

# Ejercicio 1

- Interpreta los siguientes clusters de un conjunto de datos que representa viviendas de los suburbios Boston utilizando tres variables: tasa de criminalidad per capita, proporción de acres industriales en la zona, concentración de óxido nítrico.

		cluster		
		0	1	2
CRIM	Num Elem	317	181	8
	media	0.22	7.26	55.61
	desv. tip.	0.30	6.30	18.98
	min	0.01	0.07	37.66
	max	2.64	28.66	88.98
INDUS	media	6.35	19.22	18.10
	desv. tip.	3.27	2.26	0.00
	min	0.46	18.10	18.10
	max	15.04	27.74	18.10
NOX	media	0.48	0.68	0.67
	desv. tip.	0.06	0.08	0.03
	min	0.39	0.53	0.60
	max	0.65	0.87	0.69

# Ejercicio 1

## Cluster 0

- El cluster más numeroso
- Zona de escasa criminalidad y con escasa desviación y rango
- Zona poco industrializada y de baja contaminación

## Cluster 1

- Zona de criminalidad media
- Zona industrializada
- Contaminación elevada (desviación alta)

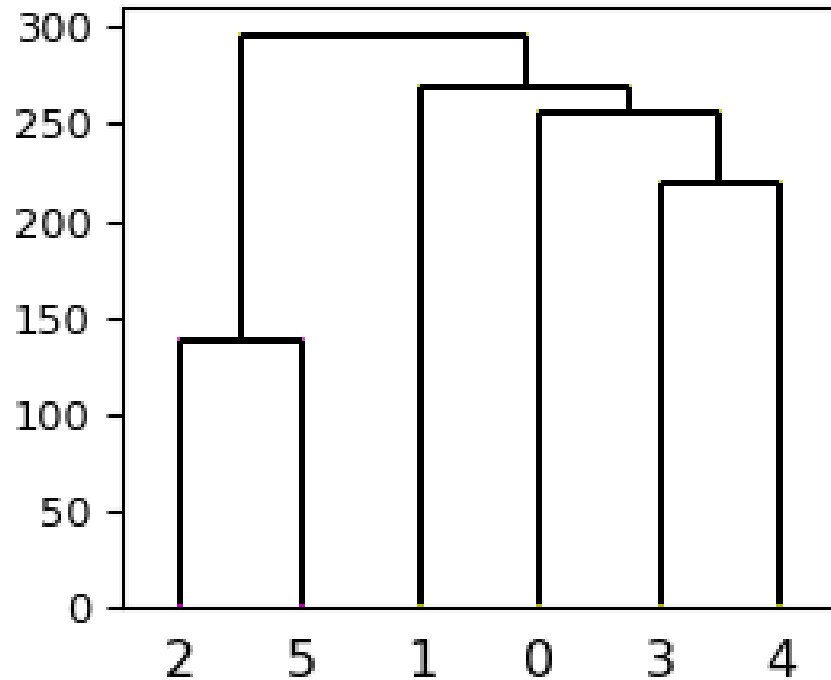
## Cluster 2

- Solo 8 viviendas
- En zona de elevada criminalidad, de zona industrial (¿la misma?), elevada contaminación

		cluster		
		0	1	2
CRIM	Num Elem	317	181	8
	media	0.22	7.26	55.61
	desv. tip.	0.30	6.30	18.98
	min	0.01	0.07	37.66
INDUS	max	2.64	28.66	88.98
	media	6.35	19.22	18.10
	desv. tip.	3.27	2.26	0.00
	min	0.46	18.10	18.10
NOX	max	15.04	27.74	18.10
	media	0.48	0.68	0.67
	desv. tip.	0.06	0.08	0.03
	min	0.39	0.53	0.60
	max	0.65	0.87	0.69

## Ejercicio 2

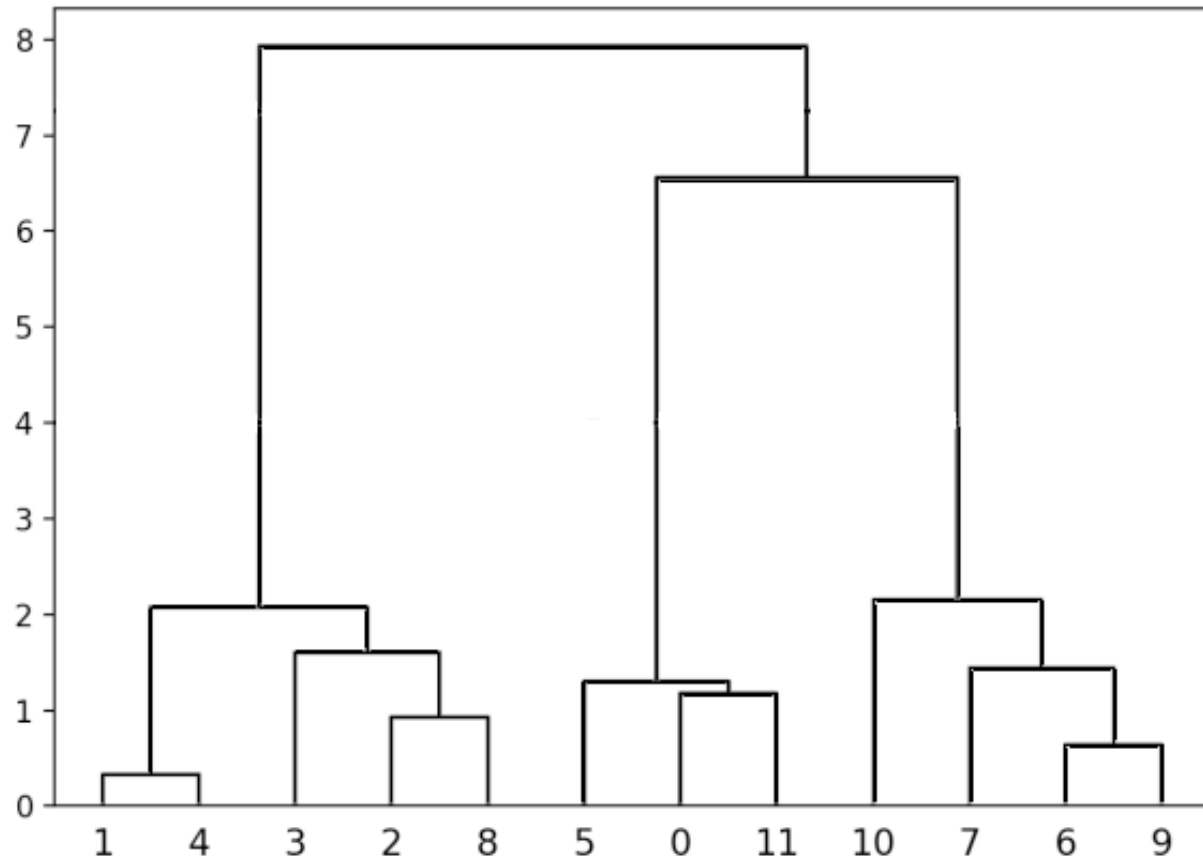
- Interpreta los siguientes dendrogramas. ¿Cuántos clusters consideras que hay en el conjunto de datos? ¿Qué puedes decir de cada uno de ellos?



## Ejercicio 2

- ❑ ¿Cuántos clusters consideras que hay en el conjunto de datos?
  - ❑ Se puede considerar que hay dos clusters el formado por los elementos 2 y 5, y el formado por 0,1,3 y 4.
- ❑ ¿Qué puedes decir de cada uno de ellos?
  - ❑ El cluster de los elementos 2 y 5 es más compacto que el otro ya que los dos elementos se unen a una distancia inferior a 150.
  - ❑ Mientras que el otro cluster es algo más numeroso, pero todos los elementos son más diferentes entre sí, porque la primera fusión que se produce lo hace para un valor de distancia superior a 200, y el resto de elementos se agregan uno a uno a distancias superiores a 250.

# Ejercicio 2

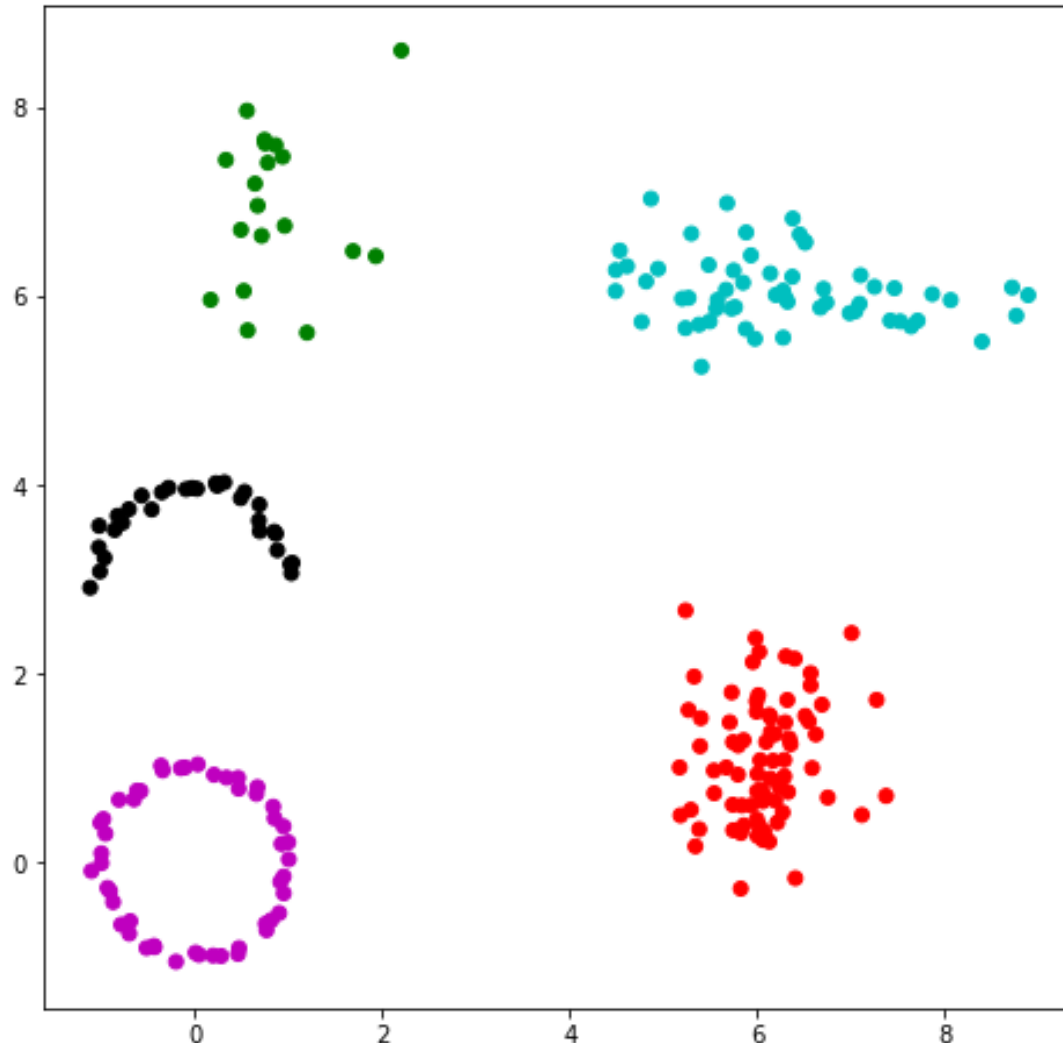


## Ejercicio 2

- ❑ ¿Cuántos clusters consideras que hay en el conjunto de datos?
  - ❑ Se puede considerar que hay tres clusters que se forman al establecer un corte a la distancia de 2.5
  - ❑ Otra alternativa sería considerar dos clusters, estableciendo el punto de corte en torno al 7. Sin embargo, la unión entre los dos subclusters de la derecha se produce a gran distancia (en torno al 6.5), mucho más cerca del corte en 1 cluster que del corte en 3 clusters
    - ❑ Esto quiere decir que ese cluster estaría agregando dos subgrupos razonablemente heterogéneos entre sí
- ❑ ¿Qué puedes decir de cada uno de ellos?
  - ❑ Los tres clusters son razonablemente homogéneos y el número de elementos es similar entre ellos.

## Ejercicio 3

- Dados los cinco grupos del siguiente gráfico de dispersión, indica cuales de las siguientes afirmaciones son ciertas.



## Ejercicio 3

- ❑ La agrupación con forma de luna no puede ser devuelta como cluster por el algoritmo de clustering jerárquico que agrupa los clusters usando el centroide (es decir, la media), si fijamos el número de clusters a 5.
- ❑ Si fijamos el número de clusters a 5, el algoritmo de clustering jerárquico agrupando clusters mediante enlace simple confundirá los dos grupos de la parte superior.
- ❑ El algoritmo de k-medias con  $k=5$  probablemente recuperaría los cinco grupos.
- ❑ La dispersión del grupo verde y del grupo azul no impide que puedan ser recuperados por el clustering jerárquico que agrupa los clusters usando el enlace completo.



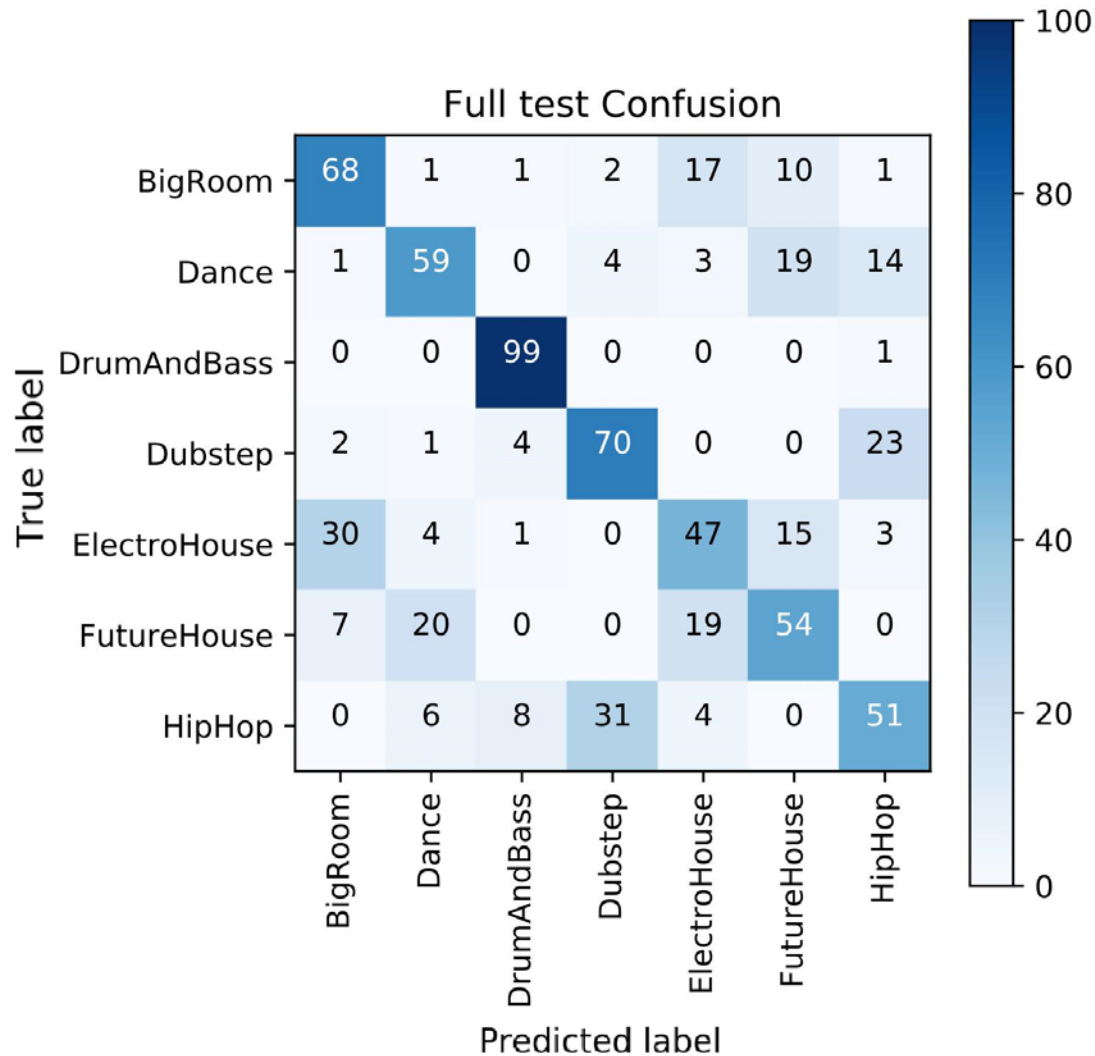
## Ejercicio 3

- ✗ ☐ La agrupación con forma de luna no puede ser devuelta como cluster por el algoritmo de clustering jerárquico que agrupa los clusters usando el centroide (es decir, la media), si fijamos el número de clusters a 5.
- ✗ ☐ Si fijamos el número de clusters a 5, el algoritmo de clustering jerárquico agrupando clusters mediante enlace simple confundirá los dos grupos de la parte superior.
- ✓ ☐ El algoritmo de k-medias con  $k=5$  probablemente recuperaría los cinco grupos.
- ✓ ☐ La dispersión del grupo verde y del grupo azul no impide que puedan ser recuperados por el clustering jerárquico que agrupa los clusters usando el enlace completo.

