

Apellidos:

Nombre:

Ejercicio 1 (4 puntos):

Contestar a las siguientes preguntas de manera clara y concisa, *usando el espacio en blanco que se deja a continuación*:

1. Aplicar el algoritmo de unificación para obtener un unificador de máxima generalidad de los términos $f(g(h(u), z), g(z, h(x)))$ y $f(g(y, y), g(h(g(v, v)), v))$, donde u, v, x, y, z son variables.

2. En el algoritmo FOIL, ¿Cuándo decimos que una regla cubre a un ejemplo? Explicar qué propiedad debe cumplir el conjunto de reglas que se construye en el algoritmo FOIL.

3. Supongamos que queremos diseñar una red neuronal que nos sirva para predecir si un valor en bolsa subirá, se mantendrá, o bajará en los próximos siete días. Detallar todos los pasos a seguir: cómo y con qué datos se obtendría un conjunto de entrenamiento, la estructura de la red neuronal, cómo se realizaría el entrenamiento y cómo se usaría la red, una vez entrenada, para predecir el comportamiento de un valor.

4. ¿De qué manera se calcula las salidas a_i de las neuronas en el algoritmo de retropropagación? ¿Y los errores Δ_i ? ¿Por qué el algoritmo de retropropagación es un algoritmo de optimización y qué es lo que trata de minimizar? Dar tres criterios de parada que se podrían usar en el algoritmo

Apellidos:

Nombre:

Ejercicio 2 (3 puntos): (Representación del conocimiento mediante reglas)

Consideremos el problema de describir el filtrado de una secuencia de elementos según cumplan o no una determinada propiedad.

1. Consideremos la siguiente base de conocimiento como aproximación a la resolución del problema anterior:

```
H1: prop(0).
H2: prop(2).
H3: noprop(1).
H4: noprop(3).

H5: filtro(-,-,-).
R1: filtro(c(X,Wx),c(X,Wy),Wz) :- prop(X),filtro(Wx,Wy,Wz).
R2: filtro(c(X,Wx),Wy,c(X,Wz)) :- noprop(X),filtro(Wx,Wy,Wz).
```

donde `prop`, `noprop` y `filtro` son símbolos de predicado, `c` es un símbolo de función y `-` es una constante; y `X`, `Wx`, `Wy` y `Wz` son variables. Desarrollar el árbol de SLD-resolución para los siguientes objetivos, hasta encontrar la primera rama de éxito. ¿Para cuáles de ellos se podrían obtener más soluciones?:

- `filtro(c(0,c(1,c(2,c(3,-)))) ,Wy,Wz)`
- `filtro(Wx,c(0,-),c(1,c(3,-)))`

2. Consideremos ahora el siguiente sistema de producción que pretende resolver el mismo problema:

```
(deftemplate filtro
  (multislot datos)
  (multislot props)
  (multislot noprops))

(defrule filtro-prop
  ?h <- (filtro (datos ?x ?r) (props ?p) (noprops ?n))
  (prop ?x)
  =>
  (modify ?h (datos ?r)
            (props ?p ?x)))

(defrule filtro-noprop
  ?h <- (filtro (datos ?x ?r) (props ?p) (noprops ?n))
  (not (prop ?x))
  =>
  (modify ?h (datos ?r)
            (noprops ?n ?x)))

(deffacts props
  (prop 0)
  (prop 2)
  (filtro (datos 0 1 2 3)))
```

Construir una tabla de seguimiento de la ejecución del sistema de producción anterior indicando todas las activaciones y desactivaciones de las reglas. ¿Con cuál de los objetivos del primer apartado se puede utilizar este sistema de producción con un resultado similar?

Apellidos:

Nombre:

Ejercicio 3 (3 puntos): **(Aprendizaje inductivo)**

1. Aplica el algoritmo FOIL para aprender la relación $p(A, B, C)$ a partir de los siguientes datos:

Ejemplos positivos: (1,3,2), (2,3,2) y (2,3,1)

Ejemplos negativos: (3,3,3) y (3,2,1)

Conocimiento base: $q(1)$ y $q(2)$

En cada paso, indicar qué ejemplos se cubren y especificar claramente la salida del algoritmo.

