



Universidad Nebrija

Asignatura: **CI5128 Ingeniería del conocimiento**
Cuatrimestre: **2º** Examen: **Final** Convocatoria: **Ordinaria**
Grupo: **5IT1** Curso: **2007/2008** Fecha: **22-6-2008**

ALUMNO: _____

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

La siguiente tabla consiste en un conjunto de datos de entrenamiento extraído de una base de datos de empleados. Los datos de la tabla han sido generalizados, de forma que por ejemplo, 31...35 para edad representa el rango de edad que va de 31 a 35.

Además, para cada fila, la variable *count* representa el número de ejemplos que tienen los atributos descritos en esa fila.

Departamento	Status	Edad	Salario	count
Ventas	senior	31...35	46K...50K	30
Ventas	junior	26...30	26K...30K	40
Ventas	junior	31...35	31K...35K	40
Sistemas	junior	21...25	46K...50K	20
Sistemas	senior	31...35	66K...70K	5
Sistemas	junior	26...30	46K...50K	3
Sistemas	senior	41...45	66K...70K	3
Marketing	senior	36...40	46K...50K	10
Marketing	junior	31...35	41K...45K	4
Secretariado	senior	46...50	36K...40K	4
Secretariado	junior	26...30	26K...30K	6

Sea *status* el atributo que designa la clase. Se pide:

1. **(2 puntos)** ¿Cómo modificarías el algoritmo ID3 para que tuviese en consideración el valor de *count* para cada una de las filas? Usa tu algoritmo para construir el primer nodo del árbol de decisión.
2. **(2 puntos)** Dado un ejemplo descrito por los valores (sistemas, 26...30, 46K-50K) para los atributos departamento, edad y salario respectivamente, ¿cuál sería su clasificación de acuerdo al algoritmo Naive Bayes?
3. **(2 puntos)** ¿Y mediante 3-NN?
4. **(2 puntos)** Diseña un perceptrón multicapa y entrénalo para el primer ejemplo de la tabla (sin tener en cuenta el atributo *count*) Para ello:
 - a) Elige el valor de todos los parámetros (incluidos pesos y umbrales) que creas necesarios.
 - b) Usa la regla delta generalizada y la función de activación sigmoide.
5. **(2 puntos)** Diseña una red neuronal que sea capaz de agrupar estos ejemplos en 3 clusters, y entrénala con el primer patrón. Elige el valor de todos los parámetros que creas necesarios. Una vez has procesado el primer ejemplo ¿pertencería el segundo ejemplo a este cluster? Responde a esta pregunta sin hacer cálculos y explicando cómo has llegado a esa conclusión.

NOTA: Sólo hay que tener en cuenta el atributo *count* para el primer apartado.