## Ejercicio 4

- ❑ Interpreta la siguiente matriz de confusión que ha arrojado un algoritmo de clasificación en Scikit Learn. El objetivo es clasificar canciones según el subgénero de música electrónica al que pertenecen. Son 7 subgéneros y se dispone de 100 canciones por subgénero.
  - ❑ ¿Cuál es la tasa de aciertos del clasificador?
  - ❑ ¿Las canciones de qué subgénero se identifican peor? ¿Con qué subgéneros se confunden más?
  - ❑ Interpreta la precisión y la exhaustividad del Drum and Bass, del Electro House y del Hip Hop.

# Ejercicio 4



## Ejercicio 4

- ❑ ¿Cuál es la tasa de aciertos del clasificador?
  - ❑ Número de aciertos (suma de los valores de la diagonal) entre el número de canciones totales (suma de todas las celdas)
  - ❑ Tasa de aciertos=  $448/700 \times 100 = 64\%$
  
- ❑ ¿Las canciones de qué subgénero se identifican peor? ¿Con qué subgéneros se confunden más?
  - ❑ El clasificador identifica peor las canciones de Electro House. Ya que solo identifica como tales el 47% de las canciones de Electro House (exhaustividad).
  - ❑ Un 30% de las canciones de Electro House el clasificador las confunde con canciones Big Room y un 15% con Future House
    - ❑ Hay confusiones menores con Dance, Drum and Bass y Hip Hop

# Ejercicio 4

- ❑ Interpreta la precisión y la exhaustividad del Drum and Bass, del Electro House y del Hip Hop.

$$Precisión = \frac{VP}{VP+FP}$$

- ❑ Drum and Bass

$$Exhaustividad = \frac{VP}{P}$$

- ❑ Precisión =  $99/113 * 100 = 87.6\%$

- ❑ De las veces que el clasificador predice DnB acierta el 87.6%

- ❑ Exhaustividad =  $99/100 * 100 = 99\%$

- ❑ De todas las canciones que realmente son DnB el clasificador clasifica como tales el 99% de ellas

- ❑ Electro House

- ❑ Precisión =  $47/90 * 100 = 52.2\%$

- ❑ Exhaustividad =  $47/100 * 100 = 47\%$

- ❑ Hip Hop

- ❑ Precisión =  $51/93 * 100 = 54.8\%$

- ❑ Exhaustividad =  $51/100 * 100 = 51\%$

## Ejercicio 5: ID3 para riesgo de quemaduras

- Aplicar el algoritmo ID3 para obtener el árbol de decisión que permita clasificar a las personas según su **riesgo de padecer quemaduras solares**

# Ejercicio 5: ID3 para riesgo de quemaduras

❑ Los datos de los que se dispone son los siguientes:

<i><b>Nombre</b></i>	<i><b>Pelo</b></i>	<i><b>Estatura</b></i>	<i><b>Peso</b></i>	<i><b>Loción</b></i>	<i><b>Quema- duras</b></i>
<i>Sara</i>	<i>rubio</i>	<i>media</i>	<i>bajo</i>	<i>no</i>	<i>sí</i>
<i>Diana</i>	<i>rubio</i>	<i>alta</i>	<i>medio</i>	<i>sí</i>	<i>no</i>
<i>Alejandro</i>	<i>castaño</i>	<i>baja</i>	<i>medio</i>	<i>sí</i>	<i>no</i>
<i>Ana</i>	<i>rubio</i>	<i>baja</i>	<i>medio</i>	<i>no</i>	<i>sí</i>
<i>Emilio</i>	<i>pelirrojo</i>	<i>media</i>	<i>alto</i>	<i>no</i>	<i>sí</i>
<i>Pedro</i>	<i>castaño</i>	<i>alta</i>	<i>alto</i>	<i>no</i>	<i>no</i>
<i>Juan</i>	<i>castaño</i>	<i>media</i>	<i>alto</i>	<i>no</i>	<i>no</i>
<i>Carolina</i>	<i>rubio</i>	<i>baja</i>	<i>bajo</i>	<i>sí</i>	<i>no</i>

- ❑ Datos sobre los que opera el algoritmo:
  - ❑ Conjunto de ejemplos de entrenamiento
    - ❑  $E$
  - ❑ Conjunto de las clases a las que pueden pertenecer los ejemplos
    - ❑  $Cl_1, Cl_2, \dots, Cl_N$
  - ❑ Conjunto de atributos definidos sobre los ejemplos
    - ❑  $A, B, C, \dots$
  - ❑ Conjuntos con los valores posibles para cada atributo
    - ❑  $A_1, \dots, A_K, B_1, \dots, B_L, C_1, \dots, C_M, \dots$



# ID3: datos del ejercicio

- ❑ Datos sobre los que opera el algoritmo:
  - ❑ Conjunto de ejemplos de entrenamiento
    - ❑ El formado por las 8 personas de la tabla
  - ❑ Conjunto de las clases a las que pueden pertenecer los ejemplos
    - ❑ Riesgo o SinRiesgo, según si hay riesgo o no de padecer quemaduras solares
  - ❑ Conjunto de atributos definidos sobre los ejemplos
    - ❑ Pelo, estatura, peso, loción
      - ❑ El nombre no es un atributo (*simplemente es una clave o identificador*)
  - ❑ Conjuntos con los valores posibles para cada atributo
    - ❑ Pelo: rubio, castaño, pelirrojo
    - ❑ Estatura: media, alta, baja
    - ❑ Peso: bajo, medio, alto
    - ❑ Loción: sí, no

# ID3: entropía inicial

- La **entropía inicial** de un nodo  $X$ , antes de clasificar los ejemplos que contiene en base a alguno de los atributos:

$$E(X) = - \sum_{j=1}^N P_X(Cl_j) \cdot \log_2 P_X(Cl_j)$$

donde

$$\log_2 x = \begin{cases} 0 & \text{si } x = 0 \\ \log_2 x & \text{en otro caso} \end{cases}$$

y la probabilidad de una clase  $Cl_j$  en el nodo  $X$  es

$$P_X(Cl_j) = \frac{|n^\circ \text{ de } \_ \text{ejemplos } \_ \text{correspondientes } \_ \text{a } Cl_j \_ \text{en } X|}{|n^\circ \text{ total } \_ \text{de } \_ \text{ejemplos } \_ \text{en } X|}$$

para  $j \in [1..N]$

# ID3: entropía inicial

□ Entropía inicial en la raíz del árbol:

□  $P(\text{Riesgo}) = 3/8$

□  $P(\text{SinRiesgo}) = 5/8$

□  $E(\text{raíz}) = -3/8 \log_2 3/8 - 5/8 \log_2 5/8 = 0,5306 + 0,4238 = 0,9544$

<i>Nombre</i>	<i>Pelo</i>	<i>Estatura</i>	<i>Peso</i>	<i>Loción</i>	<i>Quemaduras</i>
<i>Sara</i>	<i>rubio</i>	<i>media</i>	<i>bajo</i>	<i>no</i>	<i>sí</i>
<i>Diana</i>	<i>rubio</i>	<i>alta</i>	<i>medio</i>	<i>sí</i>	<i>no</i>
<i>Alejandro</i>	<i>castaño</i>	<i>baja</i>	<i>medio</i>	<i>sí</i>	<i>no</i>
<i>Ana</i>	<i>rubio</i>	<i>baja</i>	<i>medio</i>	<i>no</i>	<i>sí</i>
<i>Emilio</i>	<i>pelirrojo</i>	<i>media</i>	<i>alto</i>	<i>no</i>	<i>sí</i>
<i>Pedro</i>	<i>castaño</i>	<i>alta</i>	<i>alto</i>	<i>no</i>	<i>no</i>
<i>Juan</i>	<i>castaño</i>	<i>media</i>	<i>alto</i>	<i>no</i>	<i>no</i>
<i>Carolina</i>	<i>rubio</i>	<i>baja</i>	<i>bajo</i>	<i>sí</i>	<i>no</i>

