



**Jesús García Herrero**

## **TÉCNICAS DE INDUCCIÓN-I**

En esta clase se presentan los principios de los algoritmos básicos de inducción de modelos lógicos a partir de datos para aprendizaje de modelos predictivos. Se revisan los algoritmos más importantes, comenzando por el ID3, y la medida de la entropía de la información en la que se sustenta para construir árboles de decisión

El planteamiento de este tipo de algoritmos es heurístico, no es posible evaluar todas las posibles configuraciones porque nos encontraríamos con dificultades de escalabilidad al número de datos y dimensiones de estos datos. Por tanto, es una búsqueda heurística basada en esta medida de la información que permite obtener buenos resultados y escalabilidad a conjuntos de datos grandes.

El algoritmo se ilustra en ejemplos sencillos, presentando las características y limitaciones de esta estrategia.

# Inducción de Clasificadores

## Reglas y Árboles de Clasificación

Jesús García Herrero  
Universidad Carlos III de Madrid



Universidad  
Carlos III de Madrid



# Árboles de decisión. ID3

- Obtener reglas o relaciones que permitan clasificar a partir de los atributos
- Ej. de entrada:

Ejemplo	Sitio de acceso: $A_1$	1ª cantidad gastada: $A_2$	Vivienda: $A_3$	Última compra: $A_4$	Clase
1	1	0	2	Libro	Bueno
2	1	0	1	Disco	Malo
3	1	2	0	Libro	Bueno
4	0	2	1	Libro	Bueno
5	1	1	1	Libro	Malo
6	2	2	1	Libro	Malo

# Clasificación “1R”

- Medio rudimentario de una sola regla
- Selección de atributo de mínimo error total

Atributo	Reglas	Errores	Errores totales
A <sub>1</sub>	0 → Bueno 1 → Bueno (?) 2 → Malo	0/1 2/4 0/1	2/6
A <sub>2</sub>	0 → Malo (?) 1 → Malo 2 → Bueno	1/2 0/1 1/3	2/6
A <sub>3</sub>	0 → Bueno 1 → Bueno (?) 2 → Bueno	0/1 3/4 0/1	3/6
A <sub>4</sub>	Libro → Bueno Disco → Malo	2/5 0/1	2/6

Árboles de clasificación

# Medida de orden. Entropía

- Seleccionar el atributo que mejor separe (ordene) los ejemplos de acuerdo a las clases. “Divide y vencerás”
- La entropía es una medida de como está ordenado el universo
- La teoría de la información (basada en la entropía) calcula el número de bits (información, preguntas sobre atributos) que hace falta suministrar para conocer la clase a la que pertenece un ejemplo

# Medida de la información

- Entropía: propuesta por Shannon en su Teoría de la Información (1948)
- Dado un conjunto de eventos  $A = \{A_1, A_2, \dots, A_N\}$ , con probabilidades  $\{p_1, p_2, \dots, p_N\}$ , información en un mensaje acerca de estos sucesos:

- Información en el conocimiento de un suceso  $A_i$  (bits)

$$I(A_i) = \log_2 \left( \frac{1}{p_i} \right) = -\log_2(p_i)$$

- Información media de  $A$  (bits)

$$I(A) = \sum_{i=1}^N p_i I(A_i) = -\sum_{i=1}^N p_i \log_2(p_i)$$

**Árboles de clasificación**

# Medida de la información

- Ej.:  $A = \{A_1, A_2, A_3, A_4\}, \{1/4, 1/4, 1/4, 1/4\}$

$$I(A_i) = -\log_2\left(\frac{1}{4}\right) = 2 \text{ bits}$$

$$I(A) = -\frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) = 2 \text{ bits}$$

- $A = \{A_1, A_2, A_3, A_4\}, \{1/5, 1/5, 1/5, 2/5\}$

$$I(A_1) = I(A_2) = I(A_3) = -\log_2\left(\frac{1}{5}\right) = 2.32 \text{ bits};$$

$$I(A_4) = -\log_2\left(\frac{2}{5}\right) = 1.32 \text{ bits};$$

$$I(A) = -\frac{1}{5}\log_2\left(\frac{1}{5}\right) - \frac{1}{5}\log_2\left(\frac{1}{5}\right) - \frac{1}{5}\log_2\left(\frac{1}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 1.92 \text{ bits}$$

**Máxima entropía con sucesos equi-probables**

**Árboles de clasificación**

# Aplicación a la clasificación

Medida de lo que se gana tras usar un atributo  $A_i$

– Información antes de utilizar atributos:

clases:  $\{C_1, \dots, C_M\}$ ; instancias:  $\{n_1, \dots, n_M\}$

$$I = - \sum_{c=1}^M \frac{n_c}{n} \log_2 \left( \frac{n_c}{n} \right)$$

– Tras utilizar atributo  $A_i$

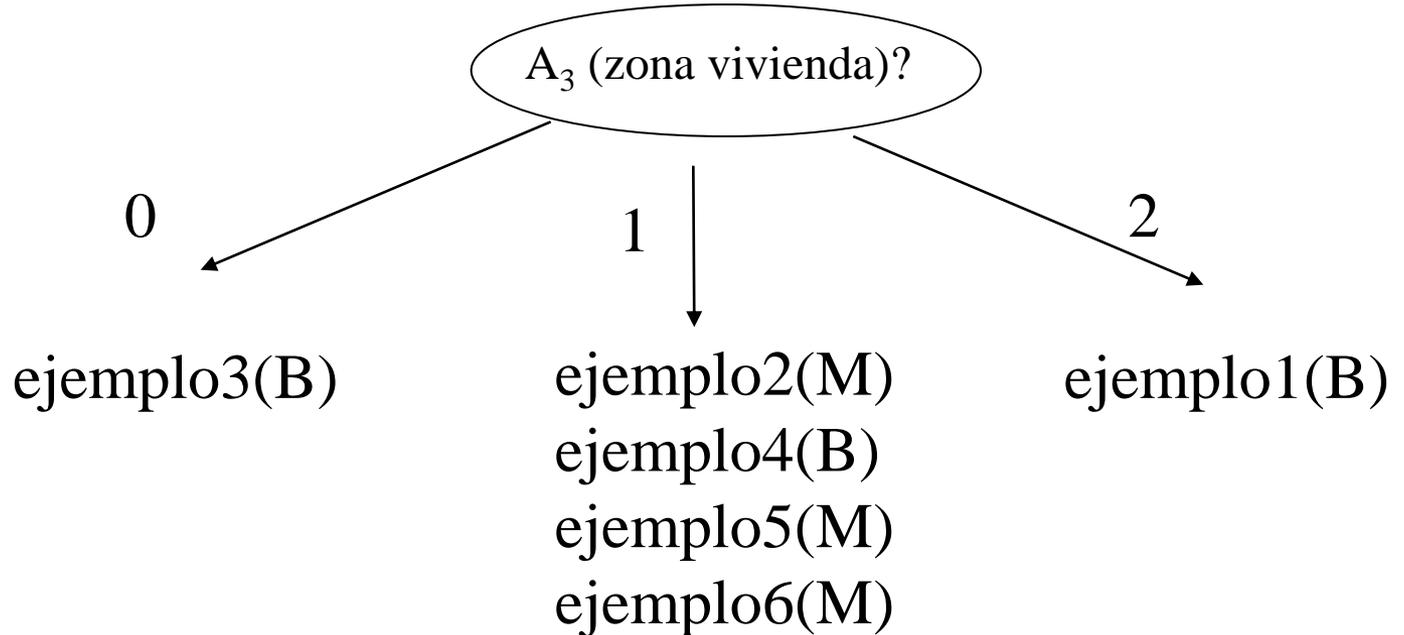
$$I(A_i) = \sum_{j=1}^{nv(A_i)} \frac{n_{ij}}{n} I_{ij}; \quad I_{ij} = - \sum_{k=1}^M \frac{n_{ijk}}{n_{ij}} \log_2 \frac{n_{ijk}}{n_{ij}}$$

