

Asignatura: CI5128 – Ingeniería del conocimiento**Cuatrimestre: 2º****Examen: Parcial****Convocatoria: Ordinaria****Grupo: 5IM1****Curso: 2004/2005****Fecha: 17-3-2005**

NOTA: Se pueden utilizar todo tipo de libros y apuntes para la realización del examen

1. **(3 puntos)** Obtén los conjuntos S y G de fórmulas más específicas que se pueden obtener del siguiente conjunto de datos usando el espacio de versiones.

Objeto	A	B	C	D	E	Tipo
1	C	A	V	P	E	+
2	Z	A	I	M	A	-
3	S	A	V	P	E	+
4	K	B	I	P	E	-

- a) ¿Cómo resolverías este problema mediante un árbol de decisión? Para ello contesta a las siguientes preguntas:
- i) **(2 puntos)** ¿Cómo lo construirías usando validación cruzada con este conjunto de entrenamiento? ¿Cómo lo harías si usases 10-fold cross validation? Describe el proceso detalladamente, explicando primero en qué consiste cada uno de las técnicas y luego cómo lo harías en un experimento real (qué habría en cada uno de los ficheros de datos, cómo harías cada una de las fases del entrenamiento y qué es lo que haría el programa con cada uno de ellos en las diferentes fases)
 - ii) **(1.5 puntos)** Como sabes, éstos métodos se utilizan para evitar el sobreajuste. Explica con detalle qué es el sobreajuste. ¿Por qué un árbol podado evita el sobreajuste?
 - iii) **(1.5 puntos)** Construye el árbol de decisión (mediante C4.5) sin realizar ninguna cuenta y explica cómo lo has obtenido. ¿En qué ha afectado escoger el algoritmo C4.5 en vez del ID3 en este caso concreto?

2. **(2 puntos)** Supongamos que ampliamos nuestro lenguaje de descripción de manera que podemos describir conceptos y ejemplos con una sola disyunción en la que los disyuntandos no están ordenados. Por ejemplo, si tomamos el dominio de los objetos geométricos podemos expresar el ejemplo ((bola, rojo, grande) o (cubo, verde, pequeño)) que es equivalente a ((cubo, verde, pequeño) o (bola, rojo, grande)), donde ahora un nuevo ejemplo puede generalizar un concepto de dos maneras posibles: por ejemplo la instancia ((cubo, rojo, pequeño) o (cubo, verde, grande)) puede generalizar el ejemplo anterior como ((x1, rojo, x3) o (cubo, verde, x3)) o como ((x1, x2, grande) o (cubo, x2, pequeño)). Lo mismo pasa con la especialización. Supongamos que tenemos en cuenta este nuevo lenguaje y la forma en que se comporta ante la generalización y la especialización. Supongamos también que tomamos como dominio el de los coches en venta de un concesionario con los siguientes atributos:

Precio: caro, medio, barato.

Motor: diesel, gasolina.

Color: azul, verde, rojo.

País: Japón, Francia, USA, Alemania.

Tamaño: grande, mediano, pequeño.

¿Cuál sería la modificación del espacio de versiones que aparece a continuación, con el siguiente contraejemplo (ejemplo negativo): ((barato, diesel, verde, Francia, grande) (caro, diesel, verde, USA, grande))

$G = \{((x1, x2, x3, x4, x5) \text{ o } (x1, x2, x3, x4, x5))\}$

$S = \{((x1, diesel, x3, USA, grande) \text{ o } (caro, diesel, x3, USA, x4))\}$