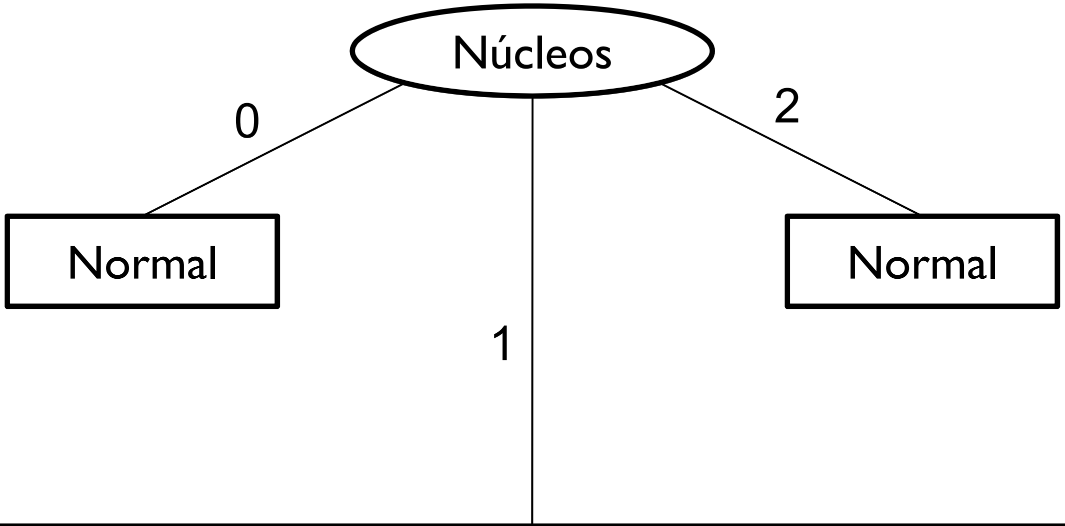
Ejemplo	Antenas	Colas	Cuerpo	Clase
1	1	0	Rayado	Normal

1

Ejemplo	Antenas	Colas	Cuerpo	Clase
2	1	0	Blanco	Cancerígena
4	0	2	Rayado	Normal
5	1	1	Rayado	Cancerígena
6	2	2	Rayado	Cancerígena

Árboles de Decisión

- ID3: Ejemplo:



Ejemplo	Antenas	Colas	Cuerpo	Clase
2	1	0	Blanco	Cancerígena
4	0	2	Rayado	Normal
5	1	1	Rayado	Cancerígena
6	2	2	Rayado	Cancerígena

ÁRBOLES DE DECISIÓN

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 0 del atributo A_1 (Antenas)
      - $I_{10} = -(1/1)\log_2(1/1) - (0/1)\log_2(0/1) = 0$
    - Entropía de información del valor 1 del atributo A_1 (Antenas)
      - $I_{11} = -(0/2)\log_2(0/2) - (2/2)\log_2(2/2) = 0$
    - Entropía de información del valor 2 del atributo A_1 (Antenas)
      - $I_{12} = -(0/1)\log_2(0/1) - (1/1)\log_2(1/1) = 0$
    - Entropía de información del atributo A_1 (Antenas)
      - $I(A_1) = \Sigma\,(n_{ij}/n)\,I_{ij} = (1/4)\cdot I_{10} + (2/4)\cdot I_{11} + (1/4)\cdot I_{12} = \mathbf{0}$

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor 0 del atributo A_2 (Colas)
      - $I_{20} = -(0/1)\log_2(0/1) - (1/1)\log_2(1/1) = 0$
    - Entropía de información del valor 1 del atributo Colas
      - $I_{21} = -(0/1)\log_2(0/1) - (1/1)\log_2(1/1) = 0$
    - Entropía de información del valor 2 del atributo Colas
      - $I_{22} = -(1/2)\log_2(1/2) - (1/2)\log_2(1/2) = 1$
    - Entropía de información del atributo Colas
      - $I(A_2) = \Sigma\,(n_{ij}/n)\,I_{ij} = (1/4)\cdot I_{20} + (1/4)\cdot I_{21} + (2/4)\cdot I_{22} = \mathbf{0.5}$

Árboles de Decisión

- ID3: Ejemplo:
  - Se calculan las entropías:
    - Entropía de información del valor "Blanco" del atributo A_4 (Cuerpo)
      - $I_{40} = - (0/1) \log_2(0/1) – (1/1) \log_2 (1/1) = 0$
    - Entropía de información del valor "Rayado" del atributo A_4 (Cuerpo)
      - $I_{41} = - (1/3) \log_2 (1/3) – (2/3) \log_2 (2/3) = 0.9183$
    - Entropía de información del atributo A4 (Cuerpo)
      - $I(A_4) = \Sigma (n_{ij}/n) I_{ij} = (1/4) \cdot I_{40} + (3/4) \cdot I_{41} = \mathbf{0.6887}$

Árboles de Decisión

- ID3: Ejemplo:
  - Se escoge el atributo con menor entropía: Antenas
  - Se dividen los patrones y se elimina ese atributo:
    - Antenas = 0

Ejemplo	Colas	Cuerpo	Clase
4	2	Rayado	Normal

    - Antenas = 1

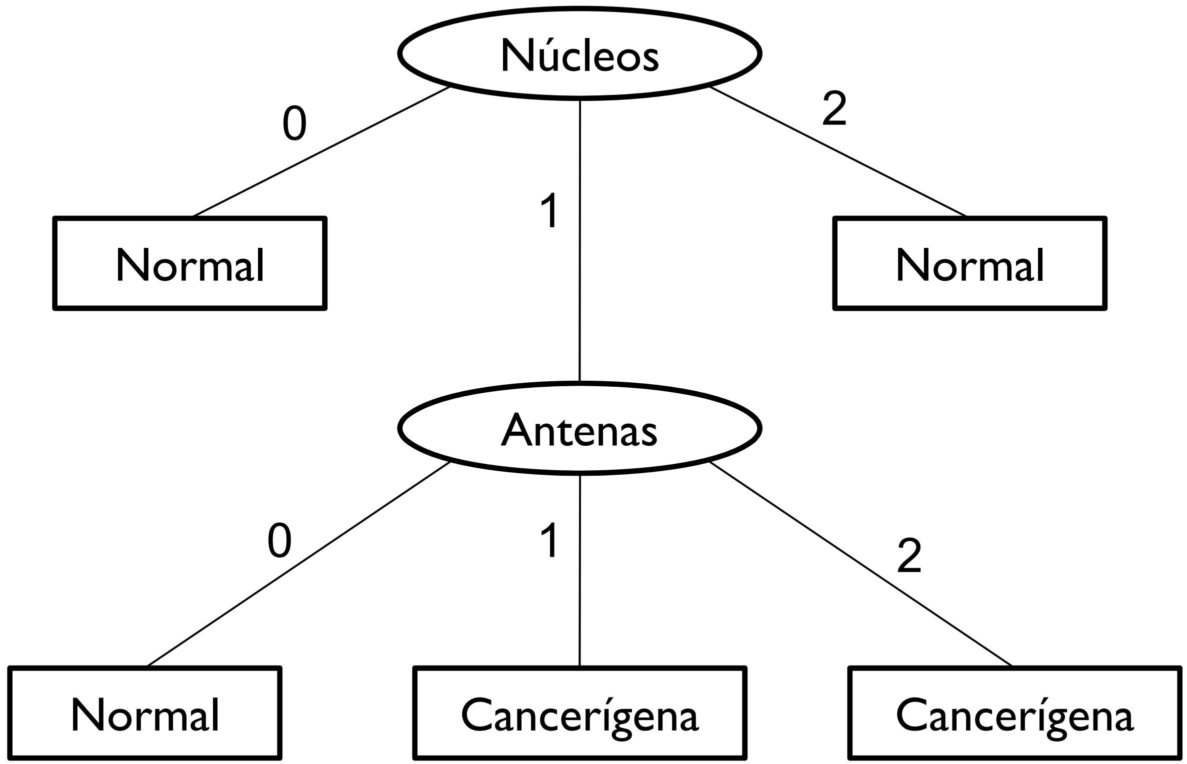
Ejemplo	Colas	Cuerpo	Clase
2	0	Blanco	Cancerígena
5	1	Rayado	Cancerígena

    - Antenas = 2

Ejemplo	Colas	Cuerpo	Clase
6	2	Rayado	Cancerígena

Árboles de Decisión

- ID3: Ejemplo:



Árboles de Decisión

- ID3: Algunas cuestiones prácticas:
  - ¿Cuánto hacer crecer el árbol?
  - ¿Es adecuada esta medida de selección de los atributos?
  - ¿Cómo manejar datos de entrenamiento con valores perdidos (desconocidos) en los atributos?
  - ¿Cómo manejar atributos con costes diferentes?
  - ¿Cómo manejar atributos continuos?

Árboles de Decisión

- ID3: ¿Cuánto hacer crecer el árbol?
  - Si se hace crecer el árbol hasta que clasifique correctamente todos los ejemplos de entrenamiento: sobreentrenamiento o sobreajuste:
    - Si hay ruido en los ejemplos, se aprende el ruido
    - Si hay pocos ejemplos asociados a los nodos hoja, no son representativos de la verdadera función
  - ID3 produce árboles que sobreajustan los ejemplos de entrenamiento, y no funcionan adecuadamente con nuevos ejemplos
    - El modelo no es capaz de generalizar

Árboles de Decisión

- ID3: Soluciones al sobreajuste:
  - Pre-poda:
  - Post-poda:

Árboles de Decisión

- ID3: Soluciones al sobreajuste:
  - Pre-poda:
    - Parar de aumentar el árbol antes de que alcance el punto en el que clasifica perfectamente los ejemplos de entrenamiento
    - ¿Cuándo hacerlo?

