

Apellidos :

Nombre :

Sólo se corregirán los ejercicios escritos a **bolígrafo azul** o **bolígrafo negro**

Una tienda quiere hacer un estudio de para determinar las características de sus posibles clientes. En una primera fase deciden considerar las cuatro características siguientes: edad (joven o adulto), nivel de ingresos (altos, medios o bajos), profesional del sector (si o no) y sexo (hombre o mujer) y encuesta a 21 personas, obteniendo los resultados que se reflejan en la siguiente tabla:

Ej.	Edad	Ingresos	Profesional	Sexo	Interesado
1	joven	bajos	si	hombre	si
2	joven	altos	si	hombre	si
3	joven	altos	no	mujer	no
4	joven	bajos	si	mujer	si
5	joven	medios	no	mujer	no
6	adulto	altos	si	hombre	no
7	adulto	altos	no	mujer	no
8	adulto	altos	si	mujer	no
9	adulto	medios	no	mujer	no
10	adulto	bajos	si	mujer	no
11	adulto	medios	no	mujer	no
12	adulto	medios	si	hombre	no
13	adulto	altos	no	hombre	si
14	joven	altos	si	mujer	si
15	joven	medios	si	hombre	si
16	adulto	medios	no	hombre	no
17	adulto	bajos	no	hombre	no
18	joven	medios	no	hombre	no
19	joven	bajos	no	mujer	no
20	adulto	medios	si	mujer	no
21	joven	medios	si	mujer	si

Ejercicio 1. [0.75 ptos.]

Aplicar el algoritmo ID3 (desarrollándolo paso a paso) para obtener un árbol de decisión que permitir describir el tipo de cliente interesado en los productos que ofrece la tienda.

Nota: Tomar como conjunto de entrenamiento los primeros 15 ejemplos de la tabla.

Ejercicio 2. [1 pto.] Tomando como conjunto de prueba los ejemplos del 16 al 21 de la tabla, calcular el rendimiento del árbol de decisión obtenido en el apartado anterior. Usando ese conjunto de prueba, aplicar (detallando los pasos) un proceso de *podado a posteriori* sobre el árbol de decisión. Expresar mediante reglas el árbol obtenido tras la poda.

1. ¿Qué rendimiento tiene este árbol sobre el conjunto de prueba?
2. ¿Y sobre el conjunto de entrenamiento?

Ejercicio 3. [1 pto.] Usando ahora **todos los ejemplos de la tabla** como conjunto de entrenamiento, predecir mediante un clasificador *naïve bayes* si un hombre joven con ingresos bajos que no es deportista profesional, va a estar interesado en los productos de la empresa.

Comparar el resultado con el ofrecido por los dos árboles de decisión anteriores.

Ejercicio 4. [0.75 ptos.] Describir el procedimiento de validación cruzada e indicar para qué se utiliza.