

Práctico 2

Árboles de decisión – Evaluación de hipótesis

Ejercicio 1

Dé árboles de decisión que representen las siguientes funciones booleanas:

- $A \wedge \neg B$
- $A \vee [B \wedge C]$
- $A \text{ xor } B$
- $[A \wedge B] \vee [C \wedge D]$
- $\neg [A \wedge B]$

Ejercicio 2

Dado el siguiente conjunto de entrenamiento:

#	a1	a2	Clasif
1	Verdadero	Verdadero	Sí
2	Verdadero	Verdadero	Sí
3	Verdadero	Falso	No
4	Falso	Falso	Sí
5	Falso	Verdadero	No
6	Falso	Verdadero	No

- a) ¿Cuál es la entropía del conjunto de entrenamiento?
- b) ¿Cuál es la ganancia de particionar por cada uno de los atributos?

Ejercicio 3

Volviendo al problema de aprender bajo qué condiciones a Pedro le gusta ir a jugar al fútbol a la playa:

#	Cielo	Temp	Humedad	Viento	Tmp. Agua	Tiempo	Juega
1	Soleado	Templado	Normal	Fuerte	Templada	Sin cambios	Sí
2	Soleado	Templado	Alta	Fuerte	Templada	Sin cambios	Sí
3	Lluvioso	Frío	Alta	Fuerte	Templada	Cambiante	No
4	Soleado	Templado	Alta	Fuerte	Fría	Cambiante	Sí

Con los siguientes posibles valores para cada atributo:

- *Cielo*: Soleado, Lluvioso, Nublado
- *Temperatura*: Templado, Frío
- *Humedad*: Normal, Alta
- *Viento*: Fuerte, Suave
- *Temp. Agua*: Templada, Fría
- *Tiempo*: Sin cambios, Cambiante.

- a) Halle el árbol de decisión utilizando el algoritmo ID3.
- b) Halle el árbol de decisión, si ahora se agrega el siguiente ejemplo al conjunto de entrenamiento:

#	Cielo	Temp	Humedad	Viento	Tmp. Agua	Tiempo	Juega
5	Soleado	Templado	Normal	Suave	Templada	Sin Cambios	No

- c) ¿Qué respuesta daría a las siguientes instancias?

#	Cielo	Temp	Humedad	Viento	Tmp. Agua	Tiempo	Juega
6	Soleado	Templado	Normal	Fuerte	Fría	Cambiante	?
7	Lluvioso	Frío	Normal	Suave	Templada	Sin cambios	?
8	Soleado	Templado	Normal	Suave	Templada	Sin cambios	?
9	Soleado	Frío	Normal	Fuerte	Templada	Sin cambios	?

- d) ¿Pertenece la solución al espacio de versiones obtenido con el algoritmo Candidate-Elimination en el práctico anterior? ¿Es esto siempre esperable?

Ejercicio 4

Luego de ver árboles de decisión y aprendizaje conceptual, un alumno decide aplicar 'lo mejor de ambos mundos': utiliza el algoritmo de Candidate-Elimination sobre el espacio de hipótesis de los árboles de decisión.

- a) Encuentre los límites S y G, luego de procesar los dos primeros ejemplos del ejercicio 3.
- b) ¿Es una buena opción la elegida por este alumno? ¿Por qué?

Ejercicio 5

- a) Suponga que verifica una hipótesis h y encuentra que comete 300 errores en una muestra de 1000 ejemplos tomados al azar. ¿Cuál es la desviación estándar de $error_S(h)$?
- b) Suponga que obtuvo una hipótesis h para un concepto booleano. Cuando se la verifica sobre 100 ejemplos, a 83 los clasifica correctamente. ¿Cuál es la desviación estándar y el intervalo de 95% de confianza para $error_D(h)$?
- c) Dada una hipótesis h que comete 10 errores sobre un conjunto de 65 ejemplos tomados al azar:
 - ¿Cuál es el intervalo de 90% de confianza para el error verdadero?
 - ¿Cuál es el intervalo $[-\infty, U]$ de 95% de confianza para el error verdadero?
 - ¿Cuál es el intervalo de $[-\infty, U]$ de 90% de confianza para el error verdadero?

Ejercicio 6 [*]

- a) Implemente el algoritmo ID3 y aplíquelo sobre un conjunto de entrenamiento dado. De ser necesario, extienda el algoritmo para manejar los casos no contemplados por el algoritmo original.
- b) Separe 4/5 del conjunto de entrenamiento y evalúe su implementación utilizando validación cruzada de tamaño 10.
- c) Con el 1/5 no utilizado en la parte previa evalúe al resultado de entrenar con los 4/5 restantes.