# ID3: entropía final

- La **entropía final** del nodo  $X$  al ramificar utilizando el atributo  $A$ , es igual a la suma de las entropías de los nodos resultantes de fijar el valor del atributo multiplicadas por la probabilidad de cada valor

$$E_A(X) = \sum_{i=1}^k P_X(A_i) \cdot E(A_i)$$

donde la probabilidad  $P_X(A_i)$  es

$$P_X(A_i) = \frac{|n^\circ \text{ de } \_ \text{ejemplos } \_ \text{ en } \_ X \_ \text{ con } \_ \text{ atributo } \_ A = A_i|}{|n^\circ \text{ total } \_ \text{ de } \_ \text{ ejemplos } \_ \text{ en } \_ X|}$$

# ID3: entropía final por el atributo pelo

□ Entropía final clasificando según el atributo pelo:

□  $E_{\text{pelo}}(\text{raíz}) =$

$P(\text{rubio}) * E(\text{rubio}) + P(\text{castaño}) * E(\text{castaño}) + P(\text{pelirrojo}) * E(\text{pelirrojo})$

□  $P(\text{rubio}) = 4/8 = 1/2$

□  $E(\text{rubio}) = -1/2 \log_2 1/2 - 1/2 \log_2 1/2 = 1/2 + 1/2 = 1$

<b>Nombre</b>	<b>Pelo</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quemaduras</b>
Sara	rubio	media	bajo	no	sí
Diana	rubio	alta	medio	sí	no
Alejandro	castaño	baja	medio	sí	no
Ana	rubio	baja	medio	no	sí
Emilio	pelirrojo	media	alto	no	sí
Pedro	castaño	alta	alto	no	no
Juan	castaño	media	alto	no	no
Carolina	rubio	baja	bajo	sí	no

# ID3: entropía final por el atributo pelo

□ Entropía final clasificando según el atributo pelo:

□  $E_{\text{pelo}}(\text{raíz}) = 1/2 + P(\text{castaño}) * E(\text{castaño}) + P(\text{pelirrojo}) * E(\text{pelirrojo})$

□  $P(\text{castaño}) = 3/8$

□  $E(\text{castaño}) = 0$

<b>Nombre</b>	<b>Pelo</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quema- duras</b>
Sara	rubio	media	bajo	no	sí
Diana	rubio	alta	medio	sí	no
Alejandro	castaño	baja	medio	sí	no
Ana	rubio	baja	medio	no	sí
Emilio	pelirrojo	media	alto	no	sí
Pedro	castaño	alta	alto	no	no
Juan	castaño	media	alto	no	no
Carolina	rubio	baja	bajo	sí	no

# ID3: entropía final por el atributo pelo

□ Entropía final clasificando según el atributo pelo:

□  $E_{\text{pelo}}(\text{raíz}) = 1/2 + 0 + P(\text{pelirrojo}) * E(\text{pelirrojo})$

□  $P(\text{pelirrojo}) = 1/8$

□  $E(\text{pelirrojo}) = 0$

□  $E_{\text{pelo}}(\text{raíz}) = 1/2 + 0 + 0 = 1/2$

<b>Nombre</b>	<b>Pelo</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quema- duras</b>
Sara	rubio	media	bajo	no	sí
Diana	rubio	alta	medio	sí	no
Alejandro	castaño	baja	medio	sí	no
Ana	rubio	baja	medio	no	sí
Emilio	pelirrojo	media	alto	no	sí
Pedro	castaño	alta	alto	no	no
Juan	castaño	media	alto	no	no
Carolina	rubio	baja	bajo	sí	no

# ID3: entropía final por el atributo estatura

□ Entropía final clasificando según el atributo estatura:

□  $E_{\text{estatura}}(\text{raíz}) = P(\text{baja}) * E(\text{baja}) + P(\text{media}) * E(\text{media}) + P(\text{alta}) * E(\text{alta})$

□  $P(\text{baja}) = 3/8$

□  $E(\text{baja}) = -1/3 \log_2 1/3 - 2/3 \log_2 2/3 = 1/3 1.585 + 2/3 0.585 = 0.9183$

<b>Nombre</b>	<b>Pelo</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quema- duras</b>
Sara	rubio	media	bajo	no	sí
Diana	rubio	alta	medio	sí	no
Alejandro	castaño	baja	medio	sí	no
Ana	rubio	baja	medio	no	sí
Emilio	pelirrojo	media	alto	no	sí
Pedro	castaño	alta	alto	no	no
Juan	castaño	media	alto	no	no
Carolina	rubio	baja	bajo	sí	no



# ID3: entropía final por el atributo estatura

□ Entropía final clasificando según el atributo estatura:

□  $E_{\text{estatura}}(\text{raíz}) = 0.344 + P(\text{media}) * E(\text{media}) + P(\text{alta}) * E(\text{alta})$

□  $P(\text{media}) = 3/8$

□  $E(\text{media}) = -2/3 \log_2 2/3 - 1/3 \log_2 1/3 = 2/3 0.585 + 1/3 1.585 = 0.9183$

<b>Nombre</b>	<b>Pelo</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quemaduras</b>
Sara	rubio	media	bajo	no	sí
Diana	rubio	alta	medio	sí	no
Alejandro	castaño	baja	medio	sí	no
Ana	rubio	baja	medio	no	sí
Emilio	pelirrojo	media	alto	no	sí
Pedro	castaño	alta	alto	no	no
Juan	castaño	media	alto	no	no
Carolina	rubio	baja	bajo	sí	no

# ID3: entropía final por el atributo estatura

□ Entropía final clasificando según el atributo estatura:

□  $E_{\text{estatura}}(\text{raíz}) = 0.344 + 0.344 + P(\text{alta}) * E(\text{alta})$

□  $P(\text{alta}) = 2/8$

□  $E(\text{alta}) = 0$

□  $E_{\text{estatura}}(\text{raíz}) = 0.344 + 0.344 + 0 = 0.688$

<b>Nombre</b>	<b>Pelo</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quema- duras</b>
Sara	rubio	media	bajo	no	sí
Diana	rubio	alta	medio	sí	no
Alejandro	castaño	baja	medio	sí	no
Ana	rubio	baja	medio	no	sí
Emilio	pelirrojo	media	alto	no	sí
Pedro	castaño	alta	alto	no	no
Juan	castaño	media	alto	no	no
Carolina	rubio	baja	bajo	sí	no

# ID3: entropía final por el atributo peso

□ Entropía final clasificando según el atributo peso:

□  $E_{\text{peso}}(\text{raíz}) = P(\text{bajo}) * E(\text{bajo}) + P(\text{medio}) * E(\text{medio}) + P(\text{alto}) * E(\text{alto})$

□  $P(\text{bajo}) = 2/8 = 1/4$

□  $E(\text{bajo}) = -1/2 \log_2 1/2 - 1/2 \log_2 1/2 = 1/2 + 1/2 = 1$

<b>Nombre</b>	<b>Pelo</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quema- duras</b>
Sara	rubio	media	bajo	no	sí
Diana	rubio	alta	medio	sí	no
Alejandro	castaño	baja	medio	sí	no
Ana	rubio	baja	medio	no	sí
Emilio	pelirrojo	media	alto	no	sí
Pedro	castaño	alta	alto	no	no
Juan	castaño	media	alto	no	no
Carolina	rubio	baja	bajo	sí	no

# ID3: entropía final por el atributo peso

□ Entropía final clasificando según el atributo peso:

□  $E_{\text{peso}}(\text{raíz}) = 1/4 + P(\text{medio}) * E(\text{medio}) + P(\text{alto}) * E(\text{alto})$

□  $P(\text{medio}) = 3/8$

□  $E(\text{medio}) = -1/3 \log_2 1/3 - 2/3 \log_2 2/3 = 1/3 1.585 + 2/3 0.585 = 0.9183$

