



Asignatura: **CI5218 – Ingeniería del Conocimiento**
Cuatrimestre: **2º** Examen: **Final** Convocatoria: **Extraordinaria**
Grupo: **5IM1** Curso: **2005/2006** Fecha: **14/07/2006**

PREGUNTAS

1. Muchos métodos de análisis de datos se basan en una medida de distancia entre las distintas instancias.

Dado el conjunto de datos de la tabla :

Instancia	A1	A2	A3	A4
1	0,1	87	azul	13
2	0,12	14	azul	70
3	0,8	3	verde	12
4	0,6	90	verde	41
5	0,23	45,3	azul	93,8
6	0,97	62	verde	91

1. **(0,5 puntos)** Definir una medida de distancia adecuada para realizar agrupamiento o clustering sobre el conjunto de datos. ¿Qué método que utilice la medida de distancia anterior (o una similar) podría usarse para calcular el valor del atributo A3 para un ejemplo en el que dispones de los valores de los atributos A1, A2, y A4? Suponiendo que todos los atributos tuvieran que estar normalizados, ¿Sería útil/necesario algún tipo de preprocesado de los datos? ¿Cómo quedaría el conjunto de entrenamiento?
2. **(2 puntos)** Utiliza la medida de distancia y el algoritmo definidos en los apartados anteriores para calcular el valor del atributo A3 de una instancia que tiene los siguientes valores para el resto de atributos: A1 = 0, A2 = 56, A4 = 71.
3. **(2.5 puntos)** Discretiza el atributo A1 en dos intervalos teniendo en cuenta el cambio de valor correspondiente en el atributo A3.
4. **(2.5 puntos)** Utiliza un algoritmo de aprendizaje no supervisado para formar clusters jerárquicos, realizando las suposiciones que creas necesarias para que el problema tenga sentido según la tarea de agrupamiento que se pide.
5. **(2.5 puntos)** Resuelve el apartado 4 por medio de una red neuronal. Todos los parámetros son de libre elección (pesos sinápticos iniciales, umbrales, factor de aprendizaje,...)