

MÁSTER UNIVERSITARIO EN
LÓGICA, COMPUTACIÓN E INTELIGENCIA ARTIFICIAL
Aprendizaje Automático

Apellidos:

Nombre :

Cuestión 1. ¿Cuántas posibles hipótesis podemos representar mediante árboles de decisión si tenemos n atributos binarios?

Cuestión 2. Definimos el tamaño de un árbol como el número de sus nodos, incluida las hojas. Dar dos árboles de tamaño distinto que representen la misma hipótesis.

Cuestión 3. Dado el conjunto de entrenamiento

Cielo	Temperatura	Humedad	Viento	Agua	Previsión	Deporte
Soleado	Templada	Normal	Fuerte	Templada	Igual	Sí
Soleado	Templada	Alta	Fuerte	Templada	Igual	Sí
Lluvia	Fría	Alta	Fuerte	Templada	Cambio	No
Soleado	Templada	Alta	Fuerte	Fría	Cambio	Sí

y como conjunto de hipótesis \mathcal{H} el conjunto de *todos* los árboles de decisión, dar dos elementos del espacio de versiones VS . Uno de ellos debe clasificar de manera positiva y otro de manera negativa el ejemplo

$\langle Lluvia, Templada, Normal, Fuerte, Templada, Cambio \rangle$

Cuestión 4. En un problema de aprendizaje sobre el que vamos a aplicar el algoritmo ID3 se considera un conjunto de entrenamiento D con N ejemplos. Supongamos que hay un atributo Atr_1 que puede tomar N valores y que cada uno de los ejemplos de D toma en Atr_1 un valor distinto. Calcular $Ganancia(D, Atr_1)$.

Cuestión 5. Demostrar que si D es un conjunto de entrenamiento para un concepto C entonces el algoritmo ID3 devuelve una hipótesis consistente con D cualquiera que sea el criterio para elegir el *mejor* atributo

Cuestión 6. Sea U un universo finito y sea C un concepto sobre U donde las instancias se representan como n -uplas de pares atributo-valor. Entonces C puede representarse como un árbol de decisión.

Cuestión 7. Responder razonadamente si las siguientes afirmaciones son verdaderas o falsas. Si la respuesta **Verdadera** debes dar razones que apoyen tu decisión. Si la respuesta es **Falsa** debes dar un ejemplo en el que no se verifique la afirmación.

- Sea $D_1 = \{e_1, \dots, e_n\}$ un conjunto de entrenamiento y sea D_2 el conjunto de entrenamiento formado a partir de D_1 donde cada ejemplo se considera dos veces, esto es, $D_2 = \{e_1, \dots, e_n, e_{n+1}, \dots, e_{2n}\}$ donde $\forall i \in \{1, \dots, n\} e_i = e_{i+n}$. Afirmación: *El árbol de decisión obtenido mediante el algoritmo ID3 a partir de D_1 y D_2 son el mismo.*
- Tenemos un conjunto con $4n$ ejemplos y lo dividimos en dos partes: un conjunto de entrenamiento con $3n$ ejemplos y un conjunto de prueba con n ejemplos. En el

conjunto de entrenamiento $2n$ ejemplos son positivos y n ejemplos son negativos. En el conjunto de prueba, los n ejemplos que hay son todos positivos. Construimos el árbol de decisión A a partir del conjunto de entrenamiento mediante el algoritmo ID3. Afirmación: *El árbol A contiene al menos un nodo tal que si podamos en ese nodo mejoramos el rendimiento respecto al conjunto de prueba.*

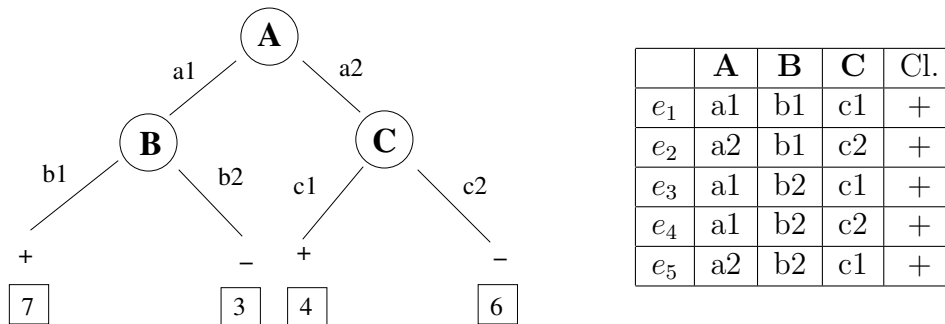
(c) Sean D_1 y D_2 dos conjuntos de entrenamiento para el mismo concepto. Entonces

$$\frac{Ent(D_1) + Ent(D_2)}{2} \leq Ent(D_1 \cup D_2)$$

(d) Sean D_1 y D_2 dos conjuntos de entrenamiento para el mismo concepto tales que $D_1 \cap D_2 \neq \emptyset$. Entonces $Ent(D_1 \cup D_2) < Ent(D_1) + Ent(D_2)$.

Cuestión 8. La ganancia de información en el algoritmo ID3 sólo sirve para hacer árboles más pequeños. Si aplicamos el algoritmo de árboles de decisión sobre un mismo conjunto de entrenamiento dos veces, la primera utilizando la ganancia de información para elegir el mejor atributo y la segunda vez utilizando otro criterio, entonces los dos árboles obtenidos pueden tener distinto tamaño, pero representan la misma hipótesis. **¿Verdadero o falso?**

Ejercicio 1. Supongamos que hemos dividido un conjunto de 25 ejemplos en dos conjuntos. El primero D con 20 ejemplos lo hemos usado para crear un árbol de decisión y el segundo $Prueba$, con 5 ejemplos, los vamos a usar para aplicar el ALGORITMO DE PODA PARA REDUCIR EL ERROR. El árbol obtenido y el conjunto de prueba son los siguientes:



Junto a la clasificación de cada hoja aparece el número de elementos del conjunto D que verifica la condición, esto es, hay 7 ejemplos con $A=a1$ y $B=b1$ que tienen clasificación +, hay 3 ejemplos con $A=a1$ y $B=b2$ que tienen clasificación -, etc. Se pide usar el ALGORITMO DE PODA PARA REDUCIR EL ERROR sobre el árbol usando el conjunto $Prueba$. **Especificar claramente cuál es el árbol obtenido.**

Ejercicio 2. Ejercicio opcional: Demostrar que la ganancia de información es siempre mayor o igual a cero.