ÁRBOLES DE DECISIÓN

- ID3: Soluciones al sobreajuste:
  - Pre-poda:
    - Parar de aumentar el árbol antes de que alcance el punto en el que clasifica perfectamente los ejemplos de entrenamiento
    - ¿Cuándo hacerlo?
      - Aplicar un test estadístico para estimar si expandiendo un nodo particular es probable producir una mejora más allá del conjunto de entrenamiento

Árboles de Decisión

- ID3: Soluciones al sobreajuste:
  - Pre-poda:
    - Parar de aumentar el árbol antes de que alcance el punto en el que clasifica perfectamente los ejemplos de entrenamiento
    - ¿Cuándo hacerlo?
  - Post-poda:
    - Permitir que sobreajuste los datos, y después podarlo reemplazando subárboles por una hoja
    - Requiere mayor coste computacional, pero ofrece mejores resultados

Árboles de Decisión

- ID3: Soluciones al sobreajuste: Post-poda:
  - Comenzando desde abajo, examinar los subárboles de los nodos no terminales
  - Podar un nodo conlleva:
    - 1.- Eliminar su subárbol correspondiente con raíz en ese nodo
    - 2.- Convertir el nodo en nodo hoja
    - 3.- Asignarle la clasificación más común de los ejemplos de entrenamiento asociados a ese nodo
  - Se poda solamente si el árbol podado resultante mejora o iguala el rendimiento del árbol original
    - Test (validación)
  - Se poda iterativamente, escogiendo siempre el nodo cuya poda implique mayor mejora, hasta que ya no haya mejora

Árboles de Decisión

- ID3: ¿Es adecuada esta medida de selección de los atributos?
  - El algoritmo ID3 favorece la elección de atributos con muchos valores
    - Por ejemplo, atributo "Fecha" en un conjunto con los ejemplos en días distintos
      - Este atributo caracteriza perfectamente cada ejemplo de entrenamiento, y se tendría un árbol perfecto y de profundidad 1, con un nodo hoja por patrón
      - La ganancia en información es la mayor
        - Al tener tantas ramas obliga a separar los ejemplos en conjuntos muy pequeños
      - Pero sería muy malo con ejemplos no vistos (test)

Árboles de Decisión

- ID3: ¿Es adecuada esta medida de selección de los atributos?
  - Hay otras medidas en lugar de la ganancia de info.
    - Ratio de ganancia: penaliza los atributos con muchos valores y muchos uniformemente distribuidos

$$\frac{G(A_i)}{H(A_i)}$$

donde $H(A_i)$ denota la entropía con respecto a los valores de A_i

$$H(A_i) = -\sum_j \frac{n_{ij}}{n} \log_2 \frac{n_{ij}}{n}$$

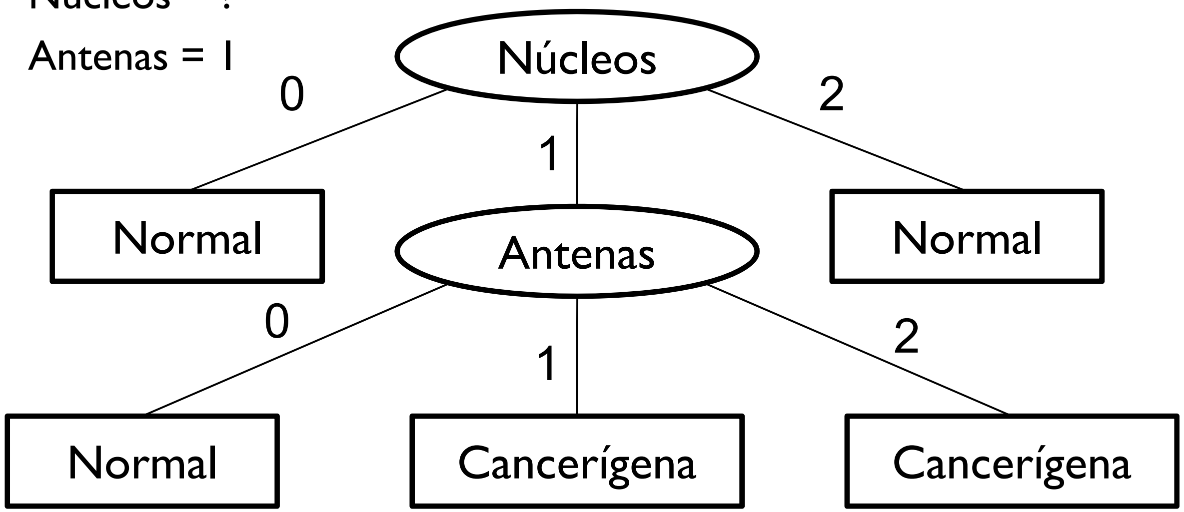
      - No tiene en cuenta las clases, solamente los diferentes valores que toma
      - Cuando un atributo tiene muchos valores distintos, tiene una entropía muy grande, con lo que se divide por un valor mayor
    - El ratio de ganancia favorece aquellos atributos que, en igualdad de ganancia, separen los datos en menos subconjuntos

Árboles de Decisión

- ID3: ¿Cómo manejar atributos con valores perdidos (*missing*)?
  - A la hora de crear el árbol:
    - Eliminar instancias incompletas
    - Estimar dichos valores (imputación)
      - Asignarles la moda:
        - Valor más común entre los ejemplos asociados al nodo donde se está calculando, o entre los ejemplos del nodo que tienen la misma etiqueta que la instancia a imputar
      - Asignar a cada posible valor una probabilidad (frecuencia) de acuerdo con el resto de ejemplos.
        - A continuación, repartir el ejemplo en cada uno de sus valores de acuerdo con la probabilidad y hacer el cálculo de la Ganancia

ÁRBOLES DE DECISIÓN

- ID3: ¿Cómo manejar atributos con valores perdidos (*missing*)?
  - A la hora de clasificar un nuevo caso con un valor perdido:
    - Los casos con valores desconocidos se clasifican de acuerdo con la mayor probabilidad que proporcione el árbol
      - Ejemplo:
        - Núcleos = ?
        - Antenas = 1



Árboles de Decisión

- ID3: ¿Cómo manejar atributos con costes diferentes?
  - ¿Todos los atributos son igual de valiosos al hacer una clasificación?
  - Introducir el coste en la medida de selección de atributos
    - Por ejemplo:

$$\frac{G(A_j)}{Coste(A_j)} \qquad \frac{2^{G(A_j)}-1}{(Coste(A_j)+1)^w}$$

w: constante entre 0 y 1 que pondera la importancia del coste frente a la ganancia

  - Tratan de situar los de bajo coste arriba del árbol, acudiendo a los de alto coste sólo cuando sea necesario
    - Es decir, cuando se quiera mejorar la clasificación

Árboles de Decisión

- ID3: ¿Cómo manejar atributos continuos?
  - Se discretizan:
    - Se particionan en un conjunto discreto de intervalos
      - Ejemplo: 2 intervalos: A<c y A>=c
    - Buscar un valor de c que produzca la mayor ganancia de información
      - Una vez ordenados los ejemplos por el atributo A
        - Este valor está siempre donde el valor de la clase cambia

Árboles de Decisión

- ID3: ¿Cómo manejar atributos continuos?
  - Se discretizan:
    - Proceso:
      - Ordenar los ejemplos de acuerdo al valor continuo de A
      - Tomar cada par de valores contiguos de A con distinto valor de C y hallar el promedio
      - Entre estos candidatos a ser c, escoger el que produzca mayor cantidad de información en este momento
      - Escogido el c, este nuevo atributo compite con los otros discretos para ver si es el elegido
      - Si no es elegido, puede que más adelante salga otro c distinto
        - Es decir, el atributo se crea dinámicamente cada vez que se escoge atributo

Árboles de Decisión

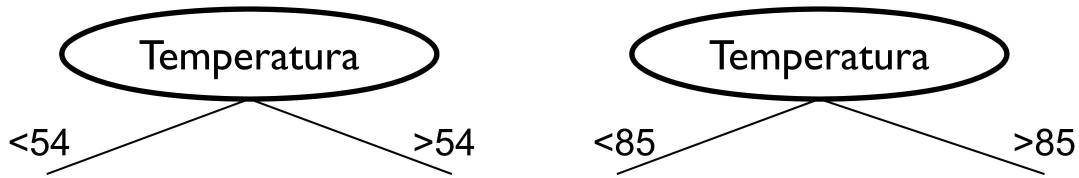
- ID3: ¿Cómo manejar atributos continuos?
  - Ejemplo:

Temperatura	80	48	90	72	40	60
Clase	A	B	B	A	B	A

Ordenado:

Temperatura	40	48	60	72	80	90
Clase	B	B	A	A	A	B

- Candidatos son $c_1 = (48+60)/2 = 54$ y $c_2 = (80+90)/2 = 85$
- Habría que calcular la ganancia al partir



- Y estas dos ganancias tienen que competir contra el resto de atributos

ÁRBOLES DE DECISIÓN

- ID3: ¿Cómo manejar atributos continuos?
  - Ejemplo:

Temperatura	80	48	90	72	40	60
Clase	A	B	B	A	B	A

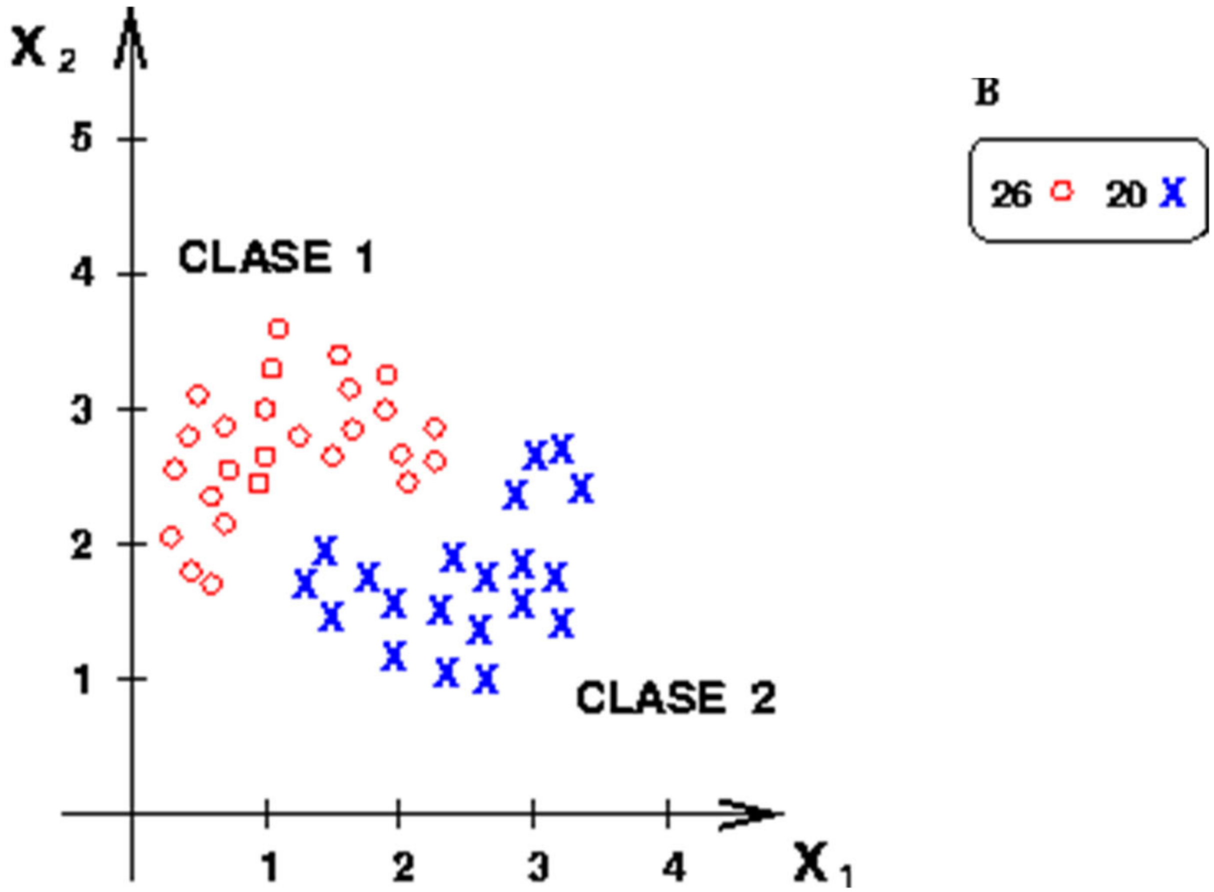
Ordenado:

Temperatura	40	48	60	72	80	90
Clase	B	B	A	A	A	B

- Candidatos son $c_1 = (48+60)/2 = 54$ y $c_2 = (80+90)/2 = 85$
- Otras alternativas:
  - Más de dos intervalos
  - Usar combinaciones lineales de varios atributos: $\alpha A + \beta B > c$ …
  - También, si es un atributo discreto pero con muchos valores
    - Agrupar los valores

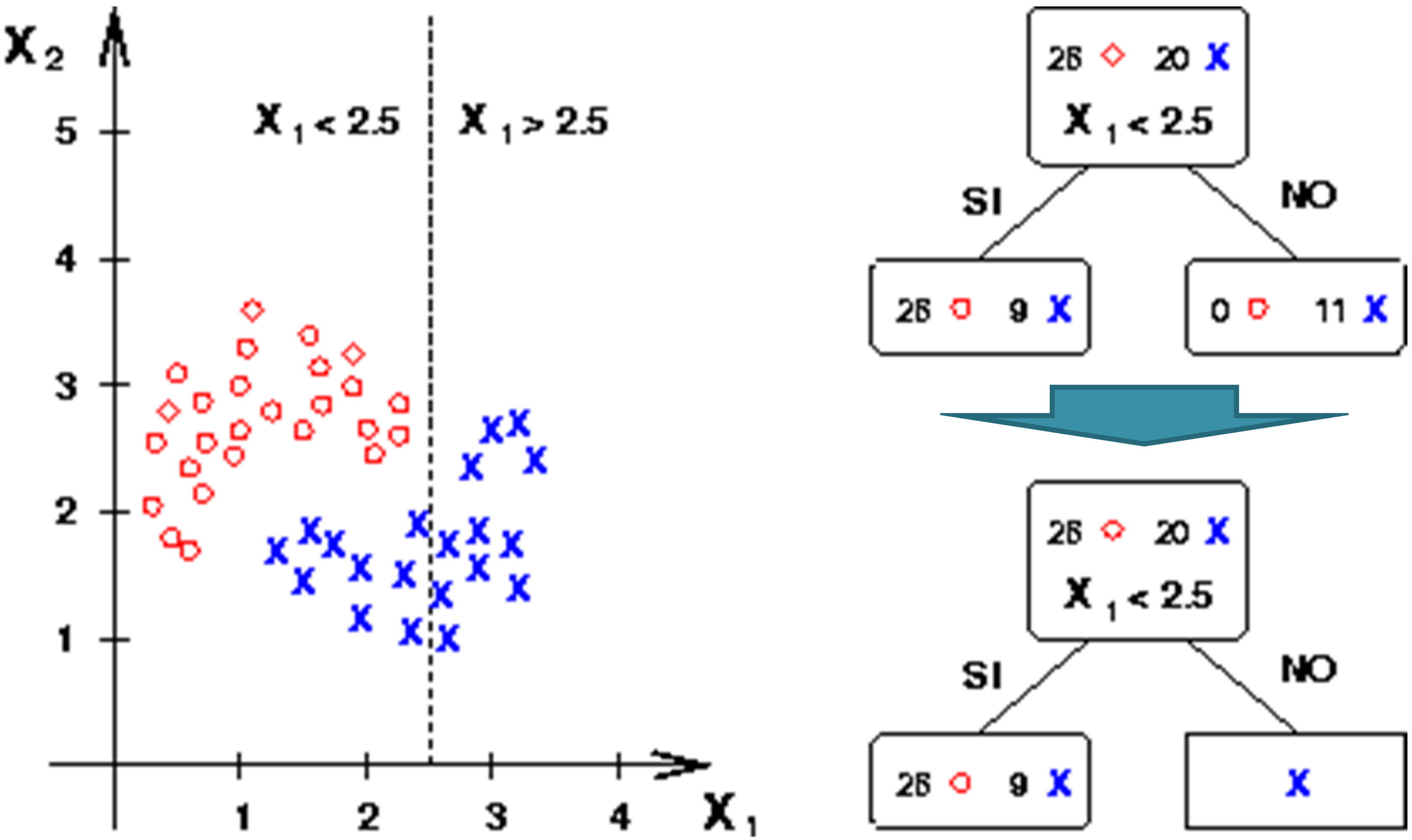
ÁRBOLES DE DECISIÓN

- Ejemplo:
  - Atributos continuos:



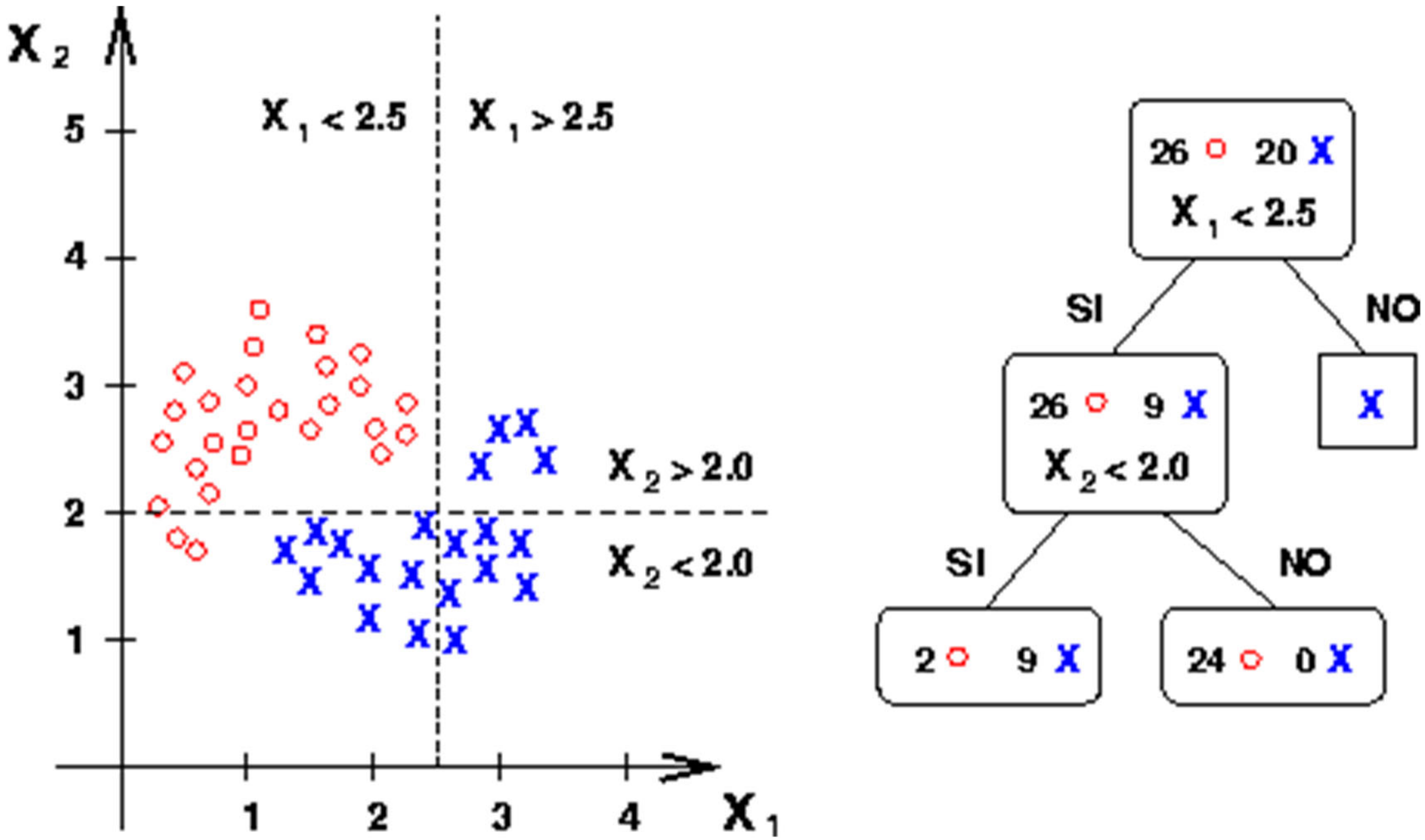
ÁRBOLES DE DECISIÓN

- Ejemplo:



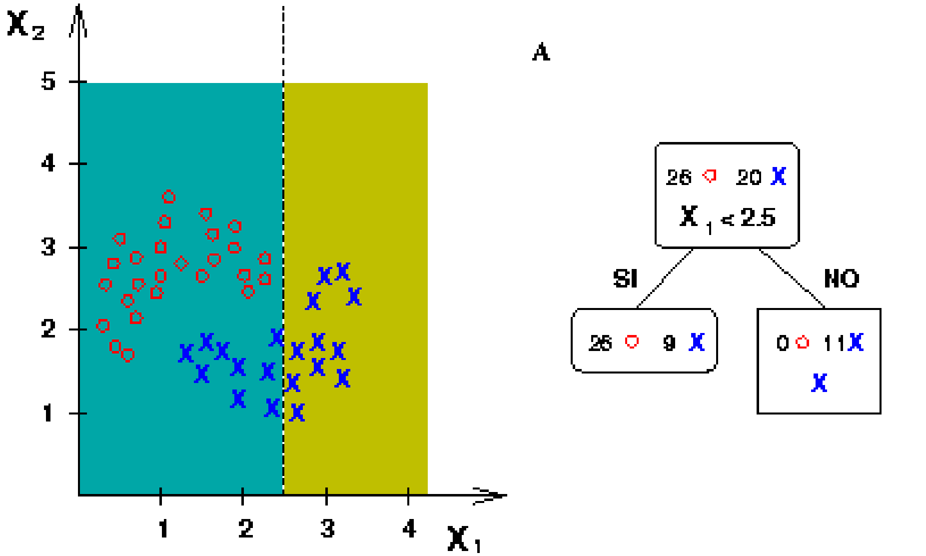
Árboles de Decisión

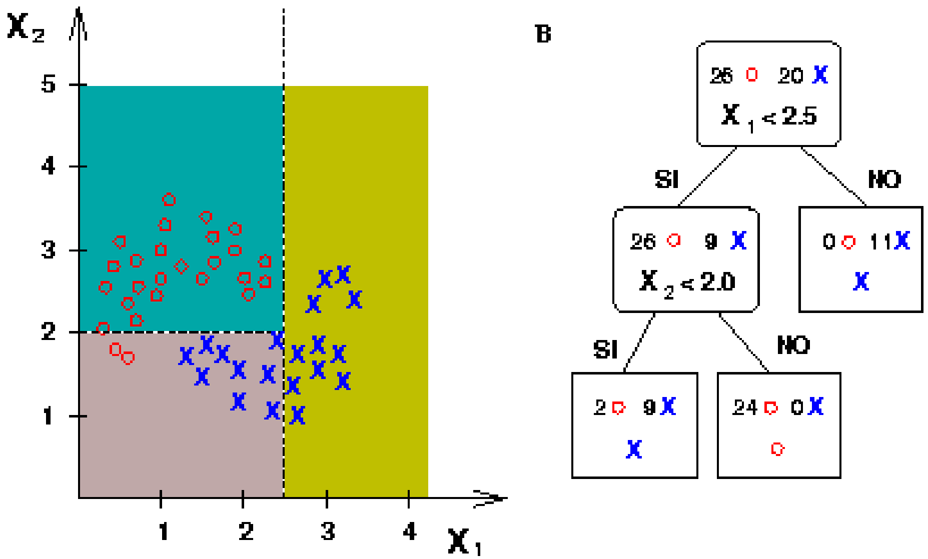
- Ejemplo:



Árboles de Decisión

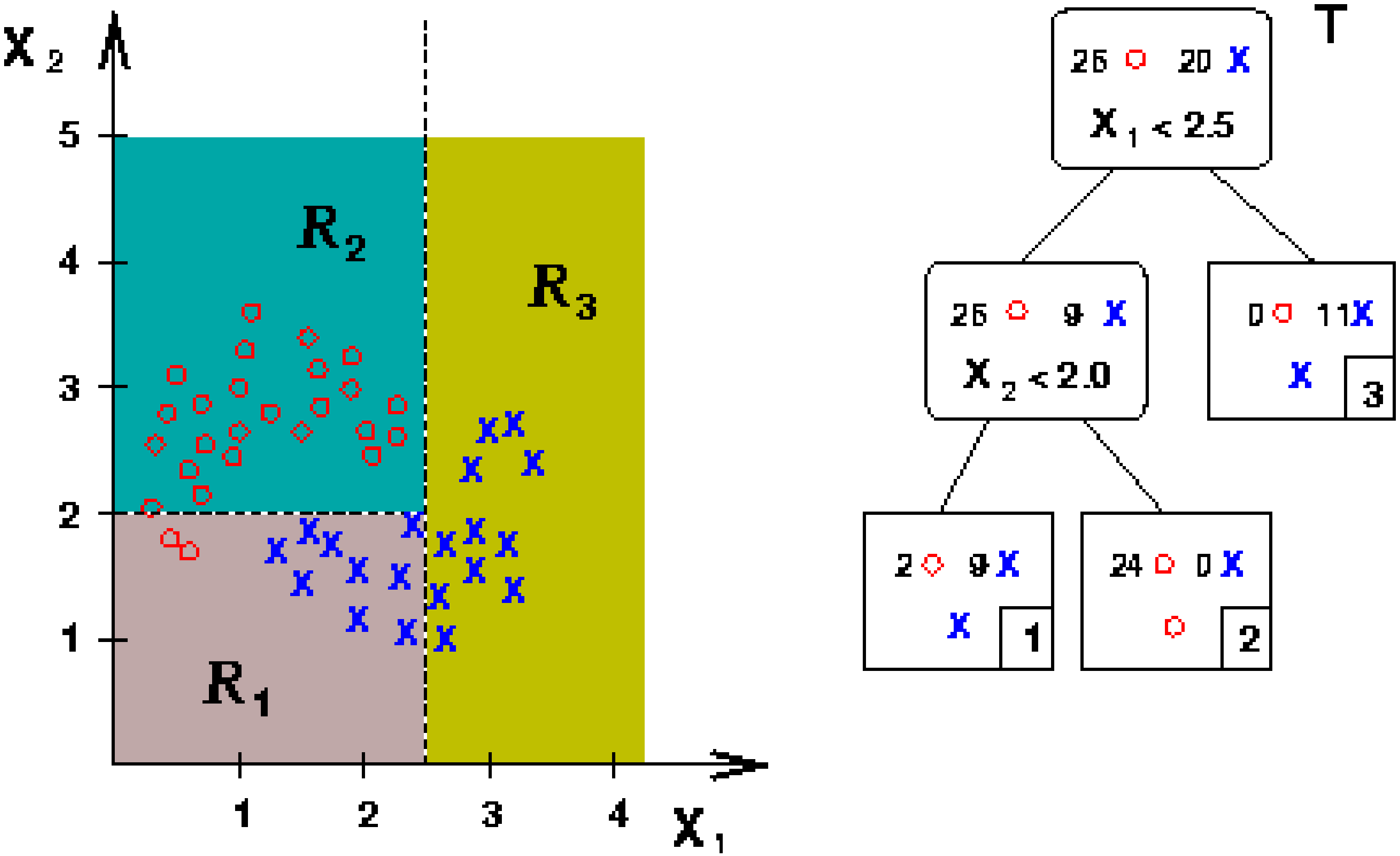
- Ejemplo:
  - Regiones:
    - Paralelepípedos





Árboles de Decisión

- Ejemplo:
  - Cada nodo hoja tiene asociada una región:

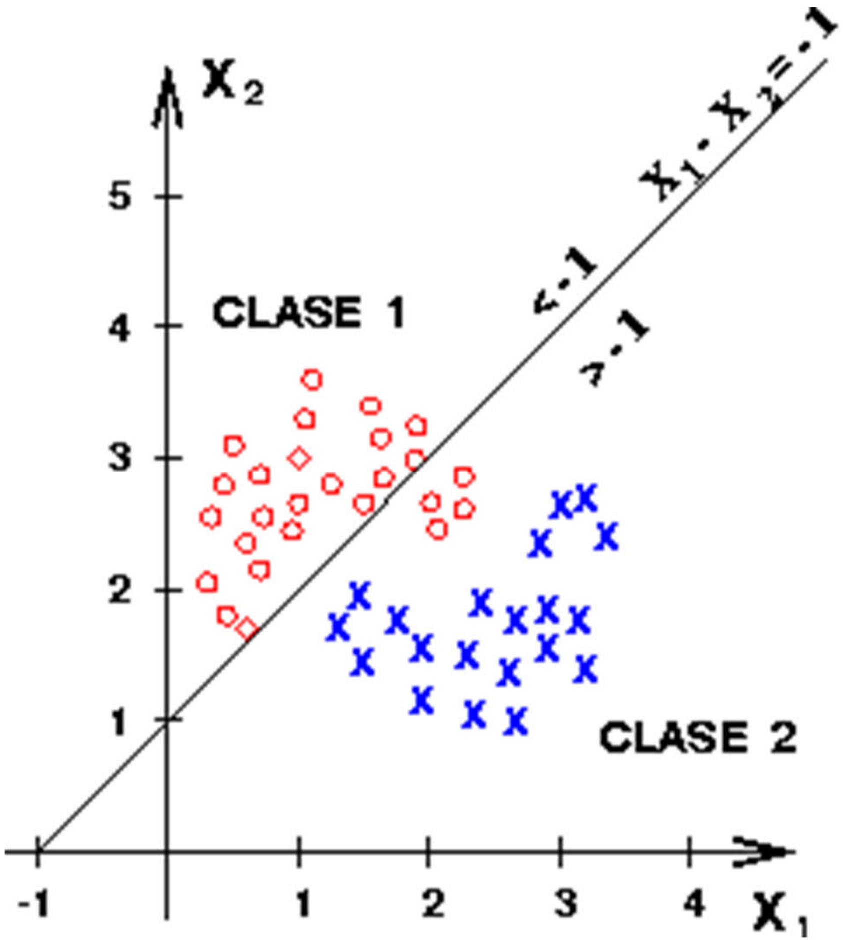


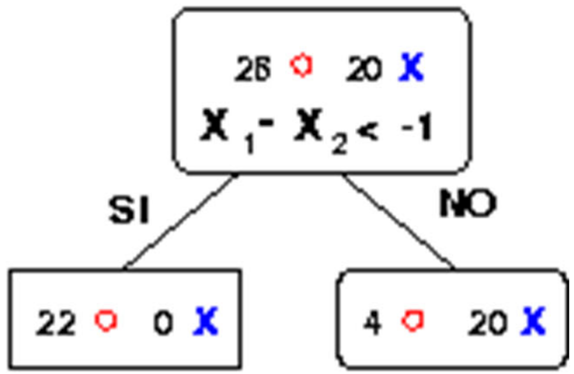
Árboles de Decisión

- Ejemplo:
  - Atributos continuos
    - Permitiendo que una condición sea una combinación lineal de atributos

ÁRBOLES DE DECISIÓN

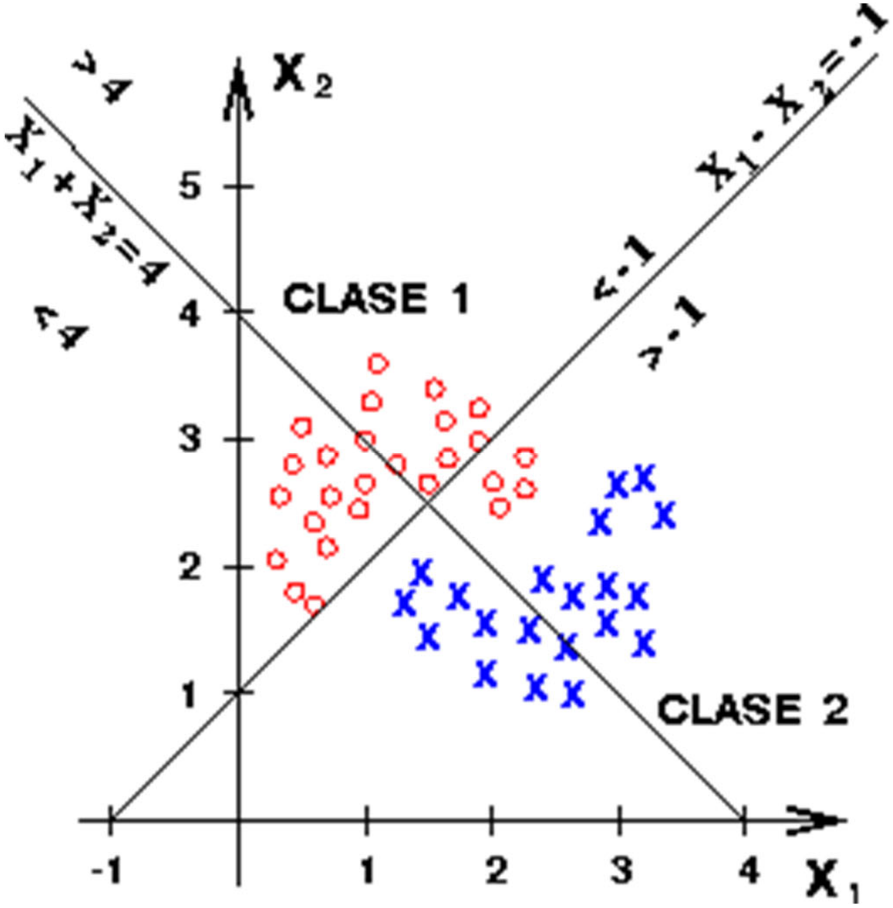
- Ejemplo:

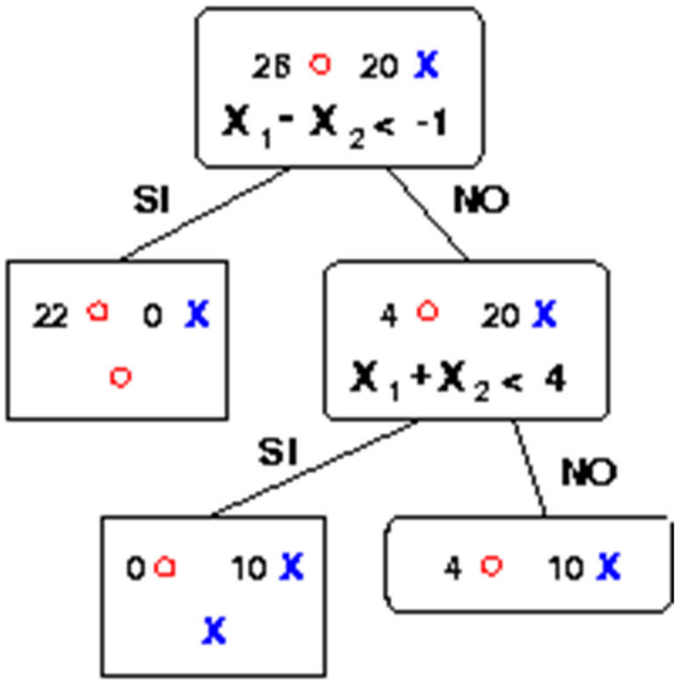




Árboles de Decisión

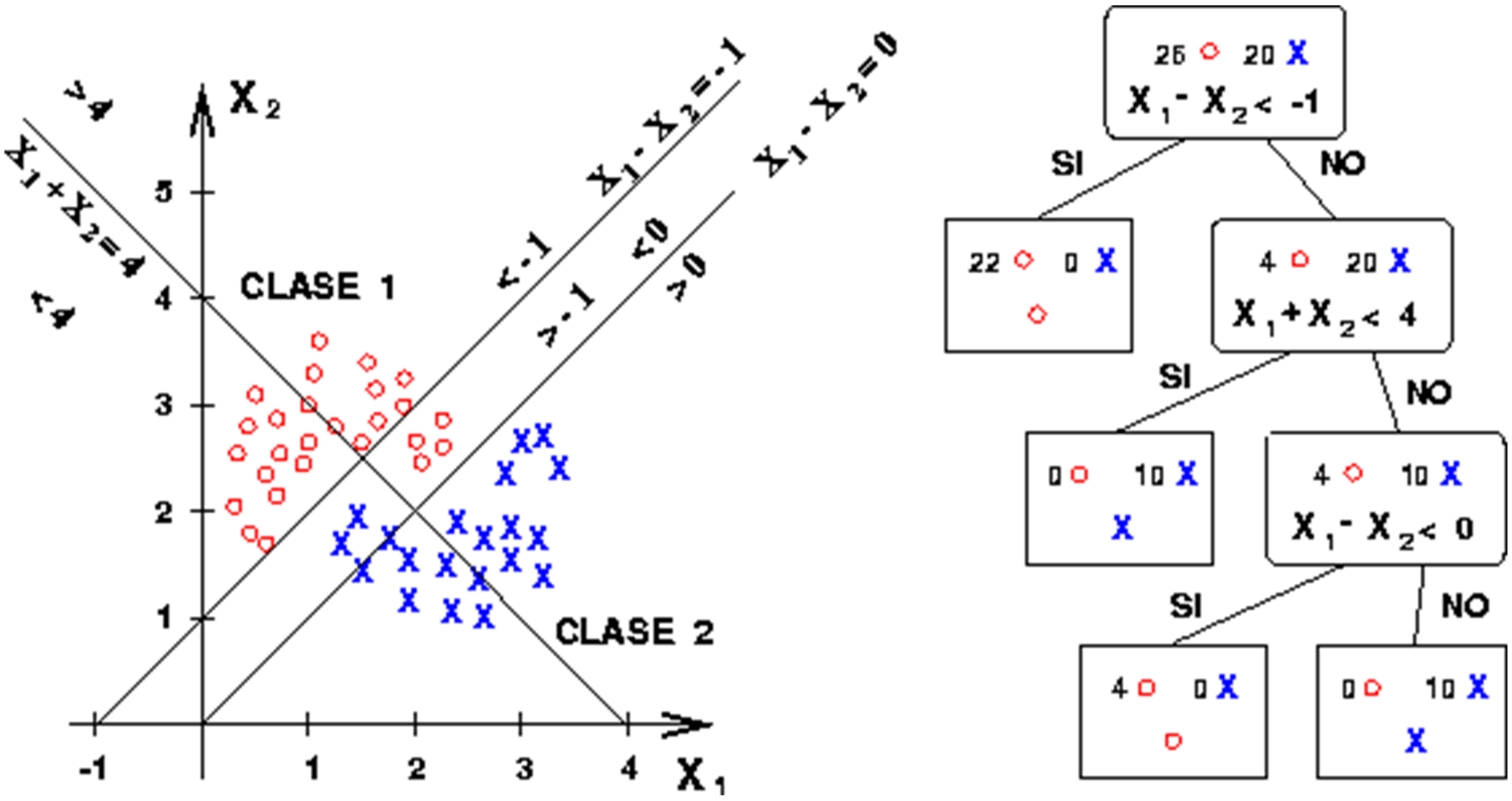
- Ejemplo:





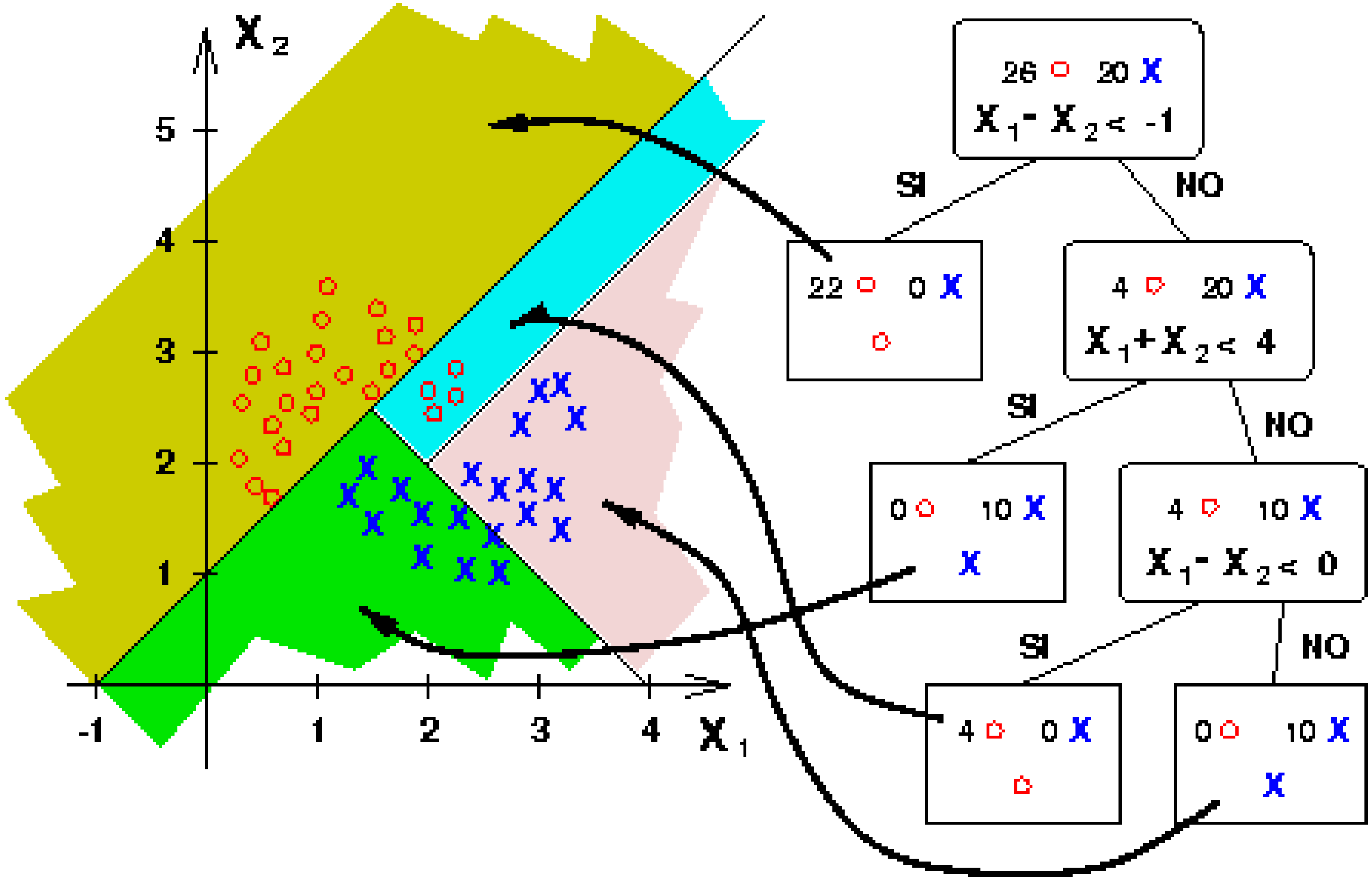
ÁRBOLES DE DECISIÓN

- Ejemplo:



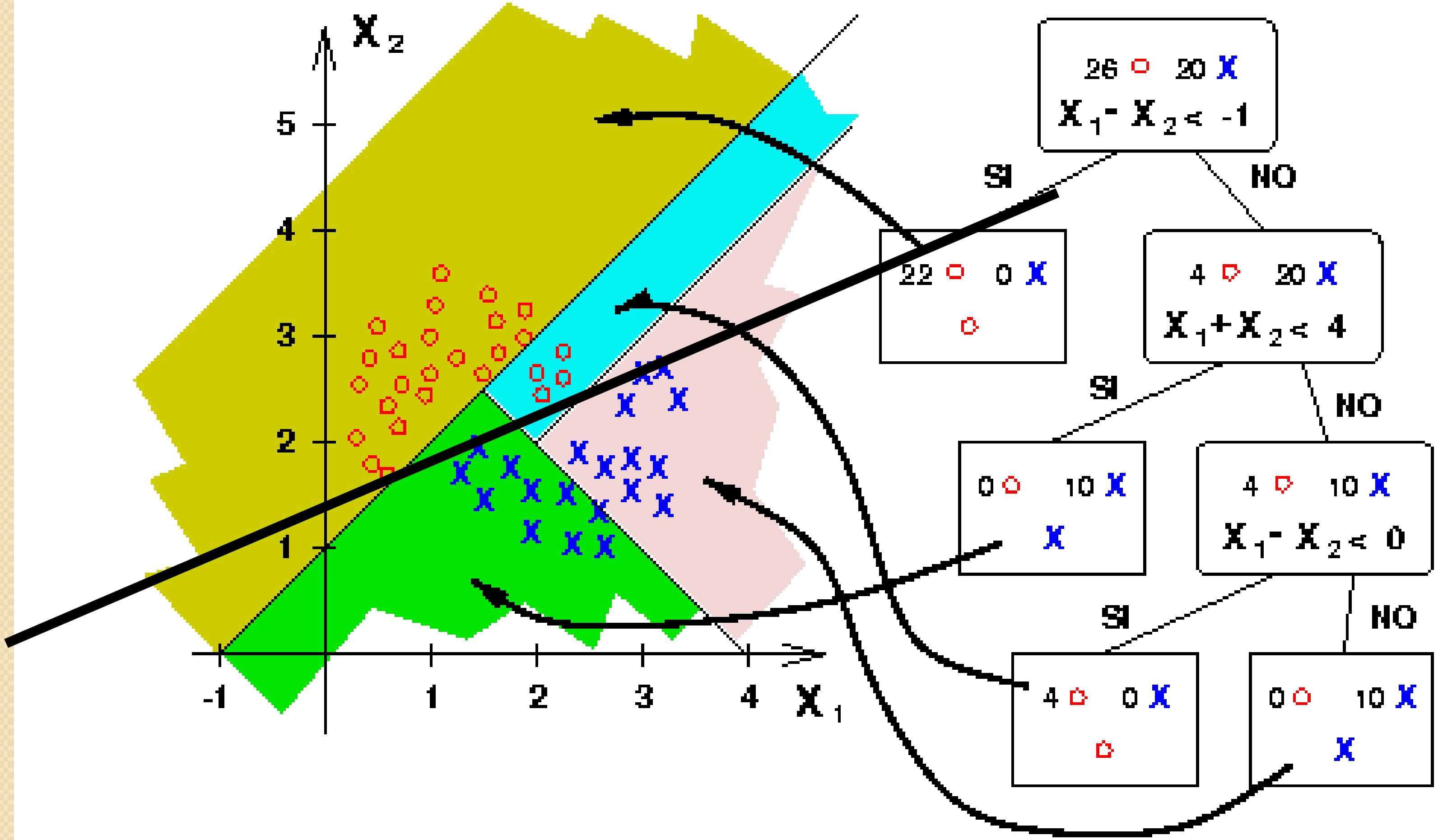
Árboles de Decisión

- Ejemplo:
  - Cada nodo hoja tiene asociada una región



Árboles de Decisión

- Ejemplo:
  - Árbol sobreajustado vs. Clasificación lineal:



ÁRBOLES DE DECISIÓN

- C4.5: (Quinlan, 1993)
  - Algoritmo basado en ID3 con todas las modificaciones anteriores:
    - Permite trabajar con valores continuos para los atributos
    - Escoge atributos usando el ratio de ganancia
      - Elimina la tendencia de ID3 a escoger atributos con muchos valores
    - Incorpora post-poda de reglas:
      - Genera 1 regla por camino
      - Elimina antecedentes de cada regla siempre que se mejore o iguale el error
    - Introduce costes en los atributos
    - Maneja valores desconocidos
    - Ordena las reglas por su error estimado (menos a más)
      - Esta secuencia se considera cuando se clasifiquen instancias

Árboles de Decisión

- CART:
  - *Classification And Regression Trees*
  - Breiman y otros, 1984
  - 4 fases:
    - Construcción del árbol
    - Parada del proceso de crecimiento
      - Se constituye un árbol máximo que sobreajusta la información contenida en la base de datos
    - Podado
      - Haciendo el árbol más sencillo y dejando sólo los nodos más importantes
    - Selección del árbol óptimo con capacidad de generalización

ÁRBOLES DE DECISIÓN

- CART:
  - Cómo escoger un atributo para ser nodo:
    - Se busca una medida de su *pureza*
      - Se escoge el atributo con mayor pureza
    - Un conjunto de patrones es más puro cuando las instancias están menos repartidas entre las clases
      - Entropía menor
      - Mayor certidumbre
    - Debe usarse una función de partición que asegure que la pureza en los nodos hijos sea la máxima
      - Que los hijos sean lo más puros posibles
        - Que las instancias en los hijos tengan más homogeneidad de clases posible

