



# Universidad Nebrija

Asignatura: **CI5128 Ingeniería del conocimiento**  
Cuatrimestre: **2º** Examen: **Final** Convocatoria: **Extraordinaria**  
Grupo: **5IT1** Curso: **2007/2008** Fecha: **27-6-2008**

ALUMNO: \_\_\_\_\_

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

1. Dado el conjunto de datos de la tabla siguiente:

Instancia	A1	A2
1	1,5	87
2	3,1	88,5
3	6	67
4	1,7	88
5	9	33
6	9,1	34,7
7	5	71
8	8,1	41,2
9	9,2	35,4
10	2,5	90
11	5	59
12	5	61
13	3	98
14	8,3	37,2
15	4	62

- (a) **(1 punto)** Definir una medida de distancia adecuada para realizar agrupación o clustering sobre el conjunto de datos. ¿Sería útil aplicar algún tipo de preprocesado sobre los datos? Definir un número de grupos o clusters adecuado para este conjunto de datos, y razonar la respuesta.
- (b) **(1,5 punto)** ¿Qué algoritmo se puede utilizar para realizar agrupación sobre este conjunto de datos? Describir cuál sería la salida de dicho algoritmo sobre este conjunto de datos, teniendo en cuenta el número de grupos definido anteriormente, y sin desarrollar todo el proceso.
- (c) **(2,5 puntos)** Plantea el mismo problema mediante una red neuronal entrenándola con los dos primeros ejemplos.
- i. ¿Qué salida obtendríamos a continuación para el tercer ejemplo? Razona la respuesta sin realizar los cálculos del entrenamiento de la red.

2. Se dispone de la siguiente base de datos, compuesta por las tablas 1 y 2:

Empleado	Empresa	Sueldo	Ascendido
juan	ibm	bajo	si
luis	microsoft	alto	si
pepe	telefónica	alto	no
carmen	microsoft	alto	si
pedro	amena	medio	no
rocío	telefónica	alto	si
maría	amena	bajo	no
raquel	telefónica	bajo	no
fernando	intel	bajo	sí

**Cuadro 1:** Tabla describiendo la empresa en la que trabaja cada empleado, el sueldo que recibe, y si ha ascendido en el ultimo año o no.

Empresa	Empleados	I+D
ibm	340	si
telefónica	200	no
amena	100	no
intel	200	si

**Cuadro 2:** Tabla describiendo el número de empleados de cada empresa y si la empresa desarrolla actividades de I+D.

A partir de dicha base de datos, se pide:

- (a) **(2 puntos)** Aplicar el algoritmo A priori para obtener **sólo** la mejor regla que describa cuándo un empleado ha ascendido. Considerar para formar el conjunto de items frecuentes un soporte mínimo de 2.
- (b) **(1,5 puntos)** Diseña una red neuronal que sirva para resolver este problema de clasificación, indicando qué regla de aprendizaje utilizarías para ello y todos los parámetros necesarios.
- (c) **(1.5 puntos)** Basándose en las tablas anteriores, se pide aplicar el algoritmo de eliminación de candidatos a los 3 últimos ejemplos y en ese orden, considerando que los empleados que no han ascendido pertenecen a la clase negativa. Traduce los límites S y G resultantes a reglas.