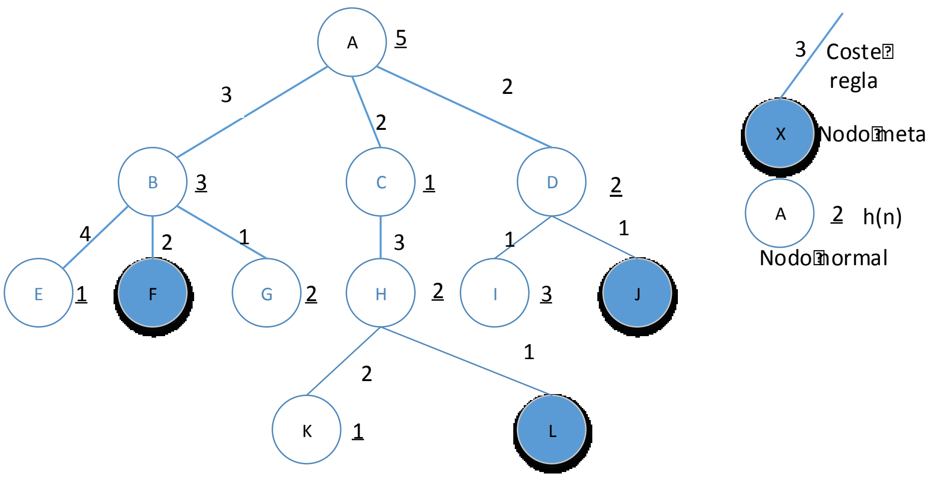
Sistemas Inteligentes – Examen Final (Bloque 1)
ETSINF, Universitat Politècnica de València
26 enero 2017 (2 puntos)

Apellidos:
Nombre:
Grupo: A B C D E F FLIP

1) Para el espacio de estados de la figura y dada una búsqueda de tipo A ($f(n)=g(n)+h(n)$) ¿cuántos nodos es necesario generar, incluyendo el nodo raíz, para encontrar la solución?



A. 7
B. 8
C. 10
D. 12

---

2) Se desea realizar una búsqueda A* en CLIPS. Para ello, las reglas no deben contener la instrucción `retract` en la parte derecha porque:

A. Al borrar los hechos no podemos calcular el valor de $g(n)$ necesario para una búsqueda A*.
B. No permitiría explorar caminos alternativos al elegido en primer lugar
C. No permitiría encontrar la solución óptima
D. Ninguna de las anteriores

---

3) Dado un algoritmo de búsqueda de tipo A, ($f(n)=g(n)+h(n)$), señala la afirmación **CORRECTA**:

A. Si $h(n)$ es consistente (y admisible), expandirá siempre menos nodos que una búsqueda no informada
B. Con $h(n)$ consistente (y admisible), expandirá siempre menos nodos que no siendo consistente.
C. Encuentran siempre la misma solución, independientemente de si $h(n)$ es admisible o no.
D. Ninguna de las anteriores

4) Dado un SBR compuesto de la siguiente regla:

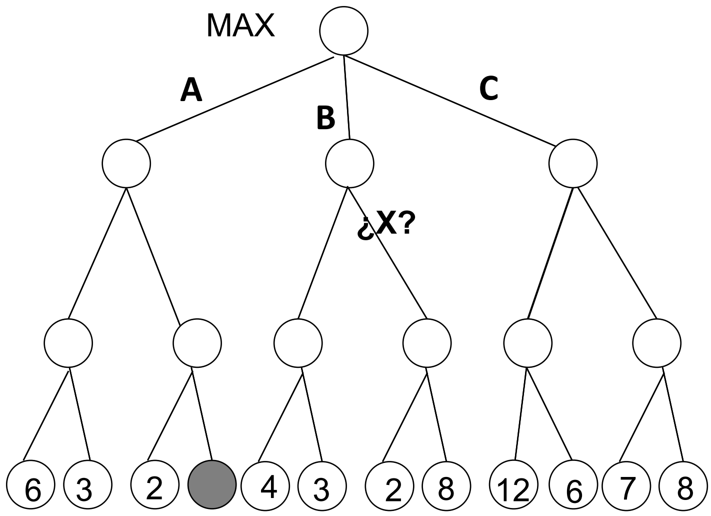
```
(defrule regla-1
   ?f <- (lista ?y $?x ?y $?x ?y)
=>
   (retract ?f)
   (assert (lista $?x)))
```

, y la BH inicial {(lista 1 2 3 2 3 2 3 2 3 2 1 2 3 2 3 2 3 2 3 2 1)}, ¿cuál será el estado final de la Base de Hechos?

