

## EJERCICIOS DE APRENDIZAJE POR ÁRBOLES DE DECISIÓN

### Algoritmo ID3 y C4.5

1. Construir el árbol de decisión basándose en el algoritmo ID3. Probar el algoritmo con los datos de la tabla 1.

Obten las reglas que definen la descripción aprendida a partir del árbol generado.

¿Qué pasaría si en el ejemplo 1 que aparece en la primera fila de la tabla 1 se cambiara el valor `largo` por `¬largo`? Muestra el resultado del algoritmo y analiza las alternativas para resolver este problema.

	Color	Aspecto	Tamaño	Longitud	Clase
1	amarillo	moteado	grande	largo	venenoso
2	¬amarillo	¬moteado	¬grande	¬largo	venenoso
3	¬amarillo	moteado	¬grande	¬largo	venenoso
4	amarillo	moteado	grande	¬largo	comestible
5	¬amarillo	¬moteado	¬grande	largo	comestible
6	amarillo	¬moteado	¬grande	largo	venenoso
7	¬amarillo	moteado	grande	largo	comestible
8	amarillo	¬moteado	grande	¬largo	venenoso
9	¬amarillo	moteado	grande	¬largo	comestible
10	amarillo	moteado	¬grande	¬largo	venenoso

**Tabla 1** Clasificación de setas.

2. Se desea determinar las reglas que rigen el comportamiento de un alumno durante su vida académica para poder detectar casos que presenten riesgo de abandonar los estudios: para ello se dispone de los siguientes datos:

Nivel cultural padres	Capacidad cálculo	Trabaja	Asiste tutorías	Clase
Graduado escolar	baja	no	sí	Abandona
Graduado escolar	baja	sí	sí	Abandona
Graduado escolar	alta	sí	sí	Termina
Bachillerato	baja	no	sí	Abandona
Universitario	alta	no	sí	Termina
Universitario	alta	sí	sí	Termina

- a) Aplica el algoritmo ID3 de Quinlan a los datos que aparecen en la tabla anterior.
- b) Traduce ese árbol de decisión a reglas que me digan únicamente cuándo un alumno va a abandonar sus estudios
- c) Explicar cómo harías el entrenamiento por validación cruzada (o cross-validation)
- d) ¿Qué predicción harías para un alumno de padres universitarios que tiene una capacidad de cálculo baja, trabaja y asiste con regularidad a tutorías mediante el árbol de decisión obtenido?

3. Construir el árbol de decisión basándose en el algoritmo ID3 y en el algoritmo C4.5 Probar el algoritmo con los datos de la tabla 2.

<b>Pronóstico</b>	<b>Temperatura</b>	<b>Humedad</b>	<b>Viento</b>	<b>¿Adecuado?</b>
Soleado	Alta	Alta	Flojo	No
Soleado	Alta	Alta	Fuerte	No
Nublado	Alta	Alta	Flojo	Si
Lluvia	Moderada	Alta	Flojo	Si
Lluvia	Baja	Normal	Flojo	Si
Lluvia	Baja	Normal	Fuerte	No
Nublado	Baja	Normal	Fuerte	Si
Soleado	Moderada	Alta	Flojo	No
Soleado	Baja	Normal	Flojo	Si
Lluvia	Moderada	Normal	Flojo	Si
Soleado	Moderada	Normal	Fuerte	Si
Nublado	Moderada	Alta	Fuerte	Si
Nublado	Alta	Normal	Flojo	Si
Soleado	Moderada	Alta	Fuerte	No

**Tabla 2** Días adecuados para jugar al tenis

4. En unos laboratorios de un hospital se está investigando sobre una sustancia para la curación de una determinada enfermedad. Dicha sustancia ha sido inyectada en varias cobayas enfermas para verificar sus efectos. Los resultados de las pruebas realizadas aparecen en la siguiente tabla:

Estado de la enfermedad	Concentración de sustancia (ppm)	Número de dosis	Condición física	Efectos
Incipiente	75	70	Fuerte	Curación
Incipiente	80	90	Fuerte	Defunción
Incipiente	85	85	Débil	Defunción
Incipiente	72	95	Débil	Defunción
Incipiente	69	70	Débil	Curación
Avanzado	72	90	Fuerte	Curación
Avanzado	83	78	Débil	Curación
Avanzado	64	66	Fuerte	Curación
Avanzado	81	75	Débil	Curación
Terminal	71	80	Fuerte	Defunción
Terminal	65	70	Fuerte	Defunción
Terminal	75	80	Débil	Curación
Terminal	68	80	Débil	Curación
Terminal	70	96	Débil	Curación

Se desea determinar las reglas que rigen las condiciones en que ha de suministrarse la sustancia para que el paciente se cure de la enfermedad. Para ello:

- Construye un árbol de decisión basándose en el algoritmo ID3
- Traduce ese árbol de decisión a reglas, que permitirán obtener reglas generales a partir de los experimentos que ha realizado el hospital con las cobayas.
- ¿Cómo clasificarías un experimento en el que se han suministrado 87 dosis de una sustancia cuya concentración es de 82 ppm, a un paciente de condición física débil mediante el árbol de decisión obtenido?

**AYUDA:** En la tabla existen dos atributos, concentración y dosis, cuyos valores son continuos. Una forma de convertir estos valores continuos en discretos es introduciendo intervalos que irán desde el valor mínimo hasta el máximo. Es decir, debes coger la anchura del intervalo total para cada atributo (valor máximo – valor mínimo) y dividir ese intervalo en  $K$  partes iguales (en nuestro caso usaremos tres subintervalos). El resultado nos dará la anchura de los subintervalos (todos de la misma anchura) para ese atributo. Luego asignas una etiqueta lingüística a cada subintervalo (por ejemplo alta, baja, ...) y tendrás así los valores discretos.