

Apellidos:

Nombre:

Ejercicio 1 (2.5 puntos):

Contestar a las preguntas de este ejercicio de manera clara y concisa, *usando el reverso de esta hoja del enunciado y a lo sumo una hoja adicional*:

1. Consideremos la siguiente base de conocimiento BC :

$$H1: p(B, 0).$$

$$R1: p(x, z) \rightarrow p(f(A, x), s(z)).$$

$$R2: p(x, z) \rightarrow p(f(C, x), z).$$

donde A , B , C y 0 son constantes.

El algoritmo de SLD-resolución aplicado al objetivo $p(L, s(0))$ obtiene infinitas soluciones. Desarrollar el árbol hasta encontrar dos soluciones y explicar brevemente cuál es la razón de que se obtengan infinitas soluciones.

2. Una bodega quiere hacer un sistema automático de clasificación de tipos de uva. Suponiendo que manejan 5 tipos de uva distintos y que el tipo de uva se obtiene a partir de 10 características, todas ellas cuantificables, diseñar un sistema basado en una red neuronal para dicha tarea de clasificación. Se debe especificar la estructura de la red, el significado de las unidades de entrada y las de salida, cómo obtener unos pesos adecuados para la red y cómo usar la red finalmente obtenida para clasificar una uva determinada.
3. El método Naive Bayes está basado en decidir, usando inferencia probabilística, respecto de un modelo expresado mediante una red bayesiana. Usando los datos del ejercicio 2, dibujar dicha red bayesiana (sin calcular las tablas de probabilidad) ¿Por qué este método se denomina “naive”? ¿Cómo se calculan las probabilidades de las tablas y qué propiedad tienen? Según este modelo, ¿qué predecimos sobre la supervivencia en cautividad de un pez pequeño, solitario y de aguas cálidas y dulces? ¿por qué?.
4. En el modelo de ejecución de un sistema de producción, ¿qué es la agenda y qué es la base de hechos? Durante el ciclo de ejecución ¿cuándo y por qué se produce una modificación de la agenda? ¿Y de la base de hechos? ¿Qué es la base de reglas y cuándo se modifica durante el ciclo de ejecución?

Apellidos:

Nombre:

Ejercicio 2 (1.5 puntos):

El departamento de Biología Marina quiere analizar las características que hacen que los peces de tamaño inferior a 20 cm sobrevivan en cautividad. Para ello han recogido datos de 20 muestras en las que se indica la adaptación a la CAUTIVIDAD en función del TAMAÑO (pequeño, mediano o grande), la TEMPERATURA del hábitat natural (fría, templada o cálida), la SALINIDAD del hábitat natural (agua dulce o salada) y la SOCIABILIDAD de la especie (solitario, pareja o grupo)

Ejemplo	TAMAÑO	TEMPERATURA	SALINIDAD	SOCIABILIDAD	CAUTIVIDAD
E_1	pequeño	fría	dulce	solitario	no
E_2	pequeño	fría	salada	pareja	si
E_3	pequeño	fría	dulce	grupo	no
E_4	pequeño	templada	salada	solitario	si
E_5	pequeño	templada	dulce	grupo	no
E_6	pequeño	cálida	salada	grupo	si
E_7	mediano	fría	dulce	solitario	si
E_8	mediano	fría	salada	pareja	si
E_9	mediano	templada	dulce	grupo	no
E_{10}	mediano	templada	salada	solitario	no
E_{11}	mediano	cálida	dulce	solitario	si
E_{12}	mediano	cálida	salada	grupo	no
E_{13}	mediano	cálida	dulce	solitario	no
E_{14}	grande	fría	salada	pareja	si
E_{15}	grande	fría	dulce	grupo	si
E_{16}	grande	templada	salada	solitario	no
E_{17}	grande	templada	dulce	grupo	si
E_{18}	grande	templada	salada	grupo	si
E_{19}	grande	cálida	dulce	solitario	no
E_{20}	grande	cálida	salada	pareja	no

1. Al aplicar el algoritmo de aprendizaje inductivo **ID3**, ¿cuál es el nodo elegido como raíz del árbol de aprendizaje?
2. A partir del nodo raíz elegido en el apartado anterior, desarrolla todo el subárbol correspondiente al nodo hijo de mayor entropía.
3. Argumenta razonadamente a favor o en contra de la siguiente afirmación: “A mayor entropía mayor es la profundidad del árbol de decisión”
4. Aplicar el **algoritmo de cobertura** para encontrar reglas con el menor número de condiciones posible, que nos permitan deducir si una especie sobrevive en cautividad a partir de sus características. Detallar los pasos de dicho algoritmo hasta el momento en que se completa la primera regla. En este punto, ¿por qué es necesario seguir aplicando el algoritmo de cobertura?