$$G(A_i) = I - I(A_i)$$

**Árboles de clasificación**

# Ejemplo

Ejemplo	Sitio de acceso: $A_1$	1ª cantidad gastada: $A_2$	Vivienda: $A_3$	Última compra: $A_4$	Clase
1	1	0	2	Libro	Bueno
2	1	0	1	Disco	Malo
3	1	2	0	Libro	Bueno
4	0	2	1	Libro	Bueno
5	1	1	1	Libro	Malo
6	2	2	1	Libro	Malo



Árboles de clasificación

# Ejemplo

$$I = -\frac{3}{6} \log_2 \left( \frac{3}{6} \right) - \frac{3}{6} \log_2 \left( \frac{3}{6} \right) = 1 \text{ bit}$$

$$I(A_3) = \sum_{j=1}^3 \frac{n_{1j}}{n} I_{1j} = \frac{n_{30}}{n} I_{30} + \frac{n_{31}}{n} I_{31} + \frac{n_{32}}{n} I_{32} = \frac{1}{6} I_{30} + \frac{4}{6} I_{31} + \frac{1}{6} I_{32}$$

$$I_{10} = -\sum_{k=1}^2 \frac{n_{10k}}{n_{10}} \log_2 \frac{n_{30k}}{n_{30}} = -\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1} = 0$$

$$I_{11} = -\sum_{k=1}^2 \frac{n_{11k}}{n_{11}} \log_2 \frac{n_{31k}}{n_{31}} = -\frac{1}{4} \log_2 \frac{1}{4} - \frac{3}{4} \log_2 \frac{3}{4} = 0.81 \text{ bits}$$

$$I_{12} = -\sum_{k=1}^2 \frac{n_{12k}}{n_{12}} \log_2 \frac{n_{32k}}{n_{32}} = -\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1} = 0$$

$$I(A_3) = 0.54; \quad G(A_3) = I - I(A_3) = 0.46 \text{ bits}$$

**Árboles de clasificación**

# Algoritmo ID3 (Quinlan 93)

1. Seleccionar el atributo  $A_i$  que maximice la ganancia  $G(A_i)$
2. Crear un nodo para ese atributo con tantos sucesores como valores tenga
3. Introducir los ejemplos en los sucesores según el valor que tenga el atributo  $A_i$
4. Por cada sucesor:
  - Si sólo hay ejemplos de una clase,  $C_k$   
Entonces etiquetarlo con  $C_k$
  - Si no, llamar al ID3 con una tabla formada por los ejemplos de ese nodo, eliminando la columna del atributo  $A_i$

# Ejemplo

Ejemplo	Sitio de acceso: $A_1$	1ª cantidad gastada: $A_2$	Vivienda: $A_3$	Última compra: $A_4$	Clase
1	1	0	2	Libro	Bueno
2	1	0	1	Disco	Malo
3	1	2	0	Libro	Bueno
4	0	2	1	Libro	Bueno
5	1	1	1	Libro	Malo
6	2	2	1	Libro	Malo

$$I(A_1) = \frac{1}{6} I_{10} + \frac{4}{6} I_{11} + \frac{1}{6} I_{12} = 0.66; G(A_1) = 1,34$$

$$I(A_2) = \frac{2}{6} I_{20} + \frac{1}{6} I_{21} + \frac{3}{6} I_{22} = 0.79; G(A_2) = 1,21$$

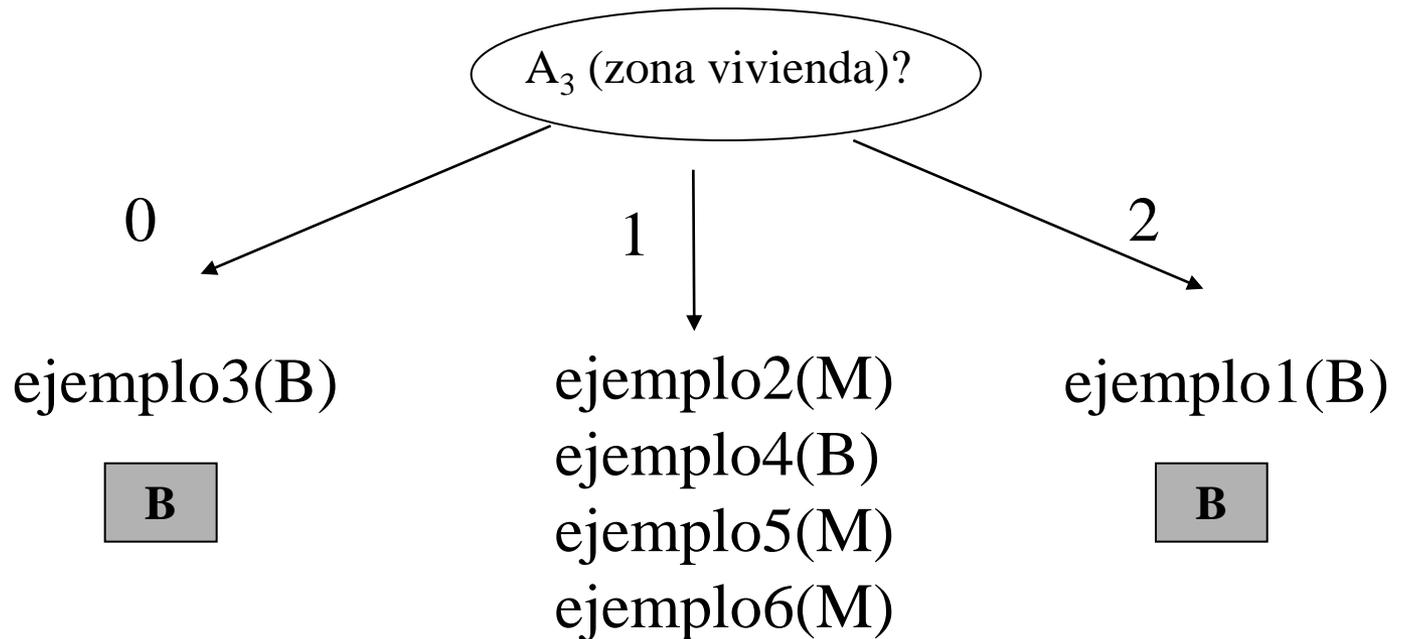
$$I(A_3) = \frac{1}{6} I_{30} + \frac{4}{6} I_{31} + \frac{1}{6} I_{32} = 0.54; G(A_3) = 1,46$$

$$I(A_4) = \frac{1}{6} I_{4Disco} + \frac{5}{6} I_{41Libro} = 0.81; G(A_4) = 1,19$$

**Árboles de clasificación**

# Ejemplo

Ejemplo	Sitio de acceso: $A_1$	1ª cantidad gastada: $A_2$	Vivienda: $A_3$	Última compra: $A_4$	Clase
1	1	0	2	Libro	Bueno
2	1	0	1	Disco	Malo
3	1	2	0	Libro	Bueno
4	0	2	1	Libro	Bueno
5	1	1	1	Libro	Malo
6	2	2	1	Libro	Malo



Árboles de clasificación

# Ejemplo

Ejemplo	Sitio de acceso: $A_1$	1ª cantidad gastada: $A_2$	Última compra: $A_4$	Clase
2	1	0	Disco	Malo
4	0	2	Libro	Bueno
5	1	1	Libro	Malo
6	2	2	Libro	Malo

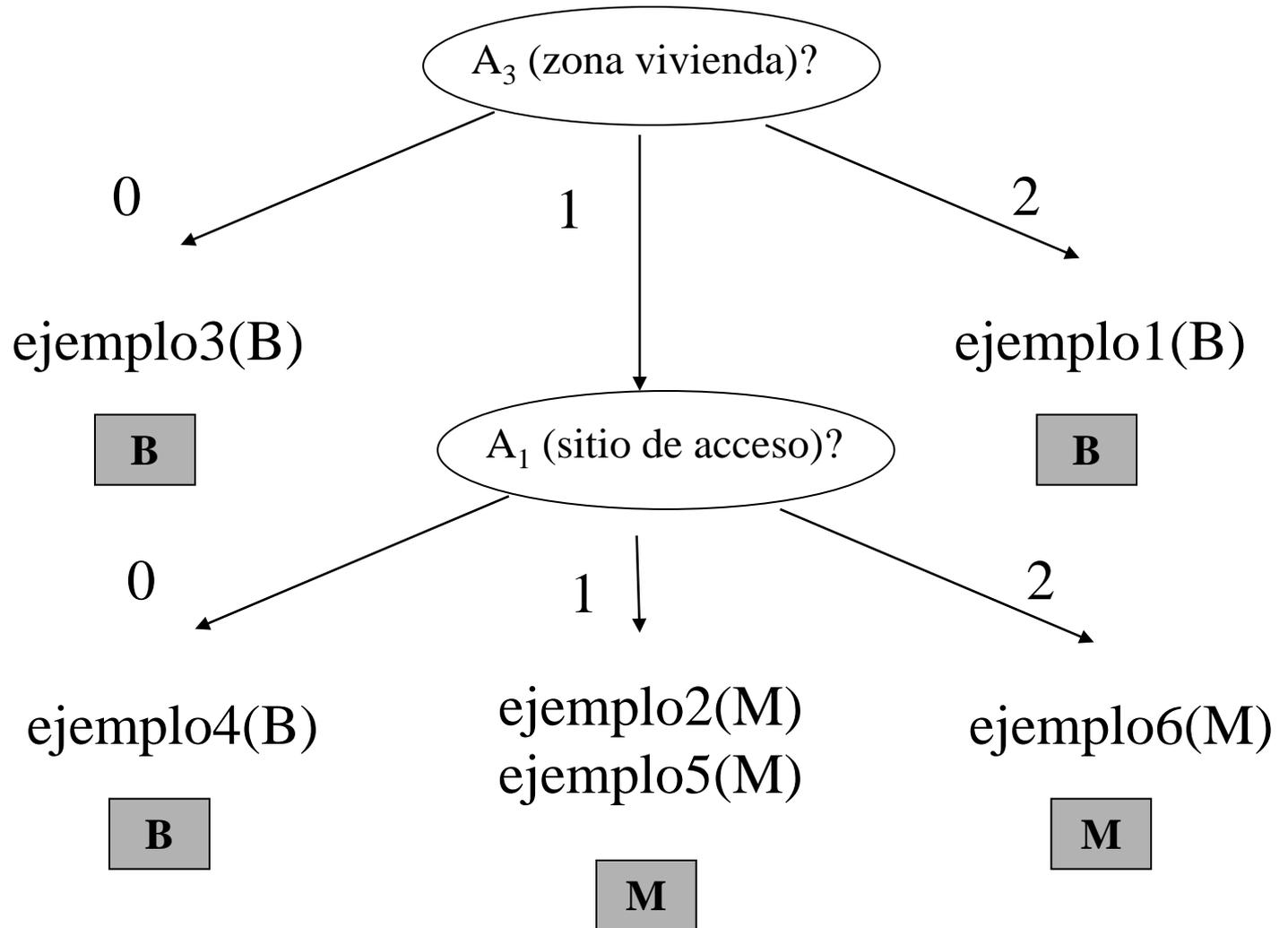
$$I(A_1) = \frac{1}{4} I_{10} + \frac{2}{4} I_{11} + \frac{1}{4} I_{12} = 0; G(A_1) = 2$$

$$I(A_2) = \frac{1}{4} I_{20} + \frac{1}{4} I_{21} + \frac{2}{4} I_{22} = 0.5; G(A_2) = 1,5$$

$$I(A_4) = \frac{1}{4} I_{4Disco} + \frac{3}{4} I_{4Libro} = 0.23; G(A_4) = 1,77$$

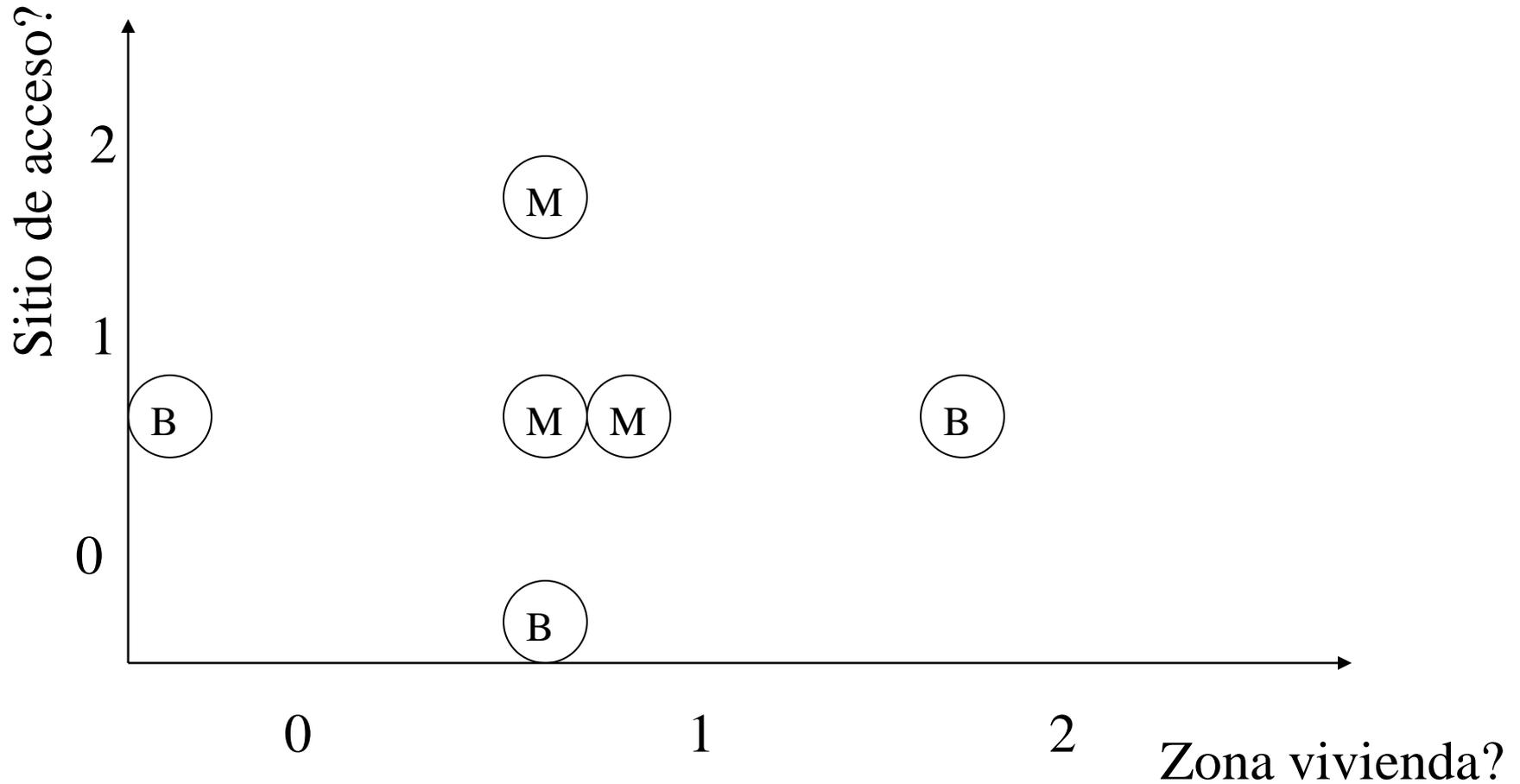
Árboles de clasificación

# Ejemplo



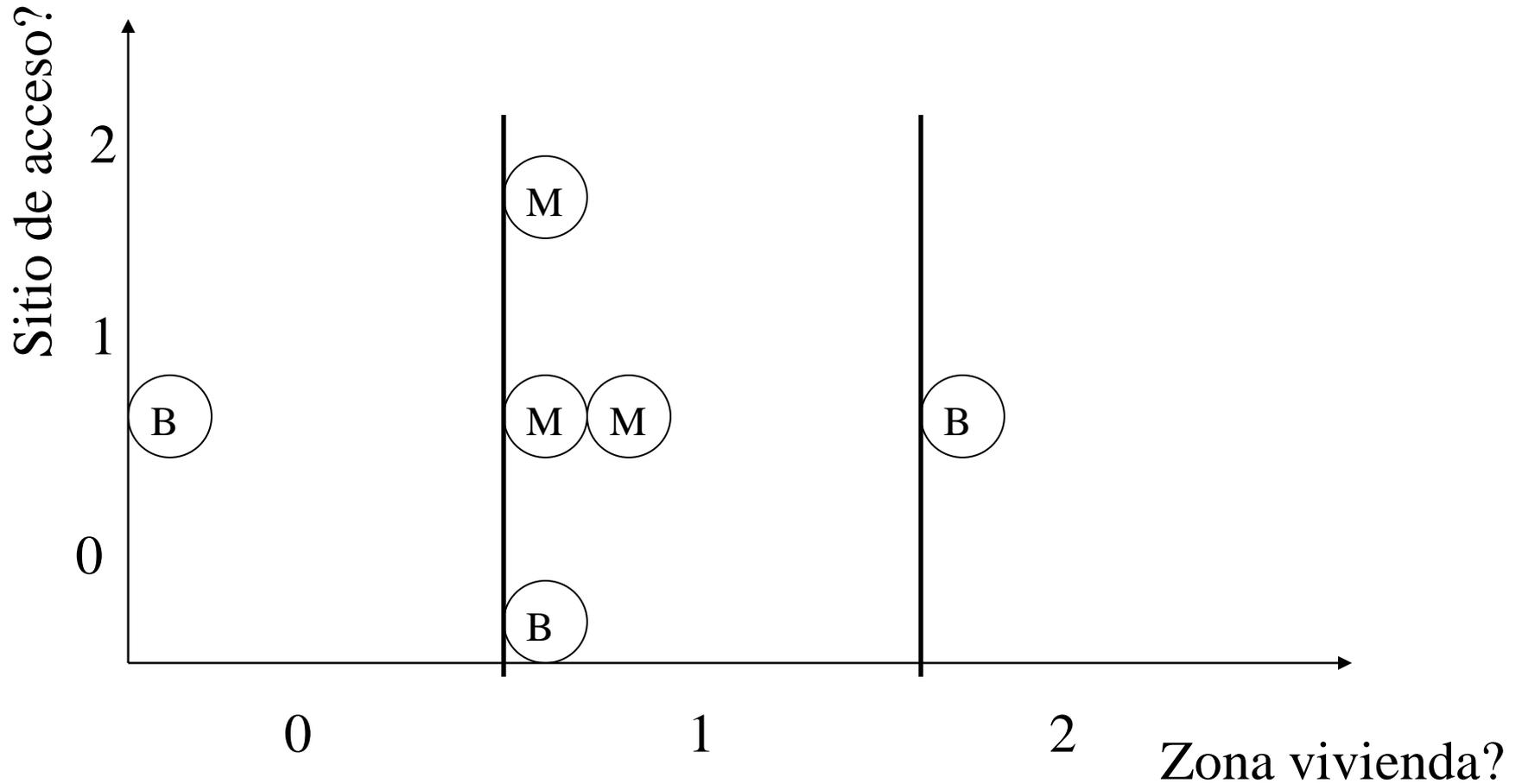
Árboles de clasificación

# Estrategia del ID3



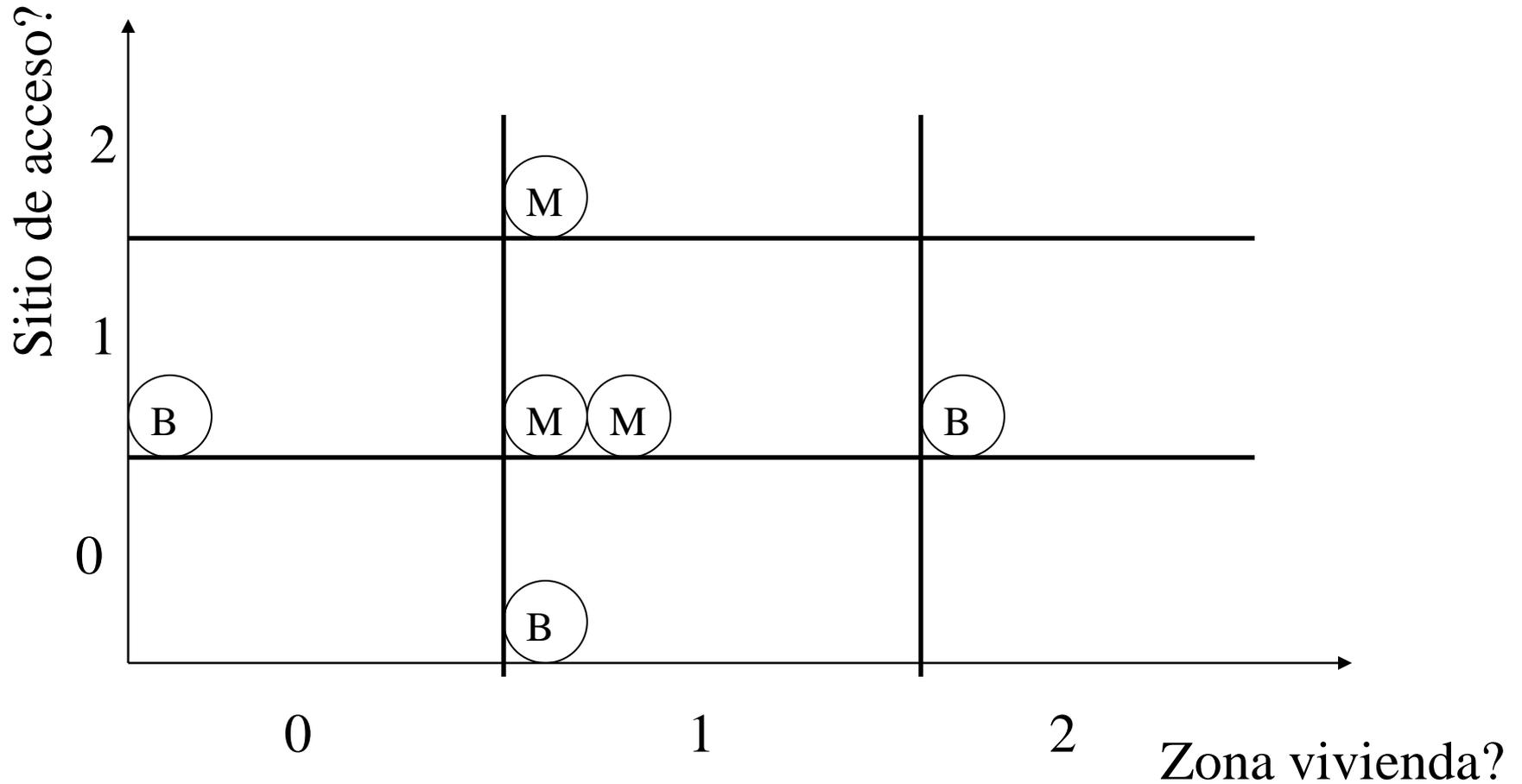
Árboles de clasificación

