



Decision Trees. Exercises

Machine Learning I

Ejercicio DT 1

Se tiene el siguiente conjunto de ejemplos pertenecientes a dos clases A y B

	X	Y	Z	Clase
E1	1	1	0	A
E2	0	0.5	0	A
E3	1	1	1	B
E4	0	0	1	A
E5	0	0	0	B

Si se utiliza un algoritmo de aprendizaje automático ID3 para clasificar los ejemplos anteriores, se pide:

1. Indicar qué atributo sería el primer mejor separador posible. Justificar la respuesta con todos los cálculos necesarios.
2. Dibujar el árbol de decisión resultante tras elegir el primer separador, indicando sus nodos, las conexiones entre éstos y los ejemplos que contienen.
3. Cuantificar el desorden de cada nodo hijo resultante en el árbol tras elegir el primer separador.
4. Indicar la clase que representa cada nodo hijo en el árbol de decisión tras aplicar el primer separador.
5. Indicar cuál es la tasa de error de clasificación en el árbol tras elegir el primer separador. (Tasa de error = Número de ejemplos mal clasificados/Número de ejemplos total)
6. Si el árbol de decisión se hubiese desarrollado hasta el nivel de elección del primer separador ¿a qué clase pertenecería el ejemplo $X=1, Y=0, Z=1$?
7. Si el ejemplo $X=1, Y=0, Z=1$ se hubiese integrado en el árbol dibujado en el punto 2 en el nodo perteneciente a la clase contestada en el punto 6, graduar de forma cualitativa, sin hacer ningún cálculo, el desorden de los nodos hijos del árbol del punto 2.

Ejercicio DT 2

Se ha construido un árbol de decisión utilizando el algoritmo de ID3. El conjunto de entrenamiento utilizado constaba de tres clases de ejemplos A, B y C. La proporción de ejemplos de la clase A era el doble que la de B y la de C el triple que la de B.

Los ejemplos del conjunto de entrenamiento estaban caracterizados por 3 atributos: AT1, AT2 y AT3. Todos los atributos tenían dos valores posibles.

Resultó que el primer mejor separador a escoger fue el Atributo AT2 que generó un nodo completamente puro con los ejemplos de la clase B.

El siguiente mejor separador del árbol fue el Atributo AT3 que generó un nodo completamente puro con 2/3 de los ejemplos de la clase C.

Finalmente al aplicar como tercer mejor separador el Atributo AT1, se separan en nodos puros todas las clases.

Expresando todos los cálculos requeridos, se pide:

- a. ¿Cuál era el grado de desorden o entropía del nodo raíz del árbol?
- b. ¿Cuál fue el decremento de entropía o desorden que provocó la aplicación como separador del Atributo AT2?
- c. ¿Cuál fue el decremento de entropía o desorden que provocó la aplicación como separador del Atributo AT3?
- d. Representar el árbol de decisión que se generó.

Ejercicio DT 3

Una entidad de crédito desea definir el perfil de sus clientes con objeto de evitar problemas de morosidad en el pago de los créditos que concede. Para ello revisa una serie de características de sus clientes y selecciona los casos siguientes con su descripción:

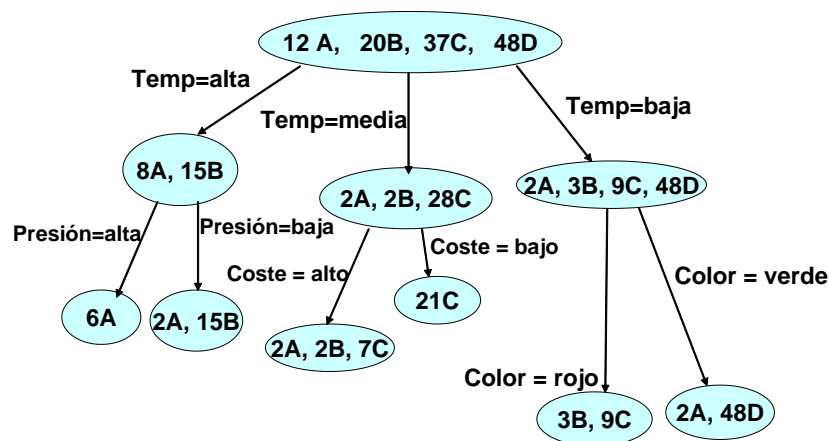
	¿Usa Internet?	¿Paga puntualmente el crédito?	Nivel de ingresos anuales	Número de hijos
Cliente 1	Sí	Sí	> 40000 y ≤ 90000 €	Ninguno
Cliente 2	Sí	Sí	> 90000 €	Entre 1 y 3
Cliente 3	Sí	No	≤ 40000 €	Ninguno
Cliente 4	No	Sí	> 90000 €	4 o más
Cliente 5	No	No	> 40000 y ≤ 90000 €	Ninguno
Cliente 6	Sí	Sí	≤ 40000 €	Entre 1 y 3

Si se utiliza un algoritmo de aprendizaje automático ID3 para conseguir el objetivo que busca alcanzar la entidad financiera, se pide:

1. Desarrollar completamente el árbol justificando la respuesta con todos los cálculos necesarios.
2. ¿Qué conocimiento se induce del árbol? ¿Un cliente con hijos sería moroso?
3. Si se hubiese utilizado como algoritmo de aprendizaje uno basado en C4.5 ¿habría cambiado el árbol desarrollado en relación con el obtenido por ID3? Razonar la respuesta indicando los cálculos y cambios, si los hubo.

Ejercicio DT 4

Dado el árbol de decisión de la figura ¿cuál fue la ratio de ganancia cuando se eligió el separador color?



Ejercicio DT 5

Un robot industrial ha de clasificar de la manera más eficiente posible en un almacén dos tipos de objetos: Redondos y Cuadrados. Para ello el robot cuenta con tres descriptores de características de esos objetos: color, anchura y longitud. Se cuenta con los siguientes 6 ejemplos con los que se quiere enseñar al robot cómo son los tipos de objetos que ha de clasificar:

	Color	anchura	longitud	Tipo de Objeto
E1	azul	a	3	Redondo
E2	azul	b	1	Cuadrado
E3	azul	a	2	Cuadrado
E4	verde	c	4	Redondo
E5	verde	a	3	Cuadrado
E6	azul	b	2	Redondo

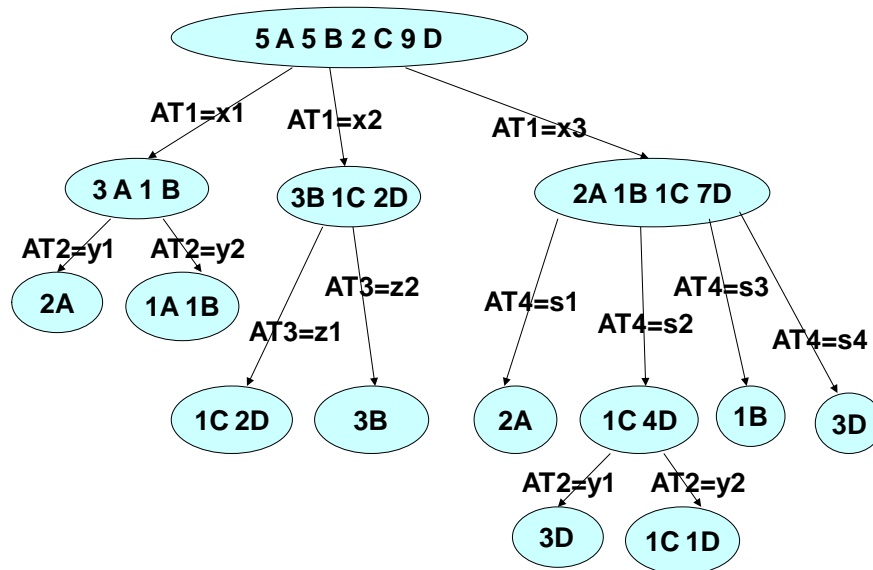
Si se utiliza un algoritmo de aprendizaje automático ID3 para clasificar los ejemplos anteriores, se pide:

1. Desarrollar completamente el árbol justificando la respuesta con todos los cálculos necesarios.
2. ¿Qué conocimiento define al objeto tipo Cuadrado?
3. ¿Cuáles fueron las razones de desarrollar el algoritmo C4.5?

- Si se hubiese utilizado como algoritmo de aprendizaje uno basado en C4.5 ¿habría cambiado el árbol desarrollado en relación con el obtenido por ID3? Razonar la respuesta indicando los cálculos y cambios, si los hubo.

Ejercicio DT 6

Supóngase que el árbol de decisión que se presenta a continuación fue obtenido mediante un algoritmo C4.5 que realizó la clasificación de los ejemplos de un conjunto de entrenamiento pero al que aún no se le ha aplicado la poda de dicho algoritmo:



En cada nodo se indica el número de casos de cada clase (Las clases son A, B, C y D)

- ¿Qué conocimiento se induciría del árbol mostrado?
- ¿A qué clase pertenecería un ejemplo con los siguientes valores de atributos: $AT1=x3$, $AT2=y2$, $AT3=z1$ y $AT4=s2$?
- ¿Cómo quedará el árbol anterior al aplicar el algoritmo de poda de C4.5? Justificar la respuesta y expresar todos los cálculos realizados para llegar a la misma.
- ¿Cuál fue la ratio de ganancia del mejor separador en este árbol desarrollado con C4.5? Justificar la respuesta y expresar todos los cálculos realizados para llegar a la misma.

Ejercicio DT 7

Se desea construir un sistema de alarmas contra inundaciones y para ello se ha recabado la siguiente información de un archivo histórico de datos

Caso	Presión (mbar)	Temperatura (°C)	Velocidad del viento (Km/h)	Dirección del viento	% Intensidad solar	Peligro inundación
E1	baja	30	20	N	10	Alto
E2	baja	10	20	O	30	Alto
E3	baja	30	0	N	30	Medio
E4	alta	20	10	E	30	Medio
E5	alta	20	10	O	10	Bajo
E6	alta	30	0	N	30	Bajo

Si se utiliza un algoritmo de aprendizaje automático ID3 para clasificar los ejemplos anteriores, se pide:

- Desarrollar completamente el árbol justificando la respuesta con todos los cálculos necesarios.
- ¿Qué conocimiento se habría aprendido tras realizar el árbol?
- Si se hubiese utilizado como algoritmo de aprendizaje uno basado en C4.5 ¿habría cambiado el árbol desarrollado en relación con el obtenido por ID3? Razonar la respuesta indicando los cálculos y cambios, si los hubo.
- A partir de los resultados obtenidos en los apartados anteriores ¿qué variables son necesarias para predecir el peligro de inundación y cuáles son superfluas? Razonar la respuesta.

Ejercicio DT 8

Un proceso industrial permite la elaboración de tres tipos de material cerámico: A, B y C. Este proceso industrial se fundamenta básicamente en elegir un tipo de tratamiento, una temperatura y un color de aditivo. Se cuenta con los siguientes 6 ejemplos de elaboración del material cerámico:

	Tratamiento	Temperatura	Color	Material cerámico
E1	reducción	media	azul	A
E2	reducción	baja	verde	C
E3	reducción	media	rojo	C
E4	oxidación	alta	azul	B
E5	oxidación	baja	rojo	B
E6	oxidación	media	verde	A

Si se utiliza un algoritmo de aprendizaje automático ID3 para clasificar los ejemplos anteriores, se pide:

1. Desarrollar completamente el árbol justificando la respuesta con todos los cálculos necesarios.
2. ¿Qué conocimiento se habría aprendido tras realizar el árbol?
3. Si se hubiese utilizado como algoritmo de aprendizaje uno basado en C4.5 ¿habría cambiado el árbol desarrollado sólo hasta el primer mejor separador en relación con el obtenido por ID3? Razonar la respuesta indicando los cálculos y cambios, si los hubo.

Ejercicio DT 9

Dado el siguiente conjunto de ejemplos:

	X1	X2	Salida
E1	1	10	X
E2	1	30	X
E3	2	10	X
E4	2	20	O
E5	2	30	O
E6	3	10	O
E7	4	30	X

Se pide:

1. Representar los ejemplos en el plano. Desarrollar completamente el árbol que generaría CART con el criterio de Gini (Mostrar cálculos).
2. Si se considera un $\alpha=0,1$ ¿Cómo quedaría el árbol tras la poda? ¿Y si se considera un $\alpha=0,15$? (Mostrar cálculos).
3. Desarrollar el primer nivel del árbol que generaría ID3 (Mostrar cálculos).

Ejercicio DT 10

Sea el siguiente conjunto de entrenamiento de 7 ejemplos descritos a través de 4 descriptores A, B, C y D de valores respectivos $A=\{\leq 1, >1\}$, $B=\{\leq 10, >10\}$, $C=\{\leq 100, >100\}$, $D=\{R, V\}$. Los ejemplos pueden pertenecer a una de dos clases diferentes etiquetadas como X y O:

	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>	<u>Clase</u>
E1	≤ 1	≤ 10	≤ 100	V	X
E2	> 1	> 10	> 100	R	X
E3	≤ 1	> 10	≤ 100	V	O
E4	≤ 1	≤ 10	> 100	V	O
E5	≤ 1	≤ 10	≤ 100	R	X
E6	> 1	≤ 10	≤ 100	V	X
E7	> 1	≤ 10	> 100	V	O

1. Desarrollar mediante el algoritmo de CART un árbol de decisión con el conjunto de entrenamiento hasta el nivel que se crea conveniente sabiendo que el parámetro α para evaluación del coste complejidad vale 0.12. Indicar todos los cálculos realizados.
2. Desarrollar mediante el algoritmo de C4.5 un árbol de decisión con el conjunto de entrenamiento hasta el nivel que se crea conveniente. Indicar todos los cálculos realizados.

Ejercicio DT 11

Una entidad de crédito desea definir el perfil de sus clientes con objeto de evitar problemas de morosidad en el pago de los créditos que concede. Para ello revisa una serie de características de sus clientes y selecciona los casos siguientes con su descripción:

	<u>¿Usa Internet?</u>	<u>¿Paga puntualmente el crédito?</u>	<u>Número de hijos</u>
Ciente 1	Sí	Sí	Ninguno
Ciente 2	Sí	Sí	Entre 1 y 3
Ciente 3	Sí	No	Ninguno
Ciente 4	No	Sí	4 o más
Ciente 5	No	No	Ninguno
Ciente 6	Sí	Sí	Entre 1 y 3

Si se utiliza un algoritmo de aprendizaje automático CART para conseguir el objetivo que busca alcanzar la entidad financiera, desarrollar el árbol hasta el nivel que se crea conveniente sabiendo que $\alpha=0.3$. Justificar la respuesta con todos los cálculos necesarios.

Ejercicio DT 12

Dado el siguiente conjunto de entrenamiento de 7 ejemplos descritos a través de 3 descriptores X_1 , X_2 y X_3 de valores respectivos $X_1=\{a, b\}$, $X_2=\{1, 15, 25\}$, $X_3=\{T, F\}$. Los ejemplos pueden pertenecer a una de dos clases diferentes etiquetadas como X y O:

	<u>X_1</u>	<u>X_2</u>	<u>X_3</u>	<u>Clase</u>
E1	a	15	T	X
E2	a	1	T	O
E3	b	15	F	X
E4	b	25	T	O
E5	a	15	F	X
E6	b	1	T	X
E7	a	25	F	X

1. Desarrollar mediante el algoritmo de CART un árbol de decisión con el conjunto de entrenamiento sin aplicar poda. Indicar los cálculos y detalles de la elaboración del árbol.
2. Si se utilizase ahora el algoritmo de poda y se empezase a analizar el árbol desarrollado desde arriba, ¿cómo quedaría si se usase un valor de 0.1 para el parámetro α ? En el mismo supuesto de antes ¿qué valor debería de tener ese parámetro para el caso en que se deseara conservar el árbol completo desarrollado en el apartado 1? Indicar todos los cálculos realizados.

Ejercicio DT 13

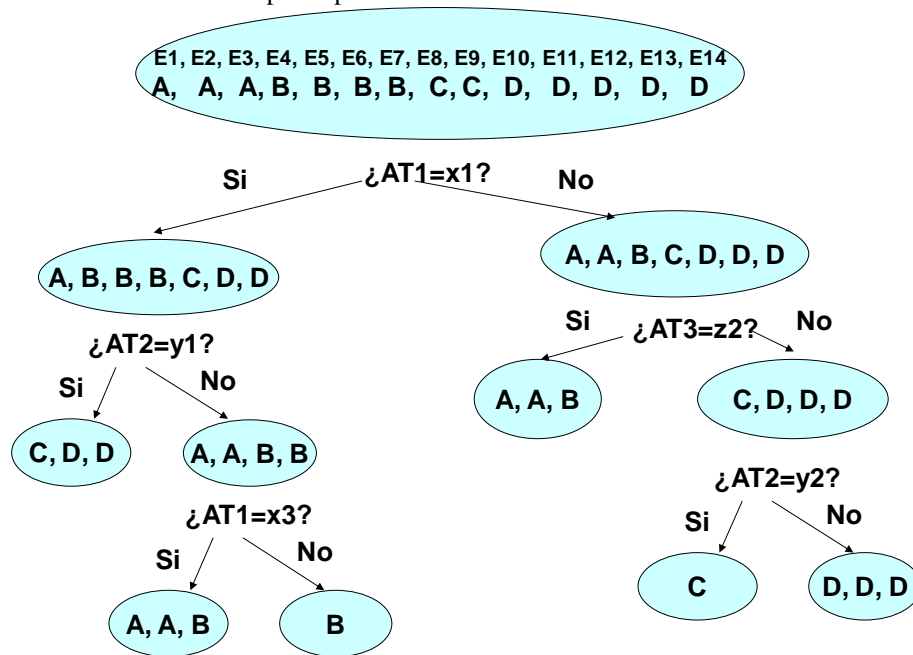
Dado el siguiente conjunto de entrenamiento de 7 ejemplos descritos a través de 3 descriptores X_1 , X_2 y X_3 de valores respectivos $X_1=\{a, b\}$, $X_2=\{10, 20, 30\}$, $X_3=\{\alpha, \beta\}$. Los ejemplos pueden pertenecer a una de dos clases diferentes etiquetadas como X y O:

	<u>X_1</u>	<u>X_2</u>	<u>X_3</u>	<u>Clase</u>
E1	a	30	α	X
E2	a	10	α	O
E3	b	20	β	X
E4	b	30	α	O
E5	a	20	β	X
E6	b	10	α	O
E7	a	30	β	X

1. Desarrollar mediante el algoritmo de CART un árbol de decisión con el conjunto de entrenamiento sin aplicar poda. Indicar los cálculos y detalles de la elaboración del árbol.
2. Si se empezase desde abajo del árbol aplicándosele el algoritmo de poda de CART al árbol de decisión desarrollado en el punto anterior ¿qué valor debería de tener el parámetro α para conservar el árbol sin podar? Indicar todos los cálculos realizados.

Ejercicio DT 14

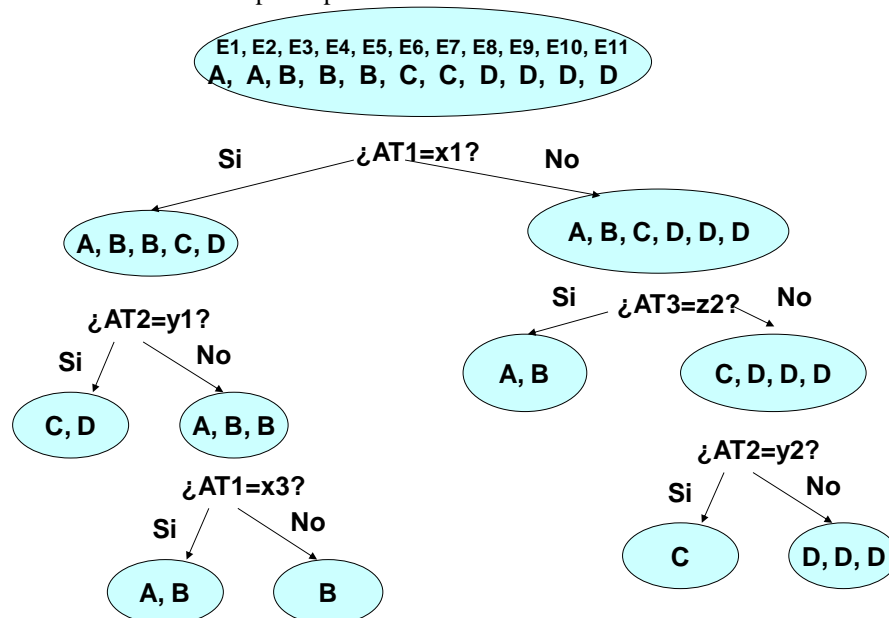
Se realizó la clasificación de los ejemplos de un conjunto de entrenamiento mediante un algoritmo CART resultando el siguiente árbol de clasificación sin aplicar poda:



1. Si el algoritmo de poda usase un valor de $\alpha=0.12$ y analizase el árbol de arriba a abajo ¿Cuál sería el árbol resultante? Justificar la respuesta y expresar todos los cálculos realizados para llegar a la misma.
2. ¿Cuál es la entropía resultante tras la elección del mejor separador en este árbol desarrollado con CART? Justificar la respuesta y expresar todos los cálculos realizados para llegar a la misma.

Ejercicio DT 15

Se realizó la clasificación de los ejemplos de un conjunto de entrenamiento mediante un algoritmo CART resultando el siguiente árbol de clasificación sin aplicar poda:



1. Si el algoritmo de poda usase un valor de $\alpha=0.03$ y analizase el árbol de arriba a abajo ¿Cuál sería el árbol resultante? Justificar la respuesta y expresar todos los cálculos realizados para llegar a la misma.
2. ¿Cuál es la entropía resultante tras la elección del mejor separador en este árbol desarrollado con CART? Justificar la respuesta y expresar todos los cálculos realizados para llegar a la misma.