<b>Nombre</b>	<b>Pelo</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quema- duras</b>
Sara	rubio	media	bajo	no	sí
Diana	rubio	alta	medio	sí	no
Alejandro	castaño	baja	medio	sí	no
Ana	rubio	baja	medio	no	sí
Emilio	pelirrojo	media	alto	no	sí
Pedro	castaño	alta	alto	no	no
Juan	castaño	media	alto	no	no
Carolina	rubio	baja	bajo	sí	no

# ID3: entropía final por el atributo peso

□ Entropía final clasificando según el atributo peso:

□  $E_{\text{peso}}(\text{raíz}) = 1/4 + 0.344 + P(\text{alto}) * E(\text{alto})$

□  $P(\text{alto}) = 3/8$

□  $E(\text{alto}) = -1/3 \log_2 1/3 - 2/3 \log_2 2/3 = 1/3 1.585 + 2/3 0.585 = 0.9183$

□  $E_{\text{peso}}(\text{raíz}) = 0.25 + 0.344 + 0.344 = 0.938$

<b>Nombre</b>	<b>Pelo</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quema- duras</b>
Sara	rubio	media	bajo	no	sí
Diana	rubio	alta	medio	sí	no
Alejandro	castaño	baja	medio	sí	no
Ana	rubio	baja	medio	no	sí
Emilio	pelirrojo	media	alto	no	sí
Pedro	castaño	alta	alto	no	no
Juan	castaño	media	alto	no	no
Carolina	rubio	baja	bajo	sí	no

# ID3: entropía final por el atributo loción

□ Entropía final clasificando según el atributo loción:

□  $E_{\text{loción}}(\text{raíz}) = P(\text{sí}) * E(\text{sí}) + P(\text{no}) * E(\text{no})$

□  $P(\text{sí}) = 3/8$

□  $E(\text{sí}) = 0$

<b>Nombre</b>	<b>Pelo</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quema- duras</b>
Sara	rubio	media	bajo	no	sí
Diana	rubio	alta	medio	sí	no
Alejandro	castaño	baja	medio	sí	no
Ana	rubio	baja	medio	no	sí
Emilio	pelirrojo	media	alto	no	sí
Pedro	castaño	alta	alto	no	no
Juan	castaño	media	alto	no	no
Carolina	rubio	baja	bajo	sí	no

# ID3: entropía final por el atributo loción

□ Entropía final clasificando según el atributo loción:

□  $E_{\text{loción}}(\text{raíz}) = 0 + P(\text{no}) * E(\text{no})$

□  $P(\text{no}) = 5/8$

□  $E(\text{no}) = -3/5 \log_2 3/5 - 2/5 \log_2 2/5 = 3/5 0.737 + 2/5 1.322 = 0.971$

□  $E_{\text{loción}}(\text{raíz}) = 0 + 0.6068 = 0.6068$

<b>Nombre</b>	<b>Pelo</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quema- duras</b>
Sara	rubio	media	bajo	no	sí
Diana	rubio	alta	medio	sí	no
Alejandro	castaño	baja	medio	sí	no
Ana	rubio	baja	medio	no	sí
Emilio	pelirrojo	media	alto	no	sí
Pedro	castaño	alta	alto	no	no
Juan	castaño	media	alto	no	no
Carolina	rubio	baja	bajo	sí	no

# ID3: disminución de la entropía

## □ Entropía final según cada uno de los tres atributos

$$\square E_{\text{pelo}}(\text{raíz}) = \frac{1}{2} + 0 + 0 = \textcolor{red}{1/2}$$

$$\square E_{\text{estatura}}(\text{raíz}) = 0.344 + 0.344 + 0 = \textcolor{red}{0.688}$$

$$\square E_{\text{peso}}(\text{raíz}) = \frac{1}{4} + 0.344 + 0.344 + 0 = \textcolor{red}{0.938}$$

$$\square E_{\text{loción}}(\text{raíz}) = 0 + 0.6068 = \textcolor{red}{0.6068}$$

## □ Disminución de la entropía:

$$\square \text{Disminución de entropía}_{\text{pelo}}(\text{raíz}) = E(\text{raíz}) - E_{\text{pelo}}(\text{raíz}) \\ = 0.9544 - \frac{1}{2} = \textcolor{red}{0,4544}$$

$$\square \text{Disminución de entropía}_{\text{estatura}}(\text{raíz}) = E(\text{raíz}) - E_{\text{estatura}}(\text{raíz}) \\ = 0.9544 - 0.688 = 0,2664$$

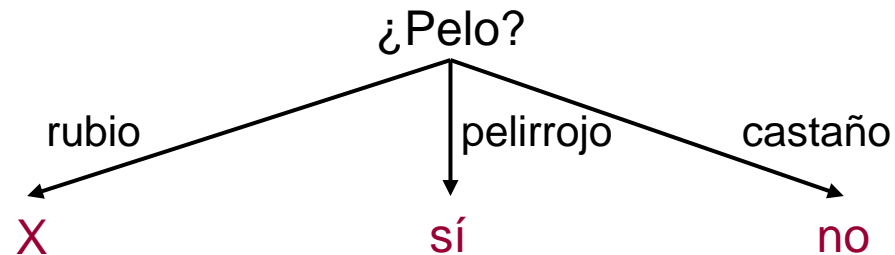
$$\square \text{Disminución de entropía}_{\text{peso}}(\text{raíz}) = E(\text{raíz}) - E_{\text{peso}}(\text{raíz}) \\ = 0.9544 - 0.938 = 0,0164$$

$$\square \text{Disminución de entropía}_{\text{loción}}(\text{raíz}) = E(\text{raíz}) - E_{\text{loción}}(\text{raíz}) \\ = 0.9544 - 0.6068 = 0,3476$$



# ID3: primer nivel del árbol

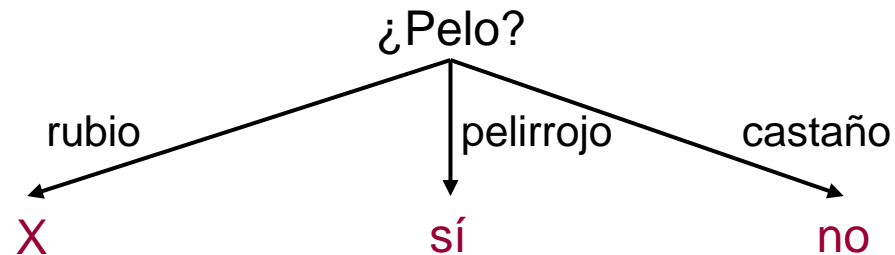
- Atributo más discriminante: **pelo**



<b>Nombre</b>	<b>Pelo</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quemaduras</b>
Sara	rubio	media	bajo	no	sí
Diana	rubio	alta	medio	sí	no
Alejandro	castaño	baja	medio	sí	no
Ana	rubio	baja	medio	no	sí
Emilio	pelirrojo	media	alto	no	sí
Pedro	castaño	alta	alto	no	no
Juan	castaño	media	alto	no	no
Carolina	rubio	baja	bajo	sí	no

# ID3: resto del árbol

□ Nodo a desarrollar: X



<b><i>Nombre</i></b>	<b><i>Estatura</i></b>	<b><i>Peso</i></b>	<b><i>Loción</i></b>	<b><i>Quema- duras</i></b>
<i>Sara</i>	<i>media</i>	<i>bajo</i>	<i>no</i>	<i>sí</i>
<i>Diana</i>	<i>alta</i>	<i>medio</i>	<i>sí</i>	<i>no</i>
<i>Ana</i>	<i>baja</i>	<i>medio</i>	<i>no</i>	<i>sí</i>
<i>Carolina</i>	<i>baja</i>	<i>bajo</i>	<i>sí</i>	<i>no</i>

□ ¿Alguna previsión de lo que va a ocurrir aplicando ID3?

# ID3: entropía inicial

□ Entropía inicial en X:

□  $P(\text{Riesgo}) = 2/4 = 1/2$

□  $P(\text{SinRiesgo}) = 2/4 = 1/2$

□  $E(X) = -1/2 \log_2 1/2 - 1/2 \log_2 1/2 = 1/2 + 1/2 = 1$

<b>Nombre</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quema- duras</b>
<i>Sara</i>	<i>media</i>	<i>bajo</i>	<i>no</i>	<i>sí</i>
<i>Diana</i>	<i>alta</i>	<i>medio</i>	<i>sí</i>	<i>no</i>
<i>Ana</i>	<i>baja</i>	<i>medio</i>	<i>no</i>	<i>sí</i>
<i>Carolina</i>	<i>baja</i>	<i>bajo</i>	<i>sí</i>	<i>no</i>

# ID3: entropía final por el atributo estatura

□ Entropía final clasificando según el atributo estatura:

□  $E_{\text{estatura}}(X) = P(\text{baja}) * E(\text{baja}) + P(\text{media}) * E(\text{media}) + P(\text{alta}) * E(\text{alta})$

□  $P(\text{baja}) = 2/4 = 1/2$

□  $E(\text{baja}) = -1/2 \log_2 1/2 - 1/2 \log_2 1/2 = 1/2 + 1/2 = 1$

□  $E_{\text{estatura}}(X) = 1/2 * 1 + 1/4 * 0 + 1/4 * 0 = 1/2$

<b>Nombre</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quema- duras</b>
Sara	media	bajo	no	sí
Diana	alta	medio	sí	no
Ana	baja	medio	no	sí
Carolina	baja	bajo	sí	no

# ID3: entropía final por el atributo peso

□ Entropía final clasificando según el atributo peso:

□  $E_{\text{peso}}(X) = P(\text{bajo}) * E(\text{bajo}) + P(\text{medio}) * E(\text{medio}) + P(\text{alto}) * E(\text{alto})$

□  $P(\text{bajo}) = 2/4 = 1/2$

□  $E(\text{bajo}) = -1/2 \log_2 1/2 - 1/2 \log_2 1/2 = 1/2 + 1/2 = 1$

□  $P(\text{medio}) = 2/4 = 1/2$

□  $E(\text{medio}) = -1/2 \log_2 1/2 - 1/2 \log_2 1/2 = 1/2 + 1/2 = 1$

□  $P(\text{alto}) = 0$

□  $E_{\text{peso}}(X) = 1/2 * 1 + 1/2 * 1 + 0 = 1$

<b>Nombre</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quemaduras</b>
Sara	media	bajo	no	sí
Diana	alta	medio	sí	no
Ana	baja	medio	no	sí
Carolina	baja	bajo	sí	no

# ID3: entropía final por el atributo loción

□ Entropía final clasificando según el atributo loción:

□  $E_{\text{loción}}(X) = P(\text{sí}) * E(\text{sí}) + P(\text{no}) * E(\text{no})$

□  $P(\text{sí}) = 2/4 = 1/2$

□  $E(\text{sí}) = 0$

□  $P(\text{no}) = 2/4 = 1/2$

□  $E(\text{no}) = 0$

□  $E_{\text{loción}}(X) = 1/2 * 0 + 1/2 * 0 = 0$

<b>Nombre</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quema- duras</b>
Sara	media	bajo	no	sí
Diana	alta	medio	sí	no
Ana	baja	medio	no	sí
Carolina	baja	bajo	sí	no

# ID3: disminución de la entropía

## □ Entropía inicial en X:

$$\square E(X) = -1/2 \log_2 1/2 - 1/2 \log_2 1/2 = 1/2 + 1/2 = 1$$

## □ Entropía final según cada uno de los tres atributos

$$\square E_{\text{estatura}}(X) = 1/2 * 1 + 1/4 * 0 + 1/4 * 0 = 1/2$$

$$\square E_{\text{peso}}(X) = 1/2 * 1 + 1/2 * 1 + 0 = 1$$

$$\square E_{\text{loción}}(X) = 1/2 * 0 + 1/2 * 0 = 0$$

## □ Disminución de la entropía:

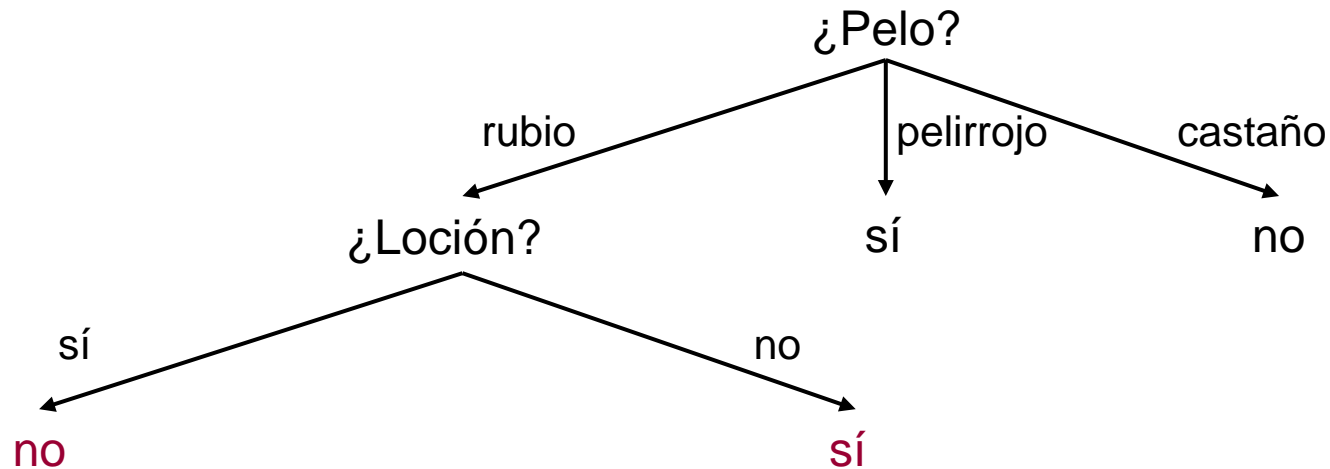
$$\square \text{Disminución de entropía}_{\text{estatura}}(X) = E(X) - E_{\text{estatura}}(X) \\ = 1 - 1/2 = 1/2$$

$$\square \text{Disminución de entropía}_{\text{peso}}(X) = E(X) - E_{\text{peso}}(X) \\ = 1 - 1 = 0$$

$$\square \text{Disminución de entropía}_{\text{loción}}(X) = E(X) - E_{\text{loción}}(X) \\ = 1 - 0 = 1$$

# ID3: árbol de decisión

- Atributo más discriminante: **loción**

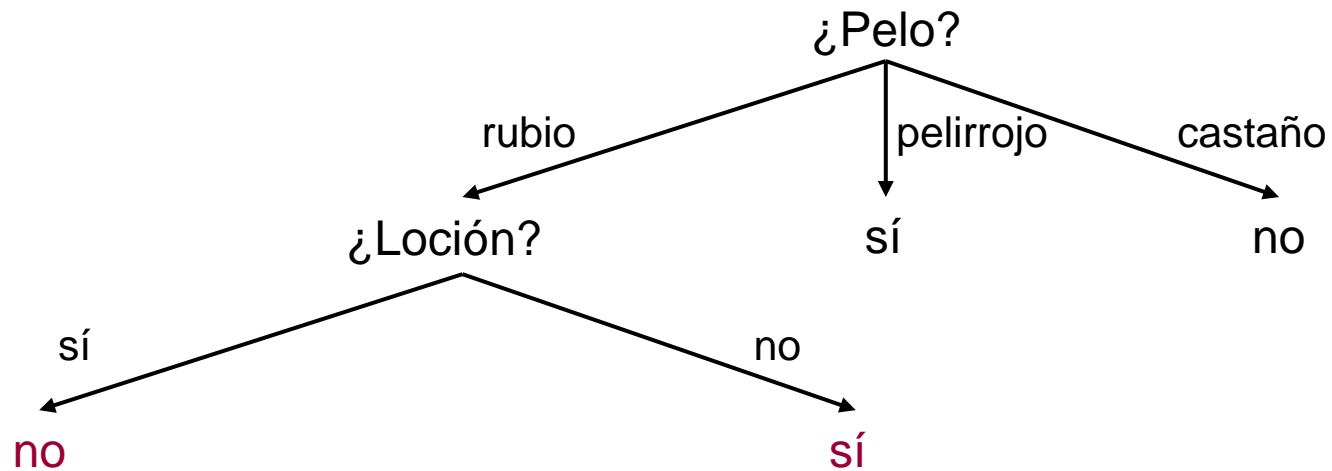


<b>Nombre</b>	<b>Estatura</b>	<b>Peso</b>	<b>Loción</b>	<b>Quemaduras</b>
<i>Sara</i>	<i>media</i>	<i>bajo</i>	<i>no</i>	<i>sí</i>
<i>Diana</i>	<i>alta</i>	<i>medio</i>	<i>sí</i>	<i>no</i>
<i>Ana</i>	<i>baja</i>	<i>medio</i>	<i>no</i>	<i>sí</i>
<i>Carolina</i>	<i>baja</i>	<i>bajo</i>	<i>sí</i>	<i>no</i>



# Ejercicio 5: ID3 para riesgo de quemaduras

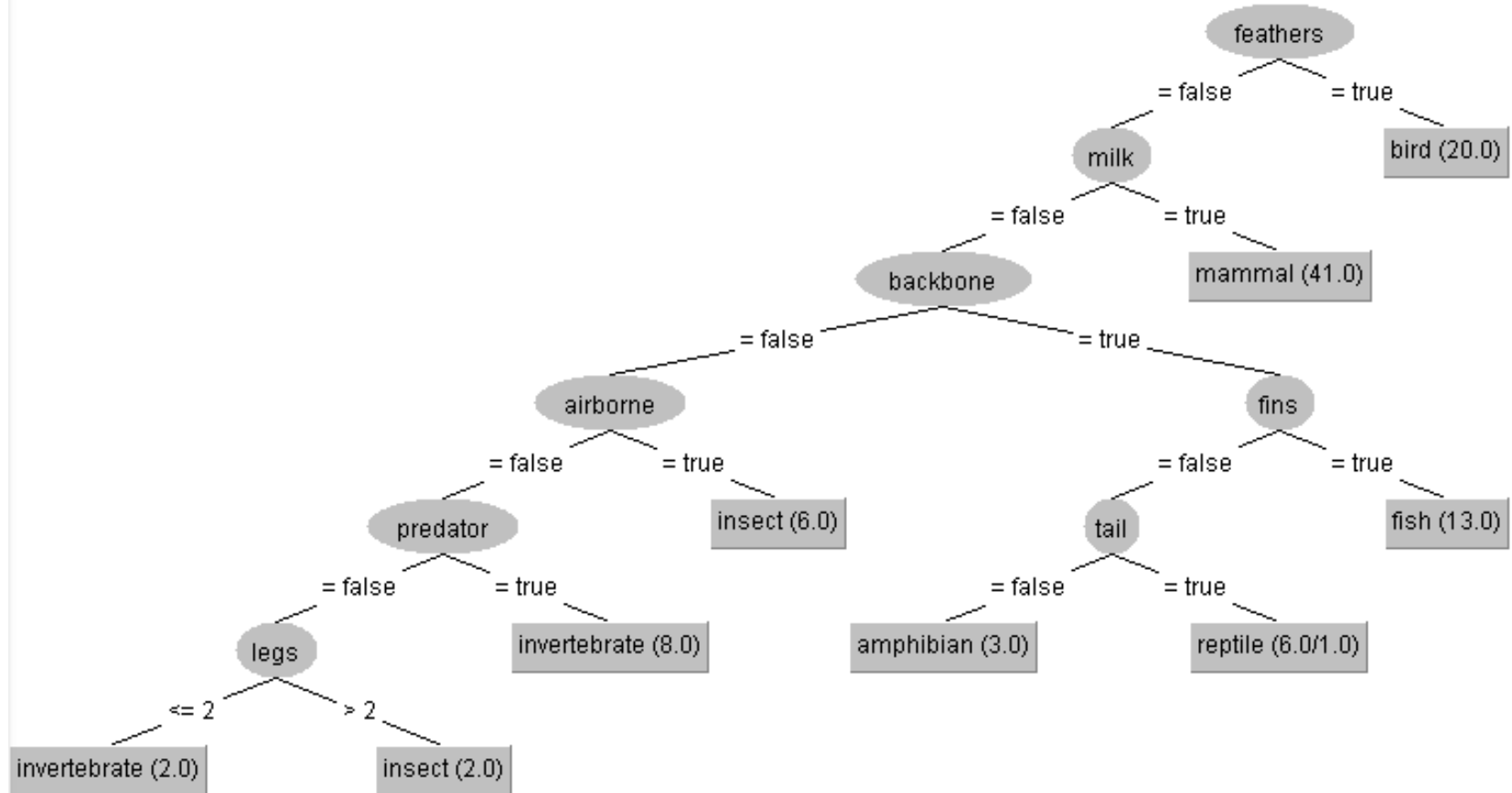
## □ Árbol de decisión



## Ejercicio 6

- ❑ Interpreta el siguiente árbol de decisión generado por WEKA que clasifica animales según su tipo (pájaros, mamíferos, peces, reptiles, anfibios, insectos e invertebrados).
  - ❑ ¿Qué variables son más relevantes?
  - ❑ ¿Qué tipos de animales le cuesta discriminar más al árbol?

# Ejercicio 6



## Ejercicio 6

- ❑ ¿Qué variables son más relevantes?
  - ❑ Las variables más relevantes son las que primero usa el árbol, es decir, “feathers” y “milk” que son bastante eficaces al generar cada una dos nodos puros de “bird” y “mammal”
- ❑ ¿Qué tipos de animales le cuesta discriminar más al árbol?
  - ❑ El árbol consigue todos nodos hoja puros, con la salvedad del nodo “reptile” donde la proporción de reptiles es 6/1
  - ❑ Sin embargo, al nodo “airborne” únicamente llegan dos tipos de animales (“insect”/“invertebrate”) y se necesita tres preguntas para discriminarlos perfectamente.

# Ejercicio 7

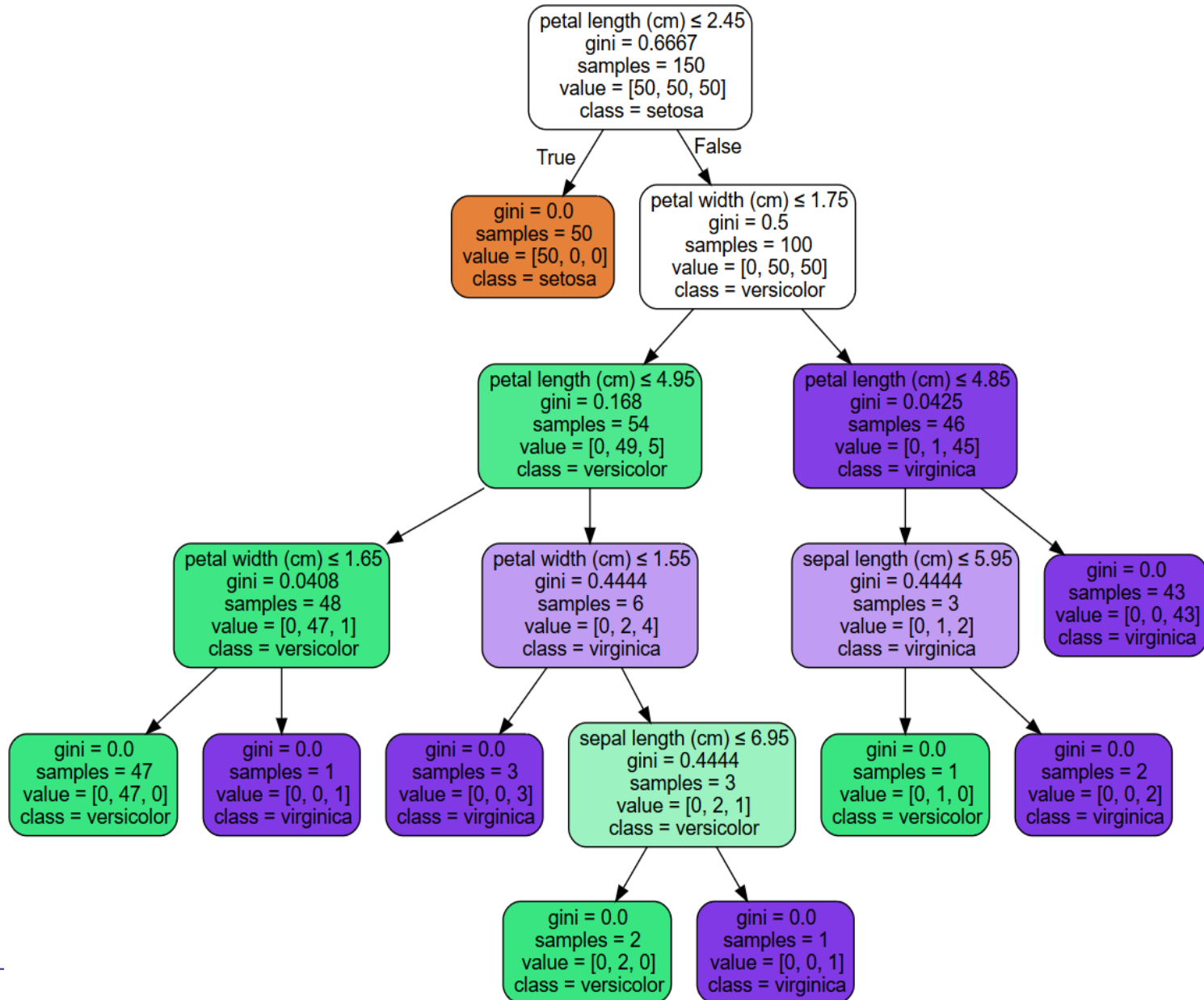
- ❑ Interpreta el siguiente árbol de clasificación generado por Scikit Learn que clasifica las especies de la flor del Iris, según el ancho y el largo del pétalo y el sépalo de la flor.
- ❑ Observa que la representación de Scikit Learn es más rica ya que indica:
  - ❑ el coeficiente de Gini de cada nodo,
  - ❑ el número de muestras que hay,
  - ❑ la distribución de las muestras entre las tres clases
    - ❑ [setosa versicolor virginica]
  - ❑ la clase del nodo, es decir, la clase mayoritaria en la distribución de muestras anterior
    - ❑ que además viene también reflejada por el color del nodo (siendo el blanco la indecisión)

# Ejercicio 7

---

- ❑ Responde a las siguientes preguntas
  - ❑ ¿Qué variables juegan un papel más relevante?
  - ❑ ¿Qué nodos consideras a tu juicio candidatos para ser podados?  
¿por qué?

# Ejercicio 7



# Ejercicio 7

- ❑ ¿Qué variables juegan un papel más relevante?
  - ❑ La longitud del pétalo es la que se usa en primer lugar porque sirve para discriminar perfectamente todas las instancias de setosa
  - ❑ Además, también se usa en el tercer nivel del árbol en ambas ramas
  - ❑ El ancho del pétalo se usa en el segundo nivel del árbol y también permite generar dos nodos con mucha pureza (Gini muy bajo) de versicolor y virginica con gran cantidad de ejemplos ambos nodos
  - ❑ En el árbol que estamos viendo la longitud del sépalo discrimina en nodos con muy pocos ejemplos por lo que o es una diferenciación muy especializada o quizás espuria (dos nodos con 3 ejemplos)
  - ❑ En este árbol el ancho del sépalo no se utiliza lo que nos hace pensar que posiblemente tenga un poder discriminante bajo en este problema
    - ❑ O que quizá exista otra variable de poder discriminante similar que ha hecho que no haya sido elegida en ningún caso



# Ejercicio 7

- ❑ ¿Qué nodos consideras a tu juicio candidatos para ser podados?  
¿por qué?
  - ❑ Posiblemente todo el subárbol desde el nodo `petal length ≤ 4.85` sea irrelevante. Ese nodo tiene ya 45 ejemplos de `virginica` por solamente 1 de `versicolor`. Para poder discriminarlo son además necesarias dos preguntas.
  - ❑ Algo similar sucede con los nodos hijo del nodo `petal length ≤ 4.95`
    - ❑ Uno de ellos tiene 47 ejemplos de `versicolor` por 1 de `virginica` y quizás la pregunta que sirve para discriminarlos sea espuria
    - ❑ El otro tiene 2 ejemplos de `versicolor` y 4 de `virginica` y son necesarias dos preguntas para discriminarlos. En este caso, el problema no es tanto el valor de Gini ya que el nodo es desordenado, sino su bajo número de ejemplos
  - ❑ En cualquier caso, para estar seguros podríamos presentarle a este árbol ejemplos que no hayan sido usados para entrenarlo y ver si los nodos que consideramos irrelevantes realmente lo son o es que han logrado aprender con éxito casos poco representativos

## Ejercicio 8

- Interpreta las siguientes dos tablas que muestran el error de la validación cruzada en 5 partes en el contexto de un problema de regresión. El error que estás viendo es la Raíz Cuadrada del Error Cuadrático Medio y los conjuntos de datos a los que pertenece cada tabla son diferentes.

Parte 1	Parte 2	Parte 3	Parte 4	Parte 5
3.3	3.8	3.5	3.6	3.9

- El error en cada una de las partes es de una naturaleza similar lo que indica que el error medio debe ser un buen indicador del error que cometerá el método de regresión ante datos no conocidos. El error medio es 3.62, y también conviene tener en cuenta el rango de variación del error oscila entre 3.3 y 3.9, que podría parecer razonable.

## Ejercicio 8

Parte 1	Parte 2	Parte 3	Parte 4	Parte 5
35.6	36.2	43.1	37.6	36.9

- ❑ El error de casi todas las partes es bastante homogéneo salvo en la parte 3 donde el error es bastante superior.
  - ❑ Esto puede deberse a que en la parte 3 existan observaciones que sean difíciles de predecir o muy diferentes a las que existen en el resto de las partes del conjunto de datos y por eso el modelo no las aprende correctamente
    - ❑ Si el conjunto de datos no se ha aleatorizado (“barajado”) antes de dividirlo, quizás tenga sentido hacerlo. De esa forma, las observaciones difíciles se repartirán mejor entre las partes y o el modelo será capaz de aprenderlas o su error se repartirá más uniformemente.
- ❑ El error también podría ser causado por alguna observación aberrante aislada (error en el valor recogido) que distorsiona el error.
  - ❑ Conviene revisar dónde se produce el error y ver si la observación es válida o no. Si no lo fuera se debe eliminar del conjunto y volver a entrenar el modelo.

## Ejercicio 9

- Interpreta la siguiente tabla que muestra Raíz Cuadrada del Error Cuadrático Medio para varios métodos de regresión sobre el mismo conjunto de datos. ¿Qué k-NN y qué perceptrón multicapa te parecen más adecuados? ¿Por qué?

Modelo	Error sobre conjunto Entrenamiento	Error sobre conjunto Validación
MLP (tasa de aprendizaje=0.8)	4.65	5.83
MLP (tasa de aprendizaje=0.3)	5.43	5.51
k-NN (k=2, 6 var. de entrada)	5.51	5.64
k-NN (k=8, 2 var. de entrada)	5.54	5.62

# Ejercicio 9

- ❑ Comparación de perceptrones
  - ❑ El primero tiene un error de entrenamiento notablemente inferior al error en validación y una tasa de aprendizaje alta. Mientras que el segundo tiene ambos errores más equilibrados y una tasa de aprendizaje menor.
  - ❑ Parece que el primer perceptrón puede haber sobreaprendido los datos y con ello generalizar peor y resentirse ante datos con los que no ha sido entrenado (como sucede con los datos de validación).
    - ❑ El Segundo perceptrón tiene una tasa de aprendizaje más baja y unos errores de entrenamiento y validación similares por lo que parece una configuración razonable.

# Ejercicio 9

## ❑ Comparación de k-NNs

- ❑ Ambos k-NNs tienen errores de entrenamiento y validación muy similares, y similares a su vez a los del perceptrón
  - ❑ Esto hace pensar que ambas configuraciones, aunque diferentes, ofrecen una buena aproximación al conjunto de datos abordado
  - ❑ Los parámetros del primero indican que busca pocos vecinos en un espacio de 6 dimensiones, mientras que el segundo busca más vecinos en 2 dimensiones
    - ❑ Cuando hay menos dimensiones los puntos se acercan y se pueden encontrar más ejemplos cercanos
- ❑ Si hubiera que elegir entre ambas, quizás la segunda sea más interesante
  - ❑ Por un lado, usa menos variables de entrada (2) con lo que es un método más sencillo y además permite visualizar el espacio de entrada mediante un gráfico en 2D
    - ❑ En aprendizaje automático se suele utilizar el principio de parsimonia (los modelos buenos si son sencillos mejor)
  - ❑ Por otro lado, usa más ejemplos lo cual hace que el valor medio que ofrece se vea menos afectado por valores extremos