

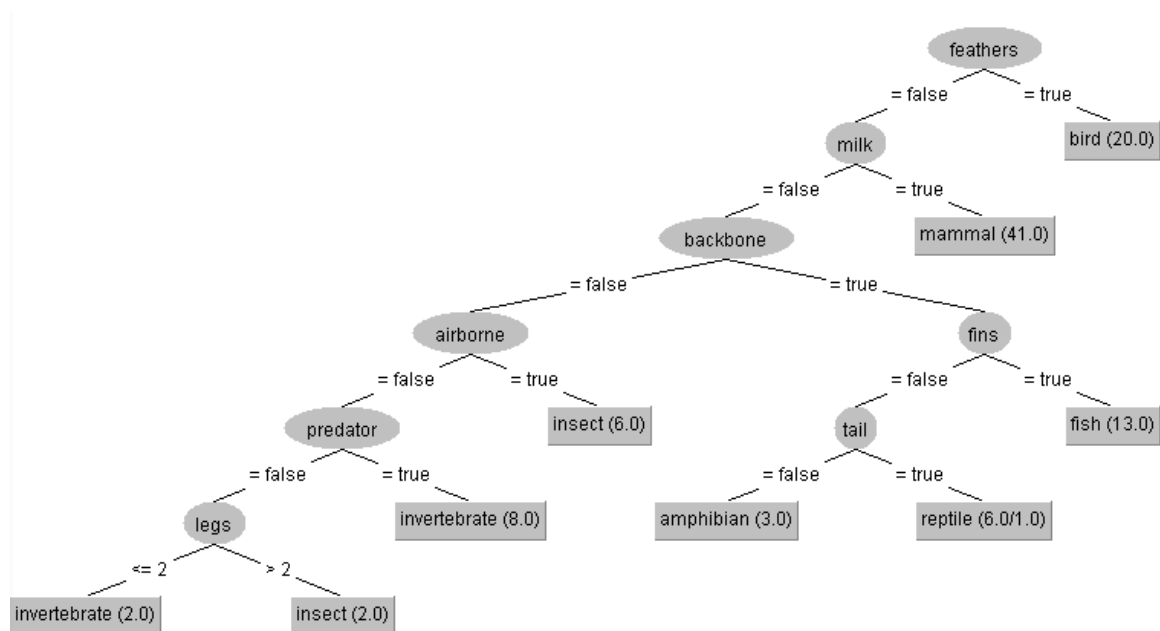
Ejercicios IA 2018-2019

Tema 6

1) 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>
<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>

2) 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?