2. Considera el siguiente conjunto de entrenamiento sobre las compras de los clientes en función de los atributos de un paquete turístico. Las ofertas dependen de los atributos *Destino*, *Tipo de Alojamiento*, *Precio* y *Distancia al domicilio*.

Ej.	DESTINO	ALOJAMIENTO	PRECIO	DISTANCIA	Compra
1	Ciudad	Hotel	Alto	Cerca	Sí
2	Montaña	Apartamento	Alto	Lejos	Sí
3	Ciudad	Hotel	Alto	Lejos	Sí
4	Ciudad	Hotel	Bajo	Lejos	Sí
5	Montaña	Hotel	Alto	Lejos	No
6	Ciudad	Hotel	Bajo	Lejos	No
7	Montaña	Hotel	Bajo	Lejos	Sí
8	Ciudad	Apartamento	Alto	Lejos	No
9	Ciudad	Hotel	Bajo	Cerca	No
10	Montaña	Hotel	Bajo	Lejos	No

¿Cuál es la mejor CONDICIÓN que completa la siguiente regla según el ALGORITMO DE APRENDIZAJE POR COBERTURA?

Si Destino = Ciudad y Alojamiento = Hotel y CONDICIÓN
entonces Compra = Sí

3. Dar dos conjuntos de entrenamiento A y B para el algoritmo ID3 de manera que $B \subseteq A$ y $2 \times \text{Entropía}(A) < \text{Entropía}(B)$

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 4 (4 puntos):

Contestar a las siguientes preguntas de manera clara y concisa, *usando el espacio en blanco que se deja a continuación*:

1. Describir el método Naive Bayes de clasificación: para qué problemas de clasificación se aplica, cómo obtiene un valor de clasificación, y cómo se estiman las probabilidades que se usan en el método ¿Qué suposiciones de independencia condicional se están asumiendo en este método?

2. ¿Qué entendemos por un modelo trigram del lenguaje? ¿Cómo se estiman las probabilidades que se usan en un modelo trigram?

3. Supongamos que hemos observado un conjunto \mathbf{d} de datos, y que queremos predecir el valor de una variable aleatoria X , teniendo en cuenta esos datos y considerando un conjunto H de posibles hipótesis. ¿Cómo se haría con el aprendizaje bayesiano? ¿Y con el aprendizaje MAP? ¿Y con el ML?

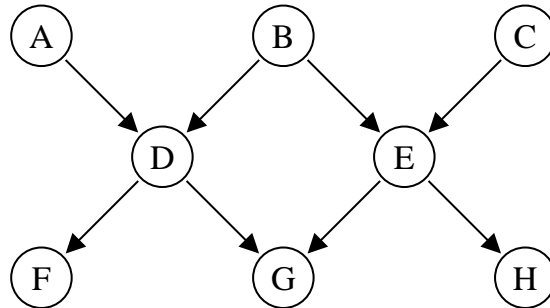
4. Describir en pseudocódigo el algoritmo de construcción de redes bayesianas.

Apellidos:

Nombre:

Ejercicio 5 (3 puntos): (Representacion del conocimiento incierto)

Consideremos la siguiente red bayesiana:



$P(a)$
0.4

$P(b)$
0.2

$P(c)$
0.7

A	B	$P(d A, B)$
a	b	0.3
a	$\neg b$	0.5
$\neg a$	b	0.7
$\neg a$	$\neg b$	0.1

B	C	$P(e B, C)$
b	c	0.9
b	$\neg c$	0.6
$\neg b$	c	0.5
$\neg b$	$\neg c$	0.1

D	$P(f D)$
d	0.4
$\neg d$	0.7

D	E	$P(g D, E)$
d	e	0.1
d	$\neg e$	0.8
$\neg d$	e	0.2
$\neg d$	$\neg e$	0.7

E	$P(h E)$
e	0.2
$\neg e$	0.7

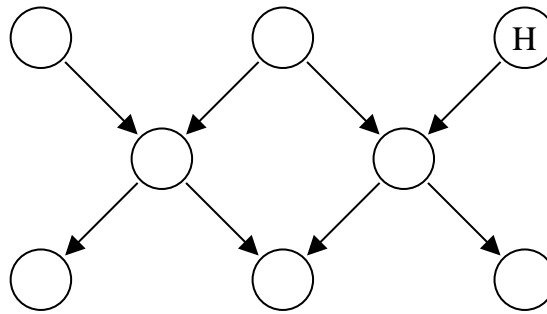
Se pide:

1. Aplicar el algoritmo de eliminación de variables para determinar la distribución de probabilidad de la variable aleatoria C condicionada a que A es falsa y G es verdadera.
2. Consideremos los siguientes eventos:
 - $(\neg a, b, c, \neg d, \neg e, \neg f, g, h)$
 - $(a, \neg b, c, d, \neg e, \neg f, g, \neg h)$
 - $(\neg a, \neg b, \neg c, \neg d, e, \neg f, g, \neg h)$

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

3. Describe de manera informal el concepto de que una variable aleatoria X es condicionalmente independiente de otra variable Y dado el valor de una tercera variable Z.

4. (**Nota:** *Este apartado es completamente independiente de los anteriores, la red bayesiana a la que se hace referencia no es la misma que la del primer apartado.*) Consideremos la siguiente red bayesiana para las variables aleatorias A, B, C, D, E, F, G y H, en la que únicamente se indica la posición de la variable aleatoria H.



Identificar la posición de cada una de las variables aleatorias restantes teniendo en cuenta la siguiente información sobre dependencias y el criterio de D-separación:

- A es condicionalmente independiente de cualquier otra variable dada B
- C es condicionalmente independiente de cualquier otra variable dada D
- B es condicionalmente independiente de D dada E
- G es condicionalmente independiente de E dados B y D
- F no es condicionalmente independiente de E dada B

Justificar razonadamente la asignación de las variables a cada uno de los nodos de la red bayesiana.

Apellidos:
 Nombre:

Ejercicio 6 (3 puntos): (Procesamiento del Lenguaje Natural)

En este ejercicio, consideraremos el poema 1 del libro de poemas *Marinero en Tierra* de Rafael Alberti publicado en 1925 y que obtuvo el Premio Nacional de Poesía. El poema consta de cinco versos. Cada uno de ellos lo consideraremos un documento:

- $D_1 \equiv$ El mar. La mar. El mar. ¡Solo la mar!
- $D_2 \equiv$ ¿Por qué me trajiste, padre, a la ciudad?
- $D_3 \equiv$ ¿Por qué me desenterraste del mar?
- $D_4 \equiv$ En sueños, la marejada me tira del corazón. Se lo quisiera llevar.
- $D_5 \equiv$ Padre, ¿por qué me trajiste acá?

Sean $\vec{W}_1, \dots, \vec{W}_5$ las representaciones vectoriales de los documentos D_1, \dots, D_5 , respecto del conjunto de términos $T = \{ \text{la, mar, me, trajiste} \}$.

Consideremos el siguiente conjunto de entrenamiento a partir de los documentos D_1, \dots, D_4

$$\mathcal{D} = \{ \langle D_1, \oplus \rangle, \langle D_2, \oplus \rangle, \langle D_3, \ominus \rangle, \langle D_4, \ominus \rangle \}$$

- (a) ¿Cuál es la clasificación del documento D_5 obtenida a partir de \mathcal{D} mediante el algoritmo k -NN, con $k = 3$ usando como medida de similitud de dos documentos el coseno del ángulo que forman sus representaciones vectoriales?
- (b) Aplicar el algoritmo de k -medias con $k = 2$ sobre $\vec{W}_1, \dots, \vec{W}_5$ **hasta la primera actualización de los centros** a partir de los centros iniciales $m_1 = (0, 0, 1, 0)$ y $m_2 = (0, 1, 0, 0)$ usando como medida de similitud de dos documentos el coseno del ángulo que forman sus representaciones vectoriales