# Estrategia del ID3



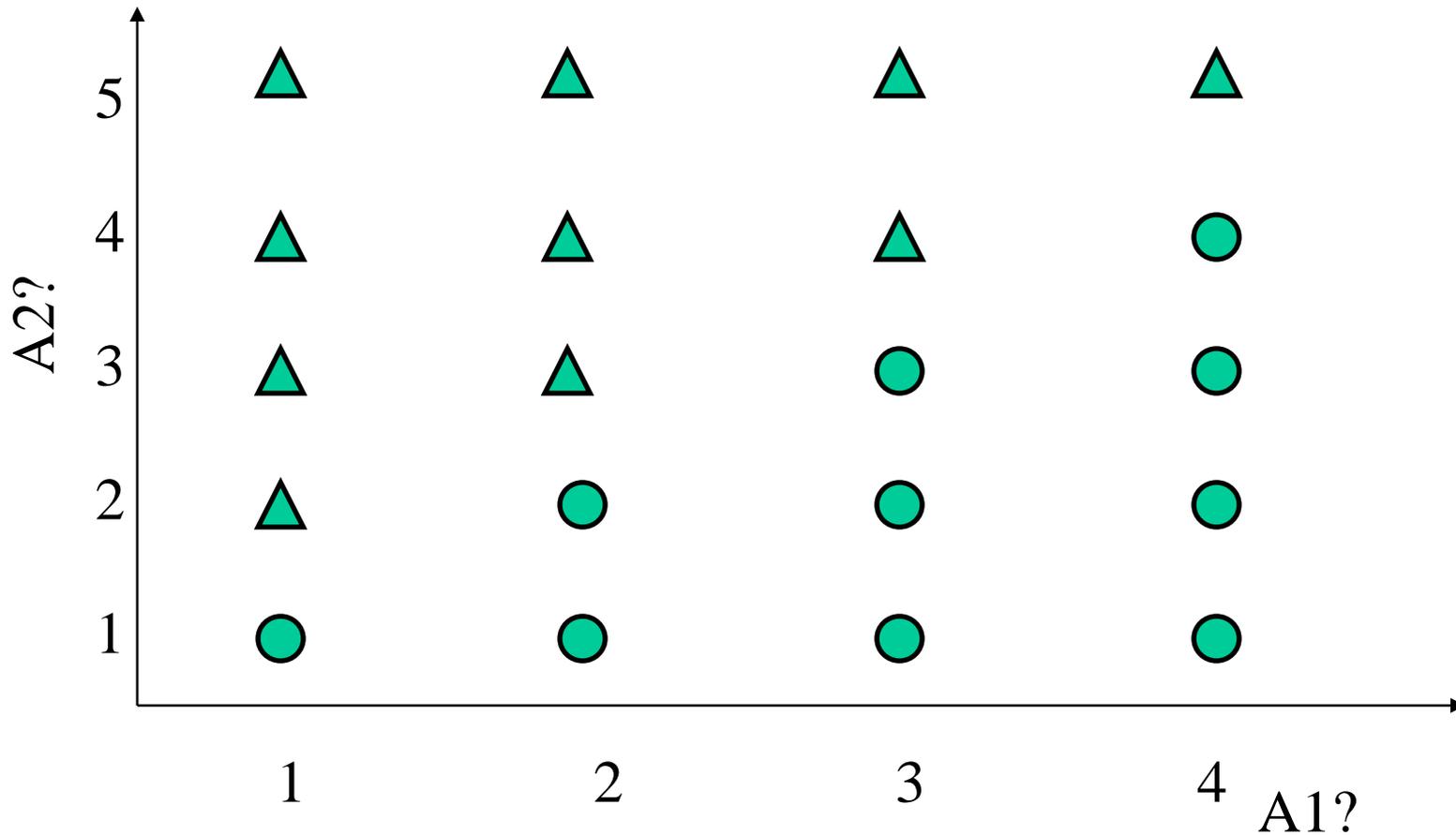
Árboles de clasificación

# Estrategia del ID3



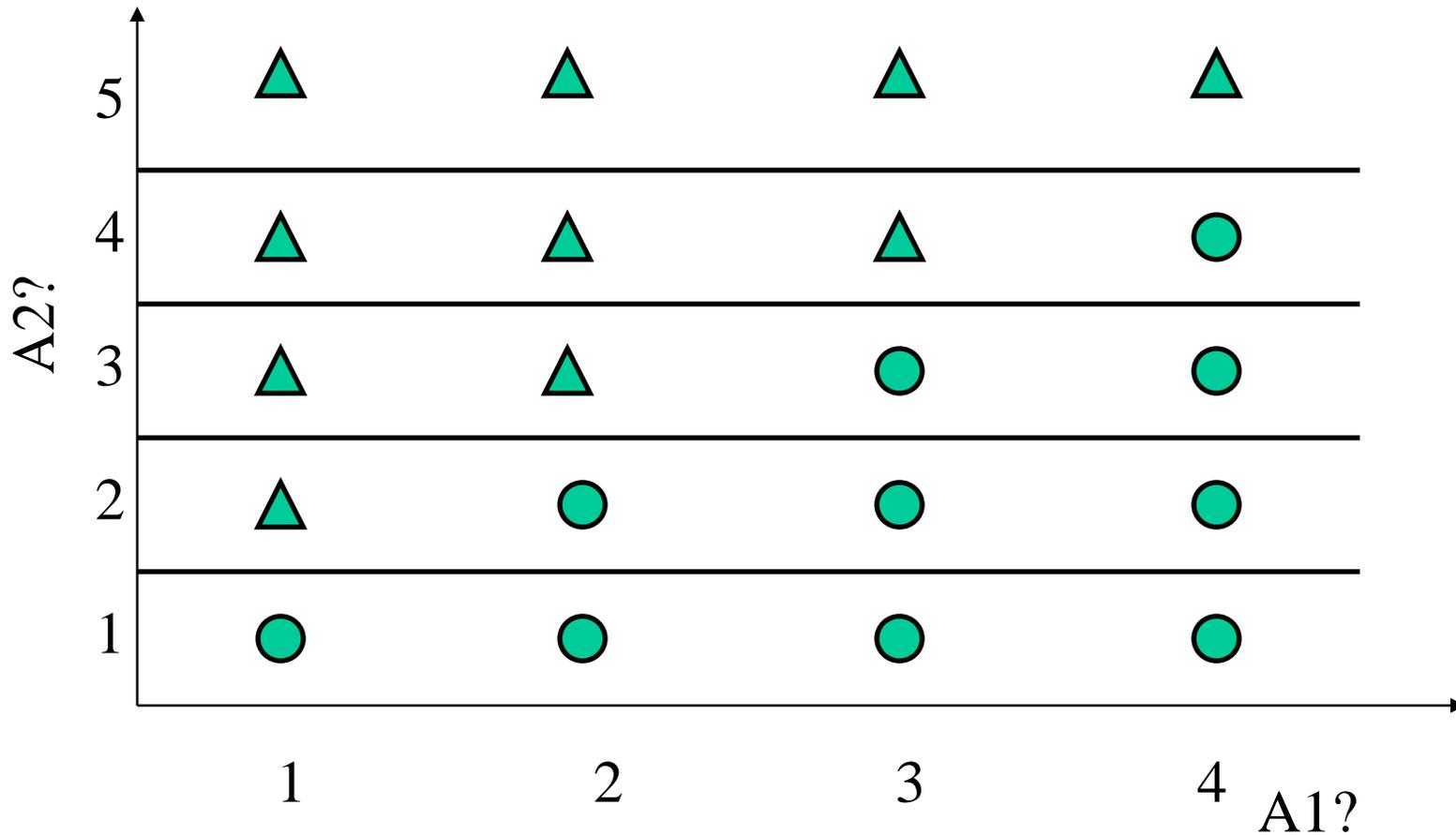
Árboles de clasificación

# Ejemplo con frontera diagonal



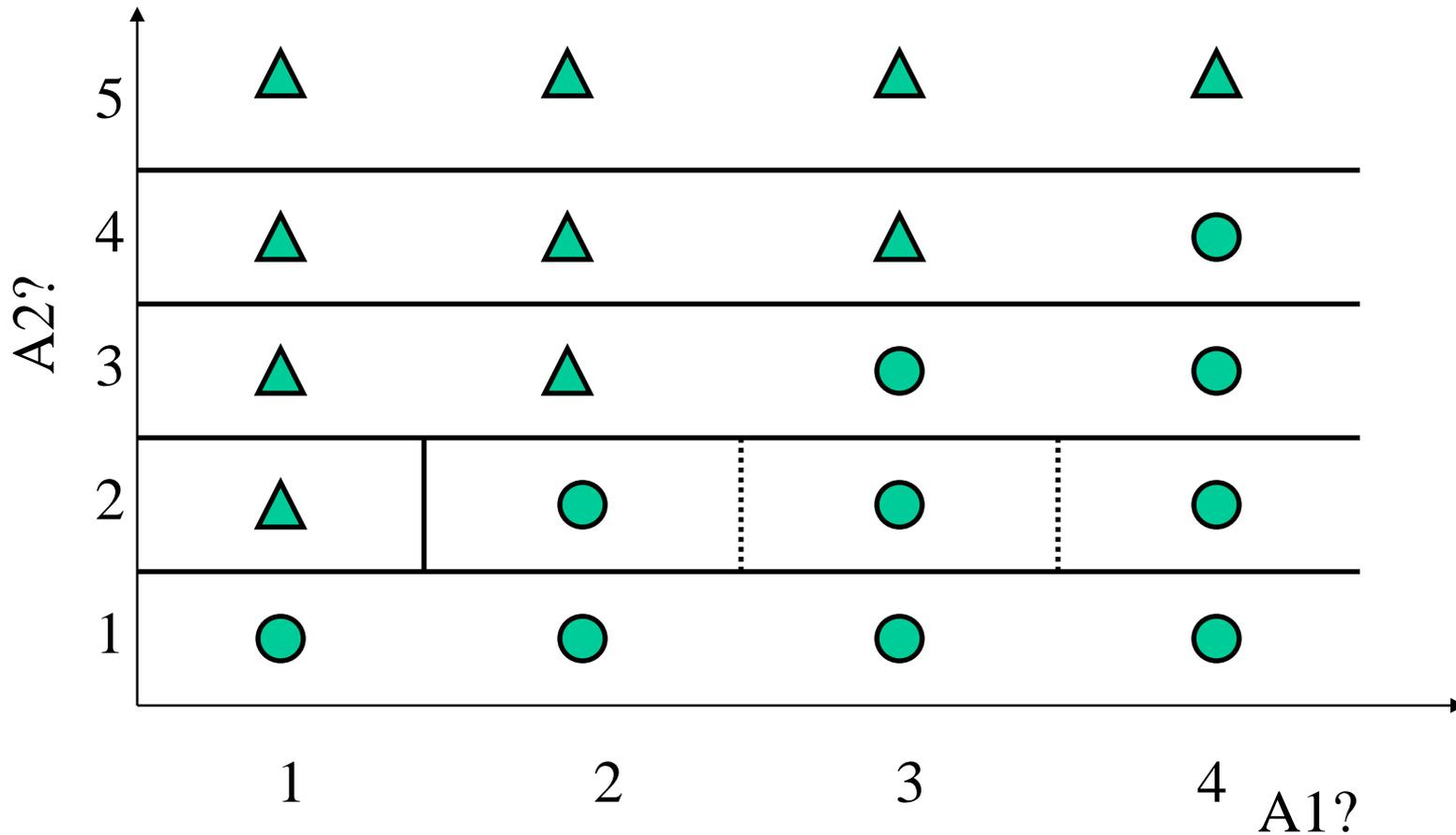
Árboles de clasificación