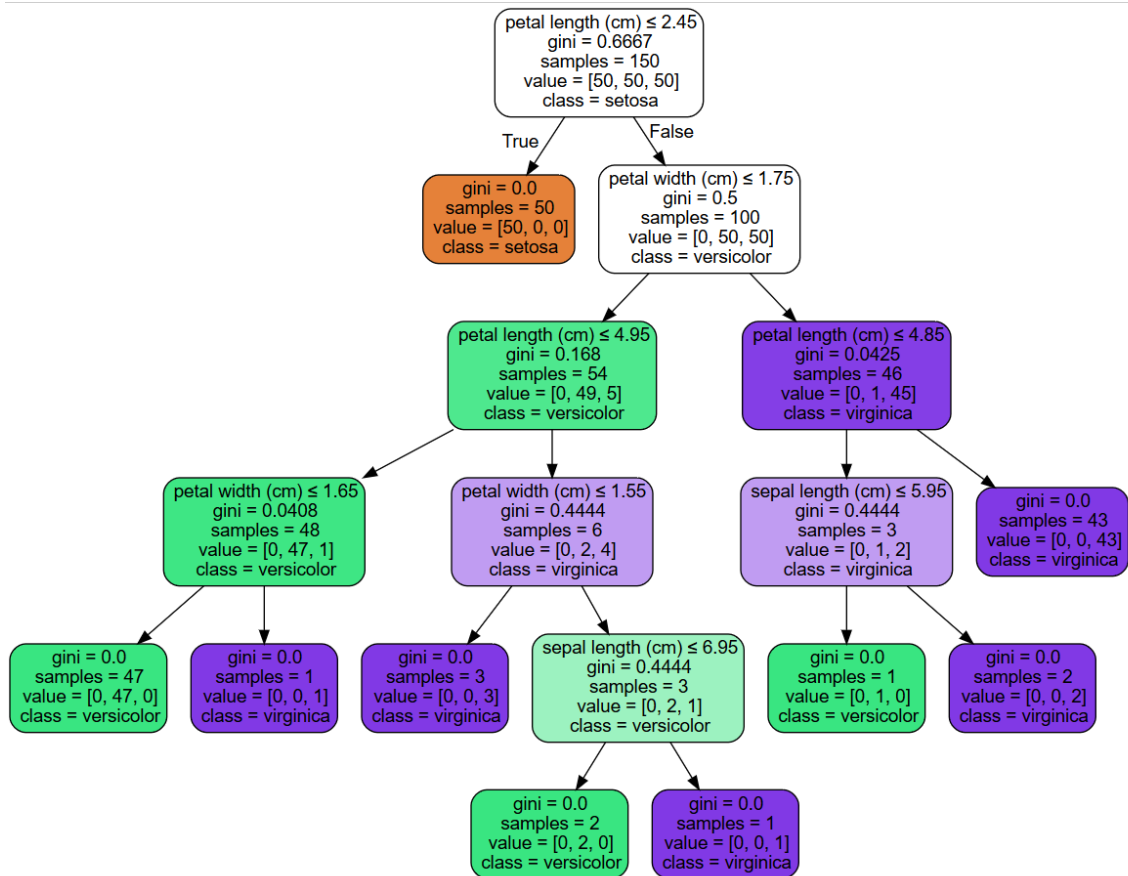


3) Interpreta el siguiente árbol de clasificación generado por Scikit Learn que clasifica las especies de la flor de Iris, según el ancho y el largo del pétalo y el sépalo de la flor. Observa que la representación de Scikit Learn es más rica ya que indica el coeficiente de Gini de cada nodo, el número 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).

Ejercicios IA 2018-2019

Tema 6

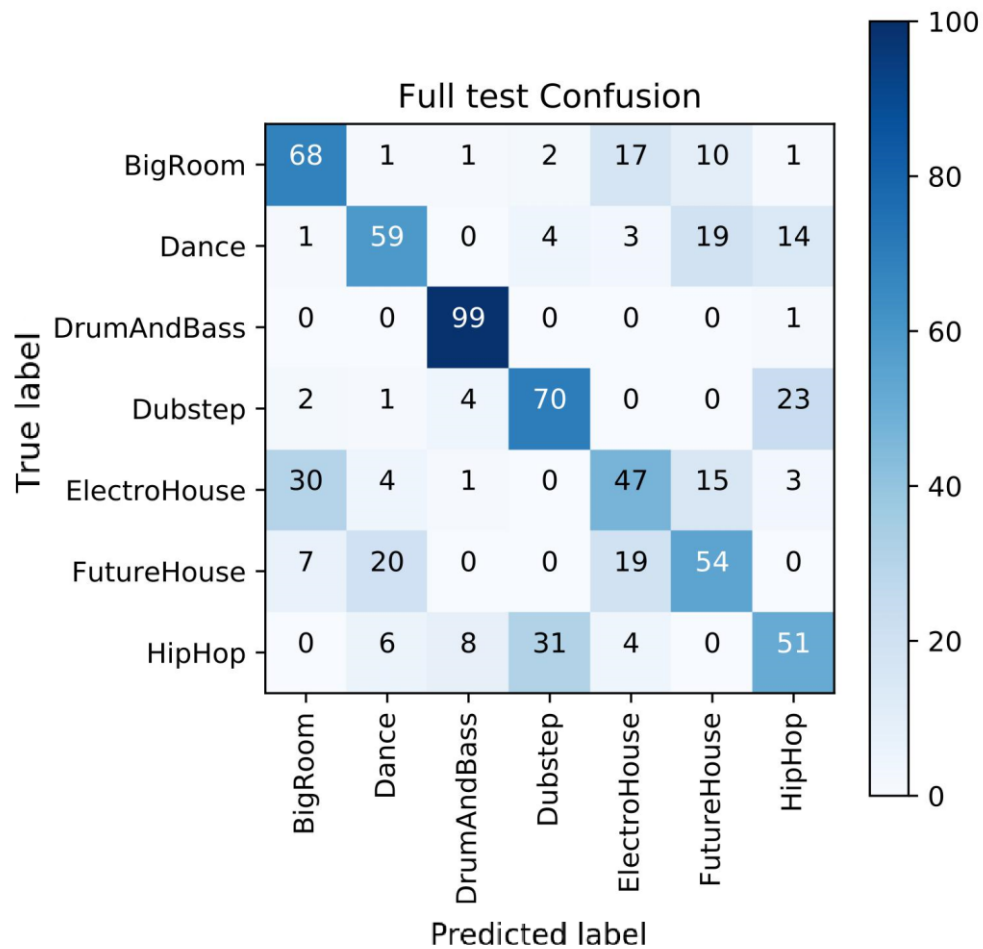
¿Qué variables juegan un papel más relevante? ¿Qué nodos consideras a tu juicio candidatos para ser podados? ¿Por qué?