A. {(lista 3 2 3)}
B. {(lista 1) (lista)}
C. {(lista 2 1 2)}
D. {(lista 1)}

---

5) Dado el árbol de juego de la figura y aplicando un procedimiento alfa-beta:



¿Qué valor debería tener el nodo terminal sombreado para que se produzca el corte indicado en la figura?

A. Con cualquier valor del nodo se produciría un corte
B. Menor que 3
C. Mayor o igual que 4
D. Nunca se podría producir el corte indicado (o ninguna de las anteriores)

---

6) Dado el árbol de juego de la figura anterior y asumiendo que se produce el corte indicado, tras la aplicación del procedimiento alfa-beta:

A. Se elige la rama A
B. Se elige la rama B
C. Se elige la rama C
D. Se elige la rama A o B

Sistemas Inteligentes – Problema Bloque 1
ETSINF, Universitat Politècnica de València,
26 de enero 2017 (3 puntos)

Se desea diseñar un Sistema Basado en Reglas (SBR) que permita manejar la colección de cromos de varios niños. Cada cromo viene representado por un identificador alfanumérico (A1, B3, etc.). Cada niño puede tener varios cromos (incluidos repetidos) y el número de niños no está limitado en la aplicación. Un posible ejemplo es:

- El niño 1 tiene los cromos: A2 A4 A5 B1 A2 B3
- El niño 2 tiene los cromos: B3 A4 C2 C1 B3 C2
- El niño 3 tiene los cromos: C2 C4 B1 A2

La información dinámica del problema se representaría con el siguiente patrón:

(colecciones [niño ?n [?id-cromo]^{m} fniño ?n]^{m})

donde:

?n ∊ INTEGER ;; Identificador del niño
?id-cromo ∊ {A1, A2, B1,…} ;;Identificador del cromo

Se pide:

a) (0.3 puntos) Escribe la Base de Hechos correspondiente al ejemplo que se muestra arriba.

```
(deffacts datos
    (colecciones  niño 1 A2 A4 A5 B1 A2 B3 fniño 1
            niño 2 B3 A4 C2 C1 B3 C2 fniño 2
            niño 3 C2 C4 B1 A2 fniño 3))
```

b) (1 punto) Escribe una única regla que permita a dos niños intercambiar un cromo. El intercambio solo es posible si el cromo que entrega cada niño es un cromo repetido en su colección, y el cromo que recibe cada niño es un cromo que no está en su colección.

```
(defrule intercambio
    (colecciones $?x niño ?n1 $?c1 ?c $?c2 ?c $?c3 fniño ?n1
            $?y niño ?n2 $?p1 ?p $?p2 ?p $?p3 fniño ?n2 $?z)
    (test (neq ?c ?p))
    (test (and (not (member$ ?c $?p1))(not (member$ ?c $?p2))(not (member$ ?c $?p3))))
    (test (and (not (member$ ?p $?c1))(not (member$ ?p $?c2))(not (member$ ?p $?c3))))
    =>
    (assert (colecciones $?x niño ?n1 $?c1 ?c $?c2 ?p $?c3 fniño ?n1 $?y niño ?n2 $?p1 ?p $?p2
?c $?p3 fniño ?n2 $?z)))
```

Another solution:

```
(defrule intercambio
  ?f1 <- (colecciones $?x niño ?n1 $?c1 ?c $?c2 ?c $?c3 fniño ?n1
                    $?y niño ?n2 $?p1 ?p $?p2 ?p $?p3 fniño ?n2 $?z)
  ?f2 <- (colecciones $?x niño ?n1 $?todo1 fniño ?n1
                   $?y niño ?n2 $?todo2 fniño ?n2 $?z)
  (test (eq ?f1 ?f2))
  (test (neq ?c ?p))
  (test (and (not (member$ ?p $?todo1))(not (member$ ?c $?todo2))))
=>
  (assert (colecciones $?x niño ?n1 $?c1 ?c $?c2 ?p $?c3 fniño ?n1 $?y niño ?n2 $?p1 ?p $?p2
?c $?p3 fniño ?n2 $?z)))
```

c) (0.7 puntos) Escribe una única regla que muestre los niños que tienen algún cromo que aparece (exactamente) tres veces en la colección. Hay que mostrar un mensaje por pantalla por cada niño y cromo; ejemplo: "El niño " ?n " tiene el cromo " ?x " tres veces".

```
(defrule acaparador
  (colecciones $? niño ?n $?x1 ?y1 $?x2 ?y1 $?x3 ?y1 $?x4 fniño ?n $?)
  (test (and (not (member ?y1 $?x1))(not (member?y1 $?x2))
             (not (member ?y1 $?x3))(not (member ?y1 $?x4))))
 =>
 (printout t "El niño " ?n " tiene el cromo " ?y1 "  tres veces " crlf))
```

d) (1 punto) Supongamos que existