



Universidad Nebrija

DII

Asignatura: **CI5128 – Ingeniería del conocimiento**

Cuatrimestre: **2º**

Examen: **Final**

Convocatoria: **Extraordinaria**

Grupo: **5IT1**

Curso: **2008/2009**

Fecha: **2-7-2009**

Alumno: _____

NOTA: Se puede usar todo tipo de libros y apuntes de manera personal e intransferible. La duración del examen es de 3 horas.

1. **(2 puntos)** En la actualidad los partidos de baloncesto se desarrollan en cuatro tiempos de 10 minutos cada uno. Pretendemos desarrollar un clasificador para predecir si un equipo va a ganar un partido concreto. Para ello nos vamos a fijar exclusivamente en los resultados de los dos primeros tiempos, más concretamente en si dicho equipo ha ganado o ha perdido cada uno de los dos primeros tiempos. Hemos recopilado datos de un buen número de partidos y nos hemos percatado de que el 60 por ciento de los equipos que ganan el partido habían ganado también el primer tiempo, mientras que para los equipos que han perdido el partido esa probabilidad de haber ganado el primer tiempo baja al 30 por ciento. Por lo que respecta al segundo tiempo, el 70 por ciento de los equipos que ganan el partido han ganado también el segundo cuarto, mientras que tan sólo el 20 de los equipos que pierden el partido han ganado dicho segundo cuarto.

1. Con la información que disponemos ¿cuál de los paradigmas de clasificación supervisada vistos durante el curso puede ser aplicado? Razona tu respuesta.

2. Efectuar el pronóstico para un equipo que ha perdido el primer tiempo, pero ha ganado el segundo.

2. Un profesor poco escrupuloso pretende ahorrarse parte de la tarea de corrección de prácticas a base de construir un clasificador que le permita asociar ciertas características de las prácticas a un resultado. En concreto se utilizarán los atributos: tamaño de la práctica (páginas), presentación (buena, regular, mala), documentación (existe o no). Se podrían dar tres resultados posibles: aprueba, suspende, corregir.

Páginas	Presentación	Documentación	Funciona	Resultado
100	Regular	Si	Si	Aprueba
100	Buena	No	No	Suspende
80	Mala	Si	Si	Aprueba
50	Regular	No	Si	Corregir
50	Buena	Si	Si	Aprueba
30	Mala	No	No	Suspende
20	Regular	Si	Si	Aprueba
20	Buena	No	Si	Suspende
20	Mala	Si	Si	Aprueba
10	Regular	No	Si	Suspende

- a) **(1.5 puntos)** Aplica el algortimo de eliminación de candidatos a los tres cuatro últimos ejemplos, suponiendo que el resultado *suspende* corresponde a la clase negativa y *aprueba* a la clase positiva.
- b) **(1.5 puntos)** Calcula el valor del ratio usado en el algoritmo C4.5 para el atributo Páginas, **teniendo en cuenta en este caso los tres valores de la clase Resultado.**

3. **(2 puntos)** La tabla adjunta contiene las distancias euclídeas de un conjunto de entrenamiento formado por 7 casos. Aplica el algoritmo de clustering jerárquico a este conjunto para obtener un dendograma resultante. ¿Qué otros modelos conoces que te servirían para hacer esta tarea de clustering? Cita al menos dos. ¿Qué diferencia hay entre estos modelos y el algoritmo de clasificación por los K vecinos más próximos? Discute esta pregunta sin entrar en los detalles del algoritmo, sólo basándote en el tipo de problemas que se pueden resolver con cada uno de ellos.

D1							
D2	0.3606						
D3	0.5000	0.4243					
D4	0.9220	0.7071	0.4472				
D5	1.3416	1.0440	0.9220	0.5000			
D6	1.8385	1.5524	1.3892	0.9434	0.5099		
D7	1.7263	1.5000	1.2369	0.8062	0.5831	0.4000	
D1	D2	D3	D4	D5	D6	D7	

4. **(3 puntos)** Resuelve mediante el perceptrón multicapa de la figura un problema de clasificación, del que sabemos que para el patrón $I_1=1$, la salida debe ser $(1,0,0)$. Elige para ello todos los parámetros de la red que creas necesarios, añadiéndolos al dibujo junto con los nombres dados a cada una de las unidades, y utiliza la regla delta generalizada para la actualización de los pesos sinápticos.