# Ejemplo con frontera diagonal



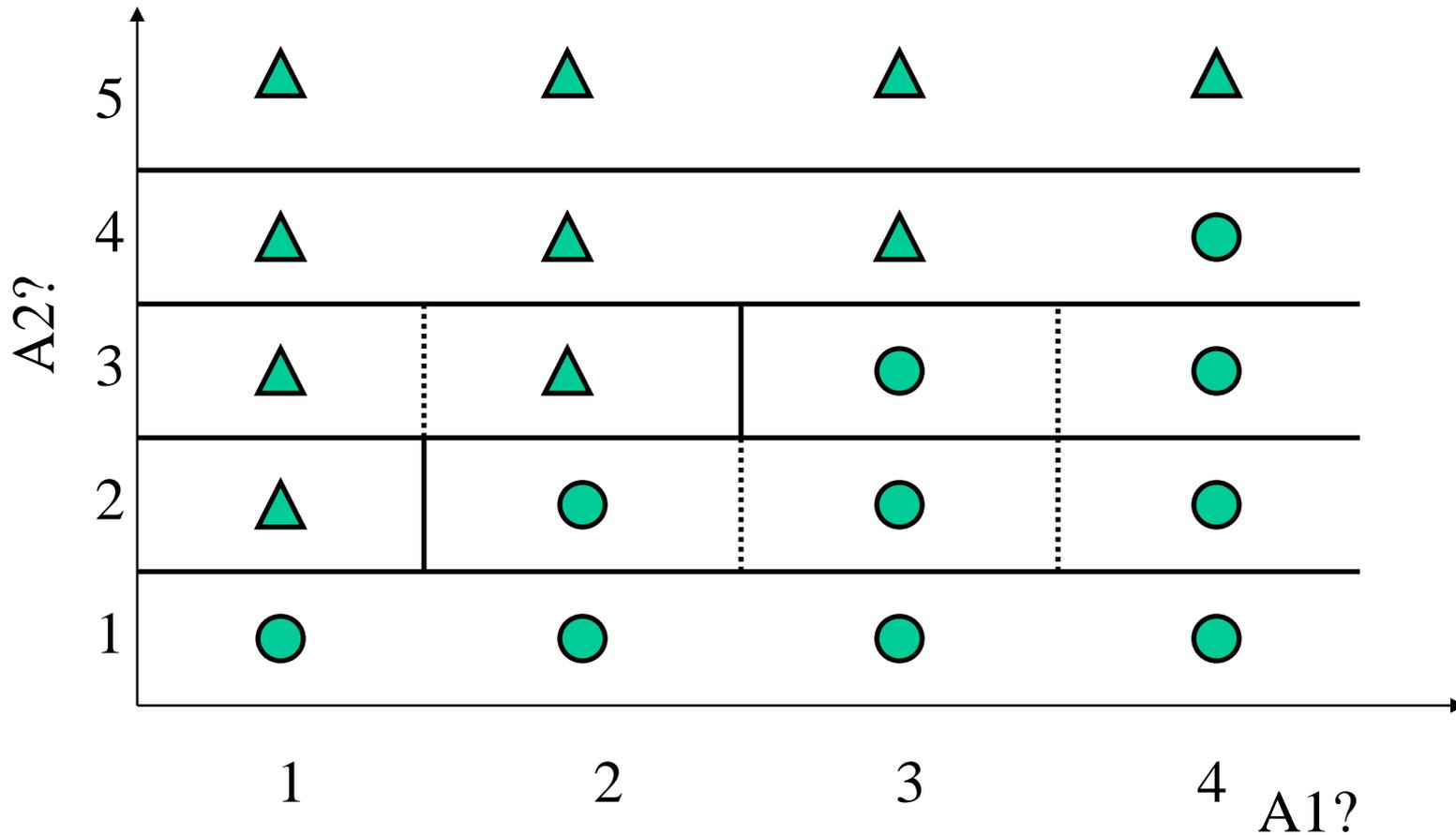
Árboles de clasificación

# Ejemplo con frontera diagonal



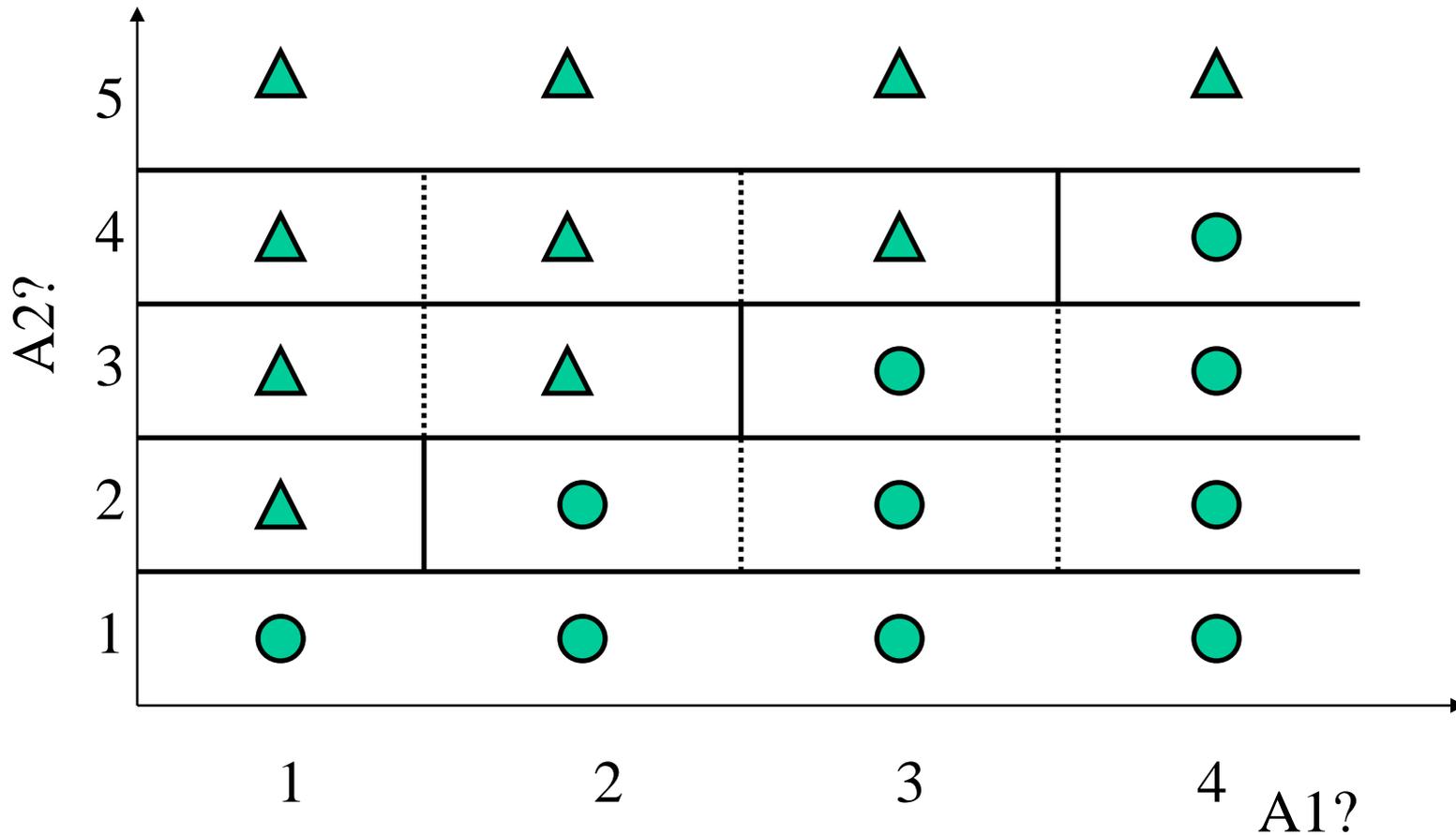
Árboles de clasificación

# Ejemplo con frontera diagonal



Árboles de clasificación