Tabla de Entropías $Ent(X, Y)$

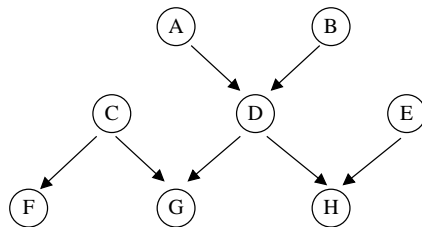
		Y									
		0	1	2	3	4	5	6	7	8	9
X	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.000	1.000	0.918	0.811	0.722	0.650	0.592	0.544	0.503	0.469
	2	0.000	0.918	1.000	0.971	0.918	0.863	0.811	0.764	0.722	0.684
	3	0.000	0.811	0.971	1.000	0.985	0.954	0.918	0.881	0.845	0.811
	4	0.000	0.722	0.918	0.985	1.000	0.991	0.971	0.946	0.918	0.890
	5	0.000	0.650	0.863	0.954	0.991	1.000	0.994	0.980	0.961	0.940
	6	0.000	0.592	0.811	0.918	0.971	0.994	1.000	0.996	0.985	0.971
	7	0.000	0.544	0.764	0.881	0.946	0.980	0.996	1.000	0.997	0.989
	8	0.000	0.503	0.722	0.845	0.918	0.961	0.985	0.997	1.000	0.998
	9	0.000	0.469	0.684	0.811	0.890	0.940	0.971	0.989	0.998	1.000

Apellidos:

Nombre:

Ejercicio 3 (1.5 puntos):

Consideremos la siguiente red bayesiana que relaciona las variables aleatorias A, B, C, D, E, F, G y H :



con las siguientes tablas de distribución:

$P(a)$
0.3

$P(b)$
0.4

$P(c)$
0.8

A	B	$P(d A, B)$
a	b	0.8
a	$\neg b$	0.7
$\neg a$	b	0.4
$\neg a$	$\neg b$	0.3

$P(e)$
0.1

C	$P(f C)$
c	0.8
$\neg c$	0.5

C	D	$P(g C, D)$
c	d	0.1
c	$\neg d$	0.7
$\neg c$	d	0.5
$\neg c$	$\neg d$	0.4

D	E	$P(h D, E)$
d	e	0.3
d	$\neg e$	0.8
$\neg d$	e	0.4
$\neg d$	$\neg e$	0.3

Se pide:

1. Supongamos que queremos calcular $P(d|a, \neg g)$ ¿Qué variables de la red podemos ignorar para esta consulta? Aplicar, detallando cada paso, el algoritmo de eliminación de variables para calcular la probabilidad anterior.
2. Consideremos los siguientes eventos:
 - (a) $(a, b, c, \neg d, \neg e, \neg f, g, h)$
 - (b) $(\neg a, \neg b, \neg c, \neg d, \neg e, \neg f, \neg g, \neg h)$
 - (c) $(a, b, c, d, \neg e, \neg f, \neg g, \neg h)$

¿Cuál(es) de ellos podría(n) ser generados por el algoritmo de ponderación por verosimilitud al intentar aproximar el valor de $P(a|\neg g, \neg h)$? ¿Por qué? ¿Qué peso llevaría(n) asociado?

Apellidos:

Nombre:

Ejercicio 4 (1 punto):

Considérense los siguientes documentos

$D_0 \equiv$ el libro de peces de juan es grueso

$D_1 \equiv$ el libro es de peces

$D_2 \equiv$ juan es grueso

$D_3 \equiv$ el libro es grueso

Ordenar los documentos D_1 , D_2 y D_3 por orden de relevancia primero el documento más relevante) usando para ello la similitud de dichos documentos respecto a D_0 según el modelo de espacio vectorial. Para ello, tomaremos como *vocabulario* el conjunto de todos los términos de todos los documentos. (**Nota:** Tomar tres cifras decimales para los cálculos.)