

Práctico 2

Árboles de decisión

Ejercicio 1

Dé árboles de decisión que representen las siguientes funciones booleanas:

1. $p \wedge \neg q$
2. $p \text{ xor } q$
3. $p \vee [q \wedge r]$
4. $[p \wedge q] \vee [r \wedge s]$
5. $\neg [p \wedge q]$

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

Donde: Cielo \in {Soleado, Lluvioso, Nublado},
 Temperatura \in {Templado, Frío},
 Humedad \in {Normal, Alta},
 Viento \in {Fuerte, Suave},
 Temp. Agua \in {Templada, Fría},
 Tiempo \in {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

Se desea construir un algoritmo para clasificar automáticamente la calidad de la fruta, de acuerdo a ciertas variables: color, tamaño, peso y mes de cosecha, y le proveen del siguiente conjunto de ejemplos:

#	Color	Tamaño	Peso	Mes	Calidad
1	rojo	mediano	200	octubre	bueno
2	rojo	grande	150	octubre	mala
3	rojo	mediano	200	noviembre	mala
4	rojo	mediano	200		bueno
5	amarillo	grande	150	noviembre	bueno
6	amarillo	mediano	220	noviembre	mala

Donde: color \in {amarillo, rojo}, tamaño \in {grande, mediano, pequeño}, peso \in [150, 250], mes \in {octubre, noviembre}, calidad \in {bueno, mala}.

- Explique por qué no puede aplicar el algoritmo ID3 básico. Dé una solución a cada uno de los problemas encontrados.
- Dé un árbol de decisión a partir de los ejemplos dados, explicando su construcción paso a paso.

Ejercicio 5

Se decide evitar el sobreajuste del modelo generado, modificando al ID3 para que, luego de construido el árbol, aplique un procedimiento de poda.

El procedimiento consiste en evaluar a los nodos interiores, desde la raíz a las hojas, con un conjunto de validación. Cuando el resultado sobre el conjunto de validación es peor que simplemente predecir el valor más común de los ejemplos de entrenamiento en ese nodo, se elimina completamente el subárbol y se lo sustituye con una hoja con ese valor. En caso contrario, se procede a evaluar recursivamente a los nodos hijos.

Dé el árbol resultante de aplicar este método con los siguientes conjuntos de datos, donde las instancias 1 al 5 son de entrenamiento y el resto de validación.

#	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
5	Soleado	Templado	Normal	Suave	Templada	Sin Cambios	No
6	Soleado	Templado	Normal	Fuerte	Fría	Suave	Sí
7	Lluvioso	Frío	Normal	Suave	Templada	Sin cambios	No

Ejercicio 6

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.

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