# Ejemplo con frontera diagonal



Árboles de clasificación

# Salida ID3

**A2 = 1: 0**

**A2 = 2**

| **A1 = 1: 1**

| **A1 = 2: 0**

| **A1 = 3: 0**

| **A1 = 4: 0**

**A2 = 3**

| **A1 = 1: 1**

| **A1 = 2: 1**

| **A1 = 3: 0**

| **A1 = 4: 0**

**A2 = 4**

| **A1 = 1: 1**

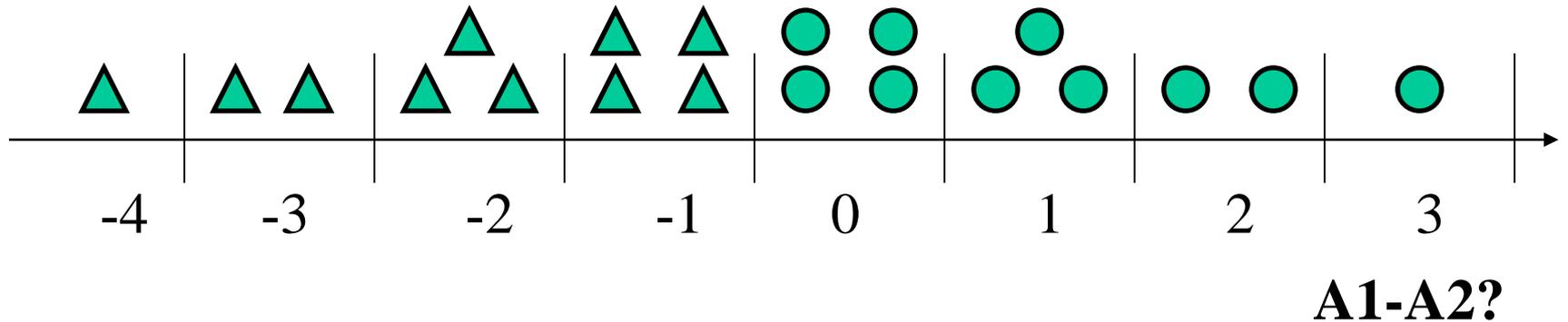
| **A1 = 2: 1**

| **A1 = 3: 1**

| **A1 = 4: 0**

**A2 = 5: 1**

# Transformación de atributos



Árboles de clasificación

# Salida ID3

Dif = -4: 1

Dif = -3: 1

Dif = -2: 1

Dif = -1: 1

Dif = 0: 0

Dif = 1: 0

Dif = 2: 0

Dif = 3: 0