



Universidad Nebrija

DII

Asignatura: **CI5128 – Ingeniería del conocimiento**

Cuatrimestre: **2º**

Examen: **Final**

Convocatoria: **Ordinaria**

Grupo: **SIT1**

Curso: **2010/2011**

Fecha: **3-6-2011**

Alumno: _____

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

1. **(2 puntos)** ACME S.A. tiene en total 500 empleados en una única sucursal. Para cada empleado, dispone de un total de 47 datos de diverso tipo, entre los que hay algunos personales, como el DNI, domicilio, etc; laborales , como su puesto, sueldo, etc; y su perfil profesional: estudios, experiencia, idiomas, etc.

La organización ha decidido romper la estructura de la empresa en 5 sucursales, por lo que quiere distribuir a los empleados entre dichas sucursales. El objetivo es que los empleados que tengan relaciones entre sí, o que tengan características similares, terminen en la misma sucursal. Además, también se espera que las distintas sucursales tengan un número parecido de empleados.

1. Definir qué algoritmo o algoritmos de aprendizaje se pueden utilizar para realizar esa tarea, y justificar la respuesta.

2. ¿Cómo preprocesarías los datos para que puedan usarse como entrada en el algoritmo elegido?

3. Describir cómo sería el proceso completo de asignación de empleados a sucursales. Pon un ejemplo de cómo se desarrollaría todo el proceso, sin detallar el resultado ni aplicar el algoritmo.

2. (2.5 puntos) Dada la siguiente tabla, donde ID-P es un identificador de producto, PVP el precio, C el porcentaje que tiene de un cierto componente, T el trimestre en que se ha vendido y PC la percepción del consumidor, responde a las siguientes preguntas:

ID-P	PVP	C	T	PC
1	2,1	0,9	primer	barato
1	1,7	0,3	segundo	normal
1	1,1	0,2	tercer	caro
1	1,4	0,25	cuarto	caro
2	2,2	0,95	primer	barato
2	1,6	0,2	segundo	caro
2	1	0,15	tercer	normal
2	1,6	0,4	cuarto	barato
3	3,1	1,95	primer	caro
3	2,6	1,25	segundo	barato
3	2,4	1,19	tercer	caro
3	2,5	1,4	cuarto	normal
4	2,2	1,5	primer	barato
4	1,6	0,8	segundo	caro
4	1,4	0,7	tercer	barato

- 1) Diseña un perceptrón multicapa que sirva para resolver la tarea de predecir si un producto va a ser caro, barato o normal, junto con la descripción del primer ejemplo de la tabla que se usaría como entrada, así como la salida deseada.
- 2) Aplica el algoritmo del espacio de versiones a los cuatro primeros ejemplos suponiendo que PC barato es equivalente a un ejemplo positivo mientras que tanto normal como caro es equivalente a un ejemplo negativo.

3. **(3 puntos)** Se dispone de datos sobre vivienda en una ciudad, tal y como se muestran en la Tabla 1. A partir de dicha base de datos, se desea construir un modelo que prediga cuándo un piso pertenece a una categoría A o a una categoría B. Para ello, se pide:

ID	Dormitorios	Baños	Tipo	Chimenea	Zona	Categoría
1	1	2	piso	no	centro	A
2	1	1	adosado	si	periferia	B
3	1	1	piso	no	centro	B
4	2	2	piso	no	centro	A
5	2	2	adosado	si	periferia	B
6	2	1	adosado	si	periferia	B
7	3	1	piso	no	centro	B
8	3	2	adosado	si	centro	A
9	2	1	adosado	si	periferia	B
10	1	2	piso	no	centro	A
11	3	1	piso	no	centro	B
12	3	2	piso	no	periferia	B

- ¿Cómo clasificaría un adosado de tres habitaciones y que no tiene chimenea según el algoritmo Naive Bayes?
 - Resuelve el punto 1 según el algoritmo k-NN, con K=3 y una ponderación por la inversa de la distancia.
4. **(2.5 puntos)** Dada la siguiente tabla que muestra las compras de 4 clientes de un supermercado (dados por un identificador TID), aplicar el algoritmo de asociación "A priori" suponiendo un soporte mínimo de 2 y obten todas las reglas de asociación de productos de esta tabla.

TID	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5