



# Universidad Nebrija

Asignatura: **CI5128 – Ingeniería del conocimiento**

Cuatrimestre: **2º**

Examen: **Parcial**

Convocatoria: **Ordinaria**

Grupo: **5IT1**

Curso: **2011/2012**

Fecha: **22-3-2012**

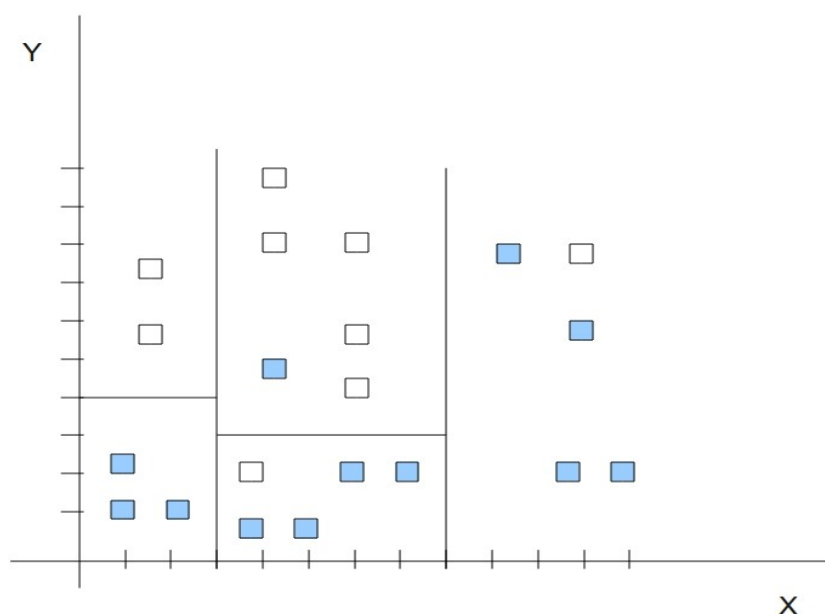
ALUMNO: \_\_\_\_\_

**NOTA:** Se puede usar todo tipo de libros y apuntes de manera personal e intransferible.

**EJERCICIO 1 (2.5 puntos)** Una región de decisión es una zona delimitada por un modelo y que clasificará a los ejemplos futuros con la clase mayoritaria de los ejemplos situados en esa región. En la figura se puede ver un gráfico con las 5 regiones que ha generado un árbol de decisión a partir del algoritmo ID3. Se pide:

a) Dibuja el árbol que ha generado estas regiones de decisión, indicando el rango de valores que has considerado para los atributos X e Y ya discretizados.

b) ¿Cuál sería el porcentaje de errores que ha cometido si lo testeamos con el mismo conjunto de entrenamiento? ¿Cuál sería su matriz de confusión?



## EJERCICIO 2

Sea la siguiente tabla que describe los datos de los clientes de una compañía de seguros.

<i>Edad</i>	<i>Hijos</i>	<i>¿Practica deporte?</i>	<i>Salario</i>	<i>¿Buen cliente?</i>
joven	sí	no	alto	sí
joven	no	sí	medio	no
joven	sí	no	medio	no
joven	sí	no	bajo	sí
mayor	sí	sí	bajo	sí
mayor	no	sí	medio	sí
joven	no	sí	medio	sí
joven	sí	no	alto	sí
mayor	sí	no	medio	sí
mayor	no	no	bajo	no

Se quiere predecir cómo va a ser un cliente nuevo que acaba de entrar de acuerdo al conjunto de entrenamiento anterior y que muestra las siguientes características:

mayor	no	no	medio	no
-------	----	----	-------	----

Para ello haz uso de los dos siguientes algoritmos:

- a) **(1.5 puntos)** Naive Bayes, haciendo uso del factor de Laplace sólo si es necesario.
- b) **(2 puntos)** K-NN, con un valor de  $K=3$  y un sistema de votos que tenga en cuenta la distancia euclídea entre ese ejemplo y los  $K$  vecinos más cercanos.
- c) **(2 puntos)** Aplica el algoritmo de eliminación de candidatos a los cuatro primeros ejemplos de la tabla anterior.

**EJERCICIO 3 (2 puntos)** Consideremos un problema de clasificación trivalente con clases 1, 2 y 3, un conjunto de entrenamiento formado por diez ejemplos  $E=\{e_1, e_2, \dots, e_{10}\}$  y A un atributo discreto tales que

Identificador	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$	$e_9$	$e_{10}$
A	$a_1$	$a_3$	$a_2$	$a_2$	$a_2$	$a_3$	$a_1$	$a_2$	$a_3$	$a_1$
Clase	2	2	1	2	3	2	2	1	1	3

Se pide calcular la ganancia del atributo A respecto al conjunto de entrenamiento E.