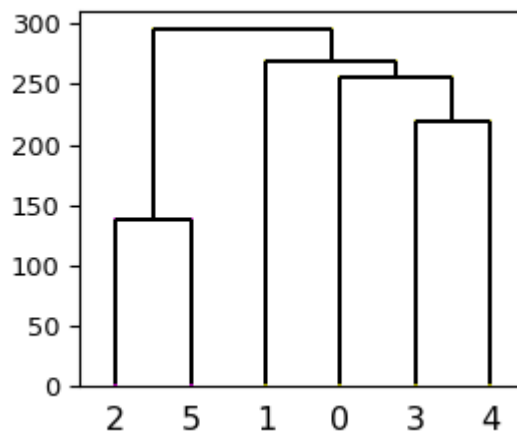
4) Interpreta la siguiente matriz de confusión que ha arrojado un algoritmo de clasificación en Scikit Learn. El objetivo es clasificar canciones según el subgénero de música electrónica al que pertenecen. Son 7 subgéneros y se dispone de 100 canciones por subgénero. ¿Cuál es la precisión global del 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 IA 2018-2019

Tema 6

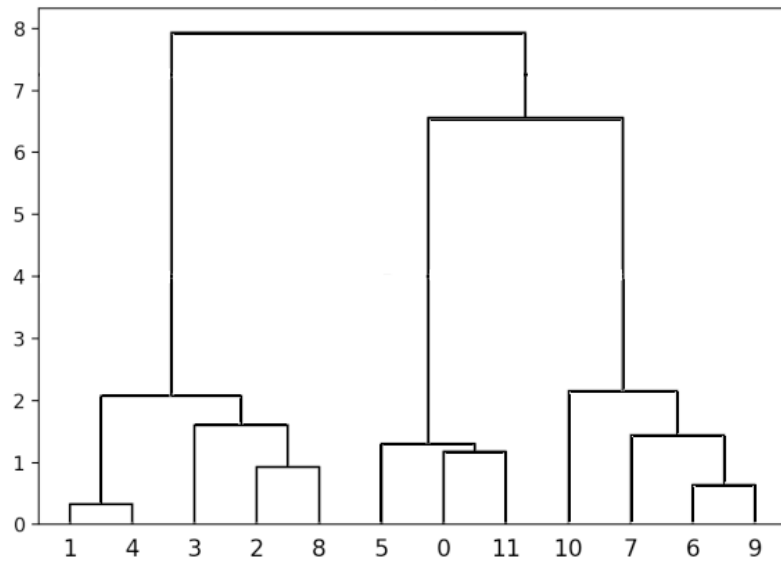


5) Interpreta los siguientes dendrogramas. ¿Cuántos clusters consideras que hay en el conjunto de datos? ¿Qué puedes decir de cada uno de ellos?



Ejercicios IA 2018-2019

Tema 6



6) 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

7) 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