Árboles de Decisión

- CART:
  - Cómo escoger un atributo para ser nodo:
    - En lugar de medirse la pureza en los nodos, se mide la impureza
    - La medida de la la impureza:
      - Debe de ser máxima cuando las instancias de un nodo están divididas equitativamente entre todas las clases
      - Debe de ser 0 cuando las instancias de un nodo son de la misma clase
    - Para medir esta impureza se podría usar:
      - Tasa de instancias mal clasificadas (no se usa)
      - Entropía (ID3, C4.5)
      - Índice de Gini (CART)

ÁRBOLES DE DECISIÓN

- CART:
  - Tasa de fallo como medida de impureza:
    - Se toma como clase el valor de clase más común de todas las instancias que llegan a un nodo
    - Se clasifican todas las instancias con ese valor de clase y se computa la tasa de fallo
    - Esa tasa de fallo será la medida de la impureza
      - Es igual a 0 si todas las instancias pertenecen a la misma clase
      - Es máxima cuando están igualmente repartidas entre las clases

      - Sin embargo, no utilizar esta medida de la impureza

ÁRBOLES DE DECISIÓN

- CART:
  - Por qué no usar la tasa de fallo como medida de impureza:
    - Ejemplo:

Posible partición

40 A	60 A
60 A	40 B

Posible partición

- 200 instancias: valor más común: A
- Tasa de fallo al clasificar todas como A: 40/200 = 0.2

Árboles de Decisión

- CART:
  - Por qué no usar la tasa de fallo como medida de impureza:
    - Ejemplo:
      - Tasa de fallo en cada parte después de dividir en cualquiera de las dos particiones:
        - 0/100 = 0
        - 40/100 = 0.4
      - Tasa de fallo como consecuencia de la división:
        - Suma de esas tasas de fallo ponderadas por el número de nodos de cada una
        - (100/200)*(0) + (100/200)*(40/100) = 0.2
        - No mejora la tasa de fallo sin dividir (0.2)
        - Sin embargo, con ambas divisiones sí que se mejoraría
    - Puede ocurrir que ninguna partición mejore esta tasa
      - Ninguna de las dos la mejora, pero ambas en conjunto sí lo hacen

Árboles de Decisión

- CART:
  - Índice de Gini:
    - El índice de Gini indica cuál es la probabilidad de clasificar erróneamente una observación
      - Medida de la impureza en un nodo
    - El índice de Gini es una medida de cuántas veces un elemento escogido aleatoriamente del conjunto sería incorrectamente etiquetado si fuera etiquetado aleatoriamente de acuerdo con la distribución de etiquetas en el subconjunto.
    - Es decir, para un conjunto de patrones, el índice de Gini mide:
      - Se escoge un patrón al azar
      - Probabilidad de equivocarse al etiquetarlo de acuerdo a la distribución de etiquetas del conjunto
      - Se realiza este proceso para los patrones de cada clase

Árboles de Decisión

- CART:
  - Índice de Gini:
    - Se calcula sumando la probabilidad de escoger cada instancia multiplicado por la probabilidad de equivocarse etiquetándola
      - Por ejemplo, si se tienen 3 clases, con 10 patrones:
        - A, A, A, A, B, B, B, C, C, C
      - Para la clase A:
        - Probabilidad de escoger un patrón de la clase A: 4/10
        - Probabilidad de equivocarse al clasificar un patrón de la clase A: 1 – 4/10
      - Para la clase B:
        - Probabilidad de escoger un patrón de la clase B: 3/10
        - Probabilidad de equivocarse al clasificar un patrón de la clase B: 1 – 3/10
      - Para la clase C:
        - Probabilidad de escoger un patrón de la clase C: 3/10
        - Probabilidad de equivocarse al clasificar un patrón de la clase C: 1 – 3/10

Árboles de Decisión

- CART:
  - Índice de Gini:
    - Se calcula sumando la probabilidad de escoger cada instancia multiplicado por la probabilidad de equivocarse etiquetándola
      - Para el nodo p:

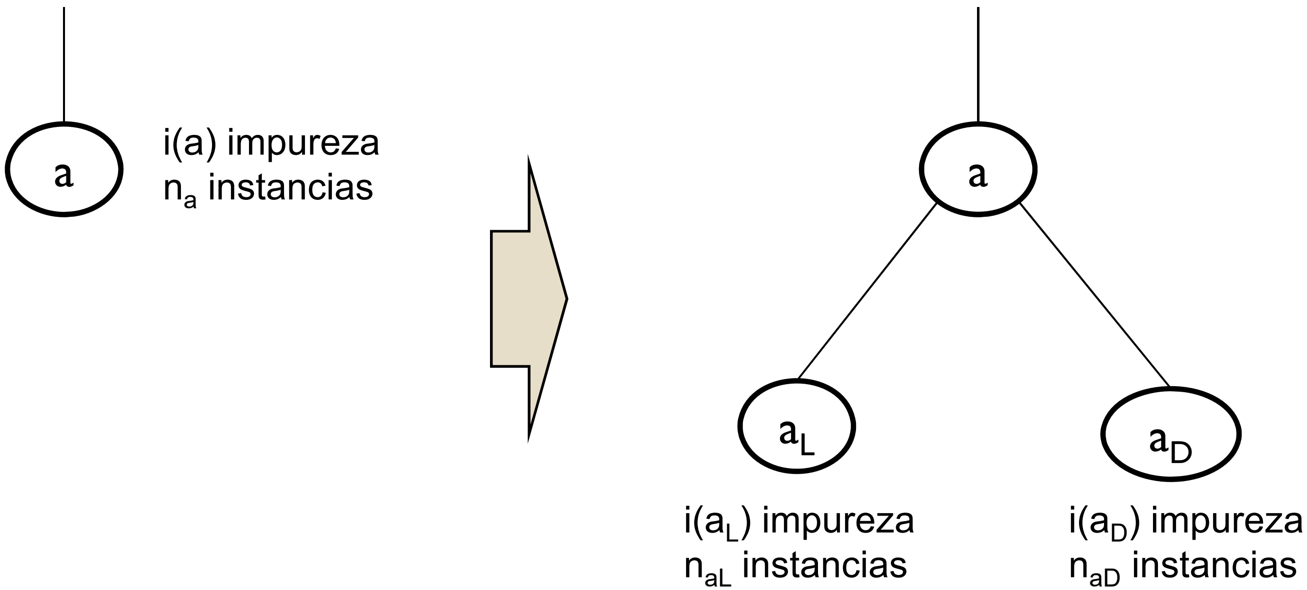
$$i(p) = \sum_{i=1}^{m} f_i(1 - f_i) = \sum_{i=1}^{m} (f_i - f_i^2) = \sum_{i=1}^{m} f_i - \sum_{i=1}^{m} f_i^2 = 1 - \sum_{i=1}^{m} f_i^2$$

donde f_i es la proporción de instancias de la clase i, de m posibles clases,

- Toma el valor de 0 cuando todas las instancias son de la misma clase

ÁRBOLES DE DECISIÓN

- CART:
  - Si el nodo a se divide en dos partes, a_L y a_D, con un número de instancias n_{aL} y n_{aD},



Árboles de Decisión

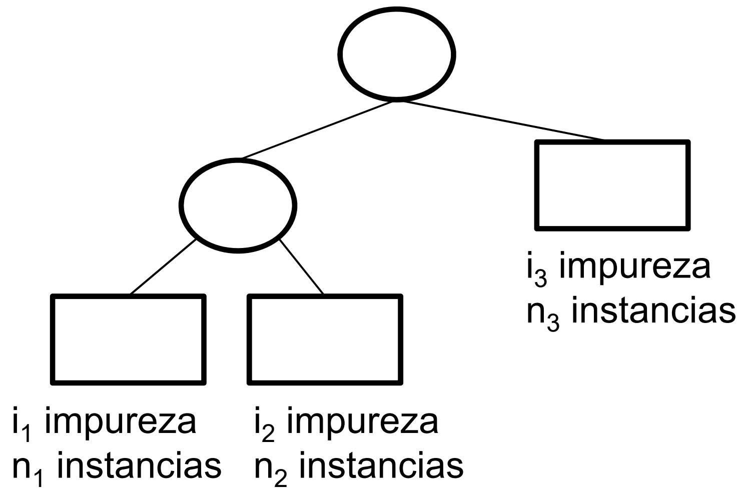
- CART:
  - Si el nodo a se divide en dos partes, a_L y a_D, con un número de instancias n_{aL} y n_{aD}, la reducción de la impureza al realizar esta división es

$$\Delta(a) = i(a) - \frac{n_{aL}}{n_a} i(a_L) - \frac{n_{aD}}{n_a} i(a_D)$$

    - La reducción de la impureza siempre es positiva
    - Se escoge la división que da la máxima reducción de la impureza
      - La que más hace decrementar el índice de Gini

Árboles de Decisión

- CART:
  - Impureza de un árbol es la suma en todos los nodos terminales de la impureza de ese nodo multiplicado por la proporción de casos que llegan a ese nodo en el árbol
  - Ejemplo:
    - Árbol A con n instancias:

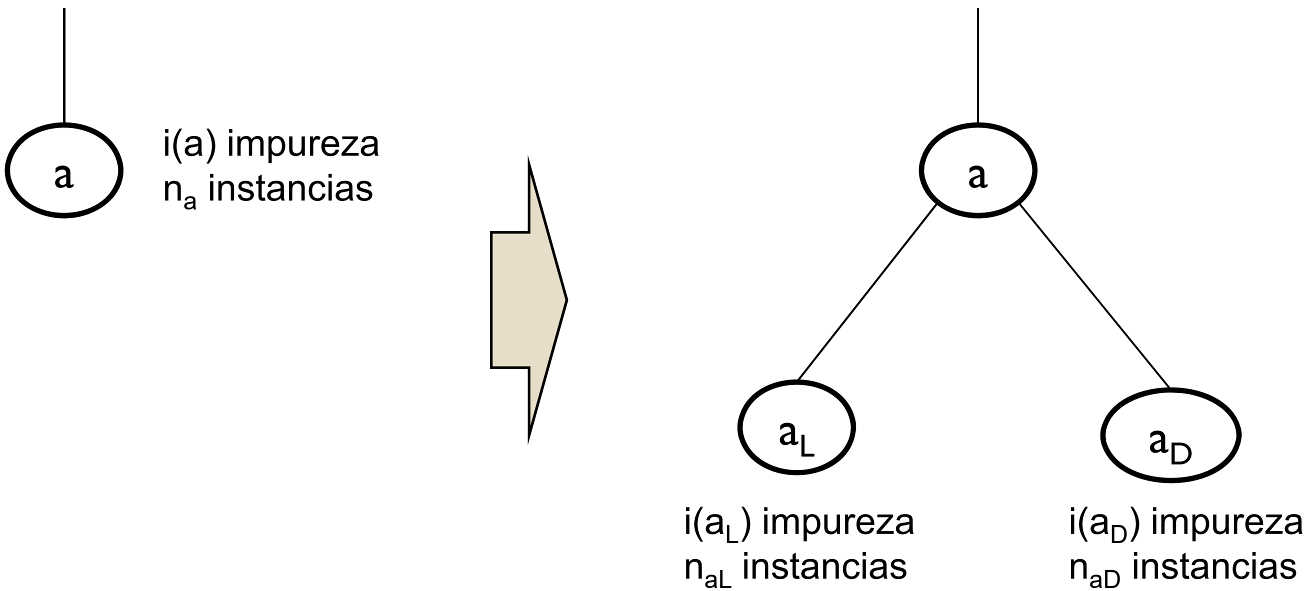


    - Impureza:

$$i(A) = \frac{n_1}{n} i_1 + \frac{n_2}{n} i_2 + \frac{n_3}{n} i_3$$

ÁRBOLES DE DECISIÓN

- CART:
  - Impureza de un árbol es la suma en todos los nodos terminales de la impureza de ese nodo multiplicado por la proporción de casos que llegan a ese nodo en el árbol
  - Si se tiene un árbol A, con a un nodo terminal del árbol, que se puede dividir en dos nodos a_L y a_D



Árboles de Decisión

- CART:
  - Impureza de un árbol es la suma en todos los nodos terminales de la impureza de ese nodo multiplicado por la proporción de casos que llegan a ese nodo en el árbol
  - Si se tiene un árbol A, con a un nodo terminal del árbol, que se puede dividir en dos nodos a_L y a_D
    - La impureza del nuevo árbol A' al dividir ese nodo será

$$i(A') = \underbrace{i(A)}_{\text{Impureza del árbol}} + \underbrace{\frac{n_{aL}}{n} i(a_L) + \frac{n_{aD}}{n} i(a_D)}_{\text{Impureza relativa a los nuevos nodos terminales}} - \underbrace{\frac{n_a}{n} i(a)}_{\text{Impureza relativa al antiguo nodo terminal a}}$$

Árboles de Decisión

- CART:
  - Por lo tanto, la reducción de la impureza de un árbol al dividir un nodo a sería:

$$i(A)-i(A')=\frac{n_a}{n}i(a)-\frac{n_{aL}}{n}i(a_L)-\frac{n_{aD}}{n}i(a_D)=$$

$$=\frac{n_a}{n}i(a)-\frac{n_a}{n_a}\frac{n_{aL}}{n}i(a_L)-\frac{n_a}{n_a}\frac{n_{aD}}{n}i(a_D)=$$

$$=\frac{n_a}{n}i(a)-\frac{n_a}{n}\frac{n_{aL}}{n_a}i(a_L)-\frac{n_a}{n}\frac{n_{aD}}{n_a}i(a_D)=$$

$$=\frac{n_a}{n}\left[i(a)-\frac{n_{aL}}{n_a}i(a_L)-\frac{n_{aD}}{n_a}i(a_D)\right]=\frac{n_a}{n}\Delta(a)$$

    - Es decir, escoger la división sobre a que maximiza la reducción de impureza del árbol es equivalente a escoger la división que maximice la reducción de impureza del nodo

Árboles de Decisión

- CART:
  - El crecimiento de un árbol continúa hasta que se cumple el criterio que declara un nodo como terminal:
    - Si es puro
      - Sólo hay una clase
    - Si no hay divisiones admisibles
      - Por ejemplo, todos los individuos del nodo toman valores idénticos para las variables explicativas
        - Pero distintas clases
    - Se podría añadir una condición sobre el tamaño mínimo del nodo
      - Por ejemplo, si el nodo contiene menos de 5 individuos
    - Se ha fijado un límite externo de la profundidad (número de niveles máximo) del crecimiento del árbol.

Árboles de Decisión

- CART:
  - El árbol que se ha generado de esta forma clasifica correctamente los registros utilizados en su proceso de aprendizaje
    - Sobreajuste
    - Se pierde capacidad de generalización
  - Poda:
    - El método CART prefiere construir en una primera etapa un árbol muy grande.
    - En una segunda etapa, se poda el árbol de tal manera que se obtenga una sucesión de subárboles anidados.
    - Para esta segunda etapa, se utiliza la **muestra-base-test**
      - Conjunto de test

ÁRBOLES DE DECISIÓN

- CART:
  - Sea A_{max} el árbol construido
    - La rama A_a del árbol A_{max} está formada por el nodo a y todos sus descendientes
    - Podar una rama A_a consiste en cortar todos los descendientes de a (dejando a)
  - El árbol máximo está construido a partir de la **muestra base**
    - Conjunto de entrenamiento

Árboles de Decisión

- CART:
  - El árbol complejo que se ha creado debe simplificarse para que alcance esta capacidad de generalización.
  - Se utiliza un método de podado del árbol.
  - El procedimiento asegura que sólo se retiran los nodos que incrementan muy poco la precisión del árbol.
  - Se utiliza una medida de coste-complejidad
    - Coste: coste de mala clasificación
    - Esta medida combina los criterios de precisión frente a complejidad en el número de nodos y velocidad de procesamiento
    - Se busca el árbol que obtiene menor valor en este parámetro
  - Los árboles más sencillos (podados con este criterio) aseguran una mayor capacidad de generalización.

Árboles de Decisión

- CART:
  - Breiman y otros proponen construir una sucesión de subárboles anidados por podas sucesivas de A_{max}, tales que el coste estimado de cada subárbol sea el más débil entre los que tienen el mismo número de nodos terminales
    - Es decir, podar sucesivamente A_{max} de tal forma que entre los árboles que tengan el mismo número de hojas (igual complejidad) se escoja el de menor error de clasificación en test
      - Coste-complejidad
    - Se necesita una muestra test para seleccionar el súbarbol de coste-complejidad mínimo
    - Se debe suprimir las ramas menos informativas
      - De tamaño grande, tal que su supresión aumente poco el coste del árbol
      - Reducir mucho la complejidad a costa de aumentar un poco el coste del árbol

Árboles de Decisión

- CART:
  - De todos los árboles podados posibles debe seleccionarse el mejor.
    - Se debe conservar uno de los subárboles entre A_{max} y su raíz:
  - El mejor árbol (árbol solución) será el que consigue menor error en el ajuste de los registros utilizados en su proceso de aprendizaje.
  - Pero esta condición no es suficiente, debe ajustar bien la base de datos utilizada en su aprendizaje, pero también debe ajustar registros no empleados en esta fase
    - Se va a seleccionar con la muestra test, buscando el árbol que tenga el coste más débil cuando se estima el coste con la muestra test
      - Coste más débil: menos errores de clasificación

Árboles de Decisión

- CART:
  - Para conseguir este objetivo de escoger el árbol de coste más débil en la muestra test hay diversos métodos
    - El método más común, que no precisa un conjunto de prueba independiente, se denomina **validación cruzada**.
    - La validación cruzada es un método de remuestreo que aprovecha el total de la información disponible en la base de datos sin prescindir de una parte de sus registros.

Árboles de Decisión

- MARS:
  - *Multivariate Adaptive Regression Splines* (Friedman, 1991)
  - Generalización continua de CART
  - Funciones base son un producto de funciones splines
  - Una variable puede ser elegida varias veces
  - Los test en los nodos pueden ser multivariantes (con combinaciones lineales de variables)

Árboles de Decisión

- ¿Para qué problemas son apropiados?
  - Instancias representadas como pares atributo-valor (atributos discretos o reales)
  - Robustez a errores en los datos (de entrenamiento), tanto en los atributos como en la variable respuesta
  - Puede haber datos perdidos (*missing*) para algunas instancias en los valores de los atributos
  - Dominios complejos donde no existe una clara separación lineal
  - Función objetivo toma valores discretos (clasificación)
    - Pero se puede extender a valores continuos
      - **Árboles de regresión**

Árboles de Decisión

- Ventajas principales del los árboles de decisión frente a otras técnicas de AA:
  - Fácilmente interpretables
    - Ventaja frente a otros modelos como RR.NN.AA. o SVM
  - Pueden ser ejecutados aunque falte algún valor en algún atributo
  - Pueden trabajar con entradas categóricas
    - No es necesario codificarlas
  - Permiten trabajar con variables con costes diferentes
    - No es necesario conocer el valor de todos los atributos
    - Conocer el valor de los atributos de alto coste sólo si es necesario
      - Estarán en la parte de abajo del árbol

Árboles de Decisión

- Ventajas principales del los árboles de decisión frente a otras técnicas de AA:
  - Ofrece información sobre las importancia de las variables
    - Más importantes: en la parte superior del árbol
    - Menos importantes: en la parte inferior del árbol
  - Dividen el espacio de búsqueda de forma recursiva
    - Pueden realizar regiones de separación complejas limitadas por la complejidad del árbol
    - Permiten trabajar de forma local en cada parte