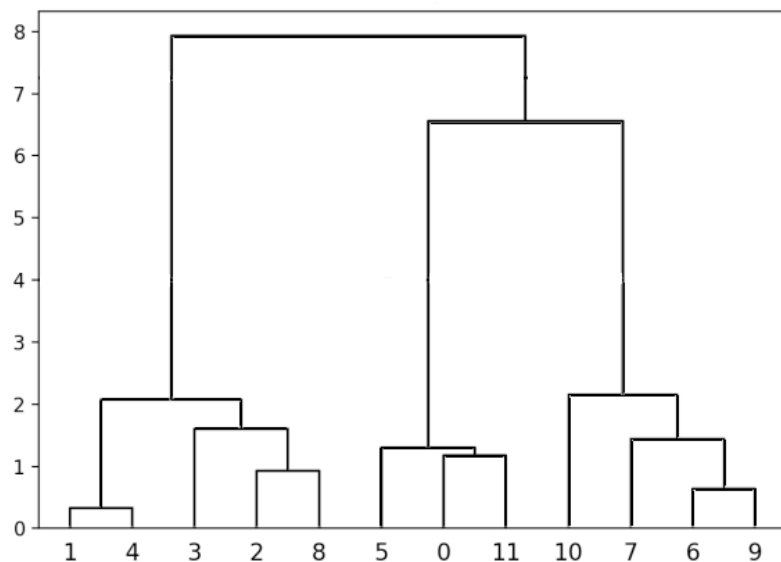


Ejercicios IA2 – Aprendizaje Automático

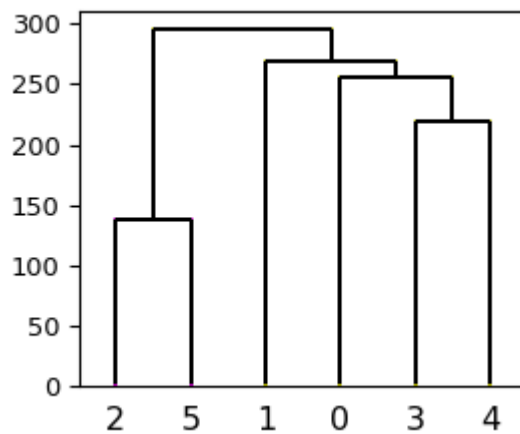
- Interpreta los siguientes clusters de un conjunto de datos que representa viviendas de los suburbios Boston utilizando tres variables: tasa de criminalidad per capita, proporción de acres industriales en la zona, concentración de óxido nítrico.

		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

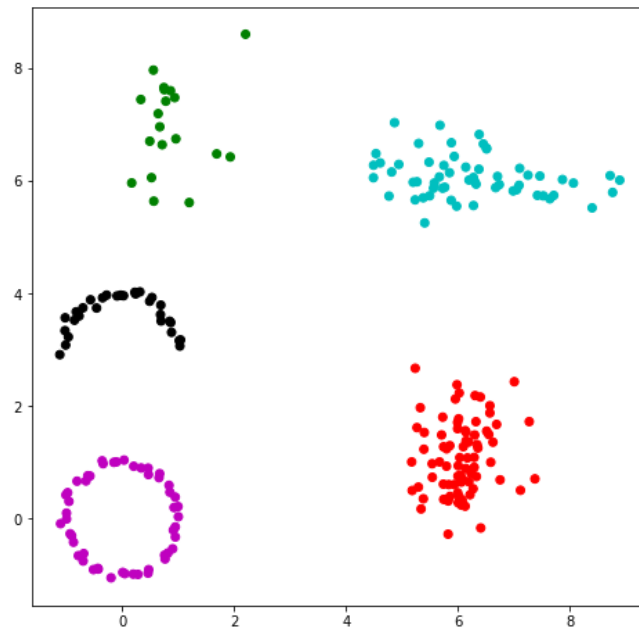
- Interpreta los siguientes dendrogramas. ¿Cuántos clusters consideras que hay en el conjunto de datos? ¿Qué puedes decir de cada uno de ellos?



Ejercicios IA2 – Aprendizaje Automático

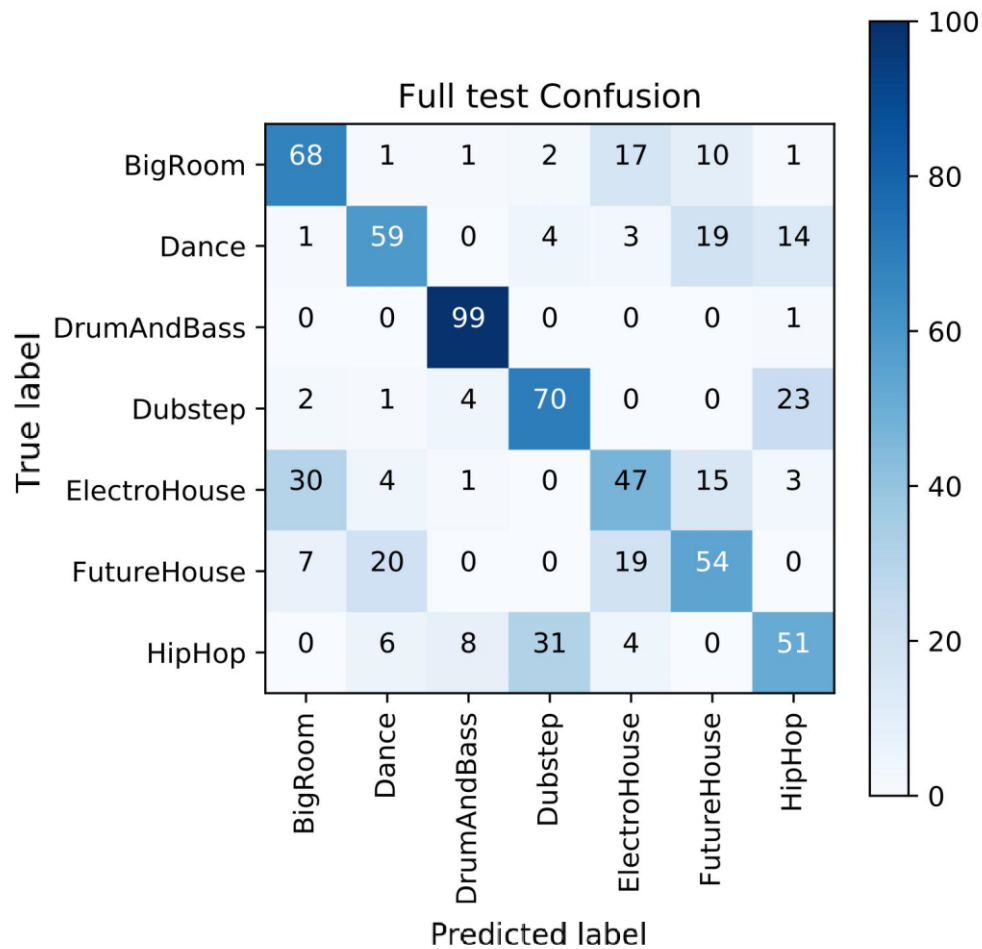


3. Dados los cinco grupos del siguiente gráfico de dispersión, indica cuales de las siguientes afirmaciones son ciertas.



- 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.
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 precisión global 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.

Ejercicios IA2 – Aprendizaje Automático

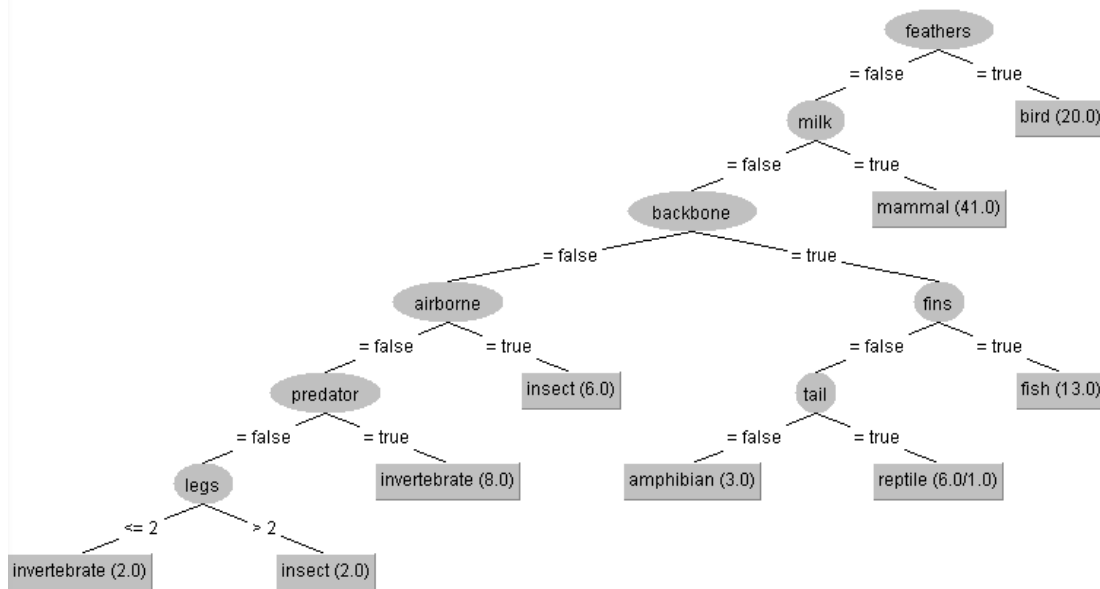


5. Aplicar el algoritmo ID3 para obtener el árbol de decisión que permita clasificar a las personas según su riesgo de padecer quemaduras solares. Los datos de los que se dispone son los siguientes:

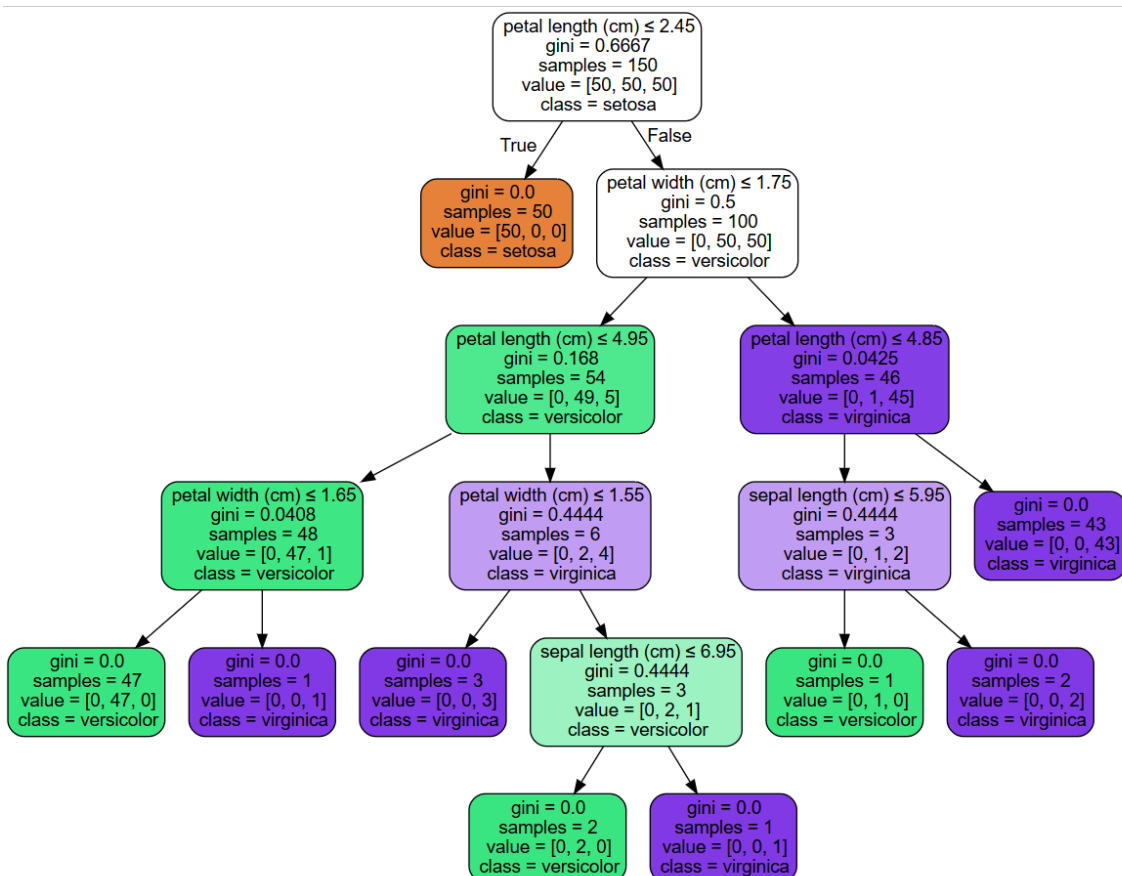
<i>Nombre</i>	<i>Pelo</i>	<i>Estatura</i>	<i>Peso</i>	<i>Loción</i>	<i>Quemaduras</i>
Sara	rubio	media	bajo	no	sí
Diana	rubio	alta	medio	sí	no
Alejandro	castaño	baja	medio	sí	no
Ana	rubio	baja	medio	no	sí
Emilio	pelirrojo	media	alto	no	sí
Pedro	castaño	alta	alto	no	no
Juan	castaño	media	alto	no	no
Carolina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

6. Interpreta el siguiente árbol de clasificación generado por WEKA que clasifica animales en (pájaros, mamíferos, peces, reptiles, anfibios, insectos e invertebrados). ¿Qué variables son más relevantes? ¿Qué tipos de animales le cuesta más al árbol discriminar?

Ejercicios IA2 – Aprendizaje Automático



7. Interpreta el siguiente árbol de clasificación generado por Scikit Learn que clasifica las especies de la flor de 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] y la clase del nodo (la mayoritaria en la distribución de muestras) que además viene también reflejada por el color del nodo (siendo el blanco la indecisión). ¿Qué variables juegan un papel más relevante? ¿Qué nodos consideras a tu juicio candidatos para ser podados? ¿Por qué?



Ejercicios IA2 – Aprendizaje Automático

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

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

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 de entrenamiento	Error sobre conjunto de 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 variables de entrada)	5.51	5.64
k-NN (k=8, 2 variables de entrada)	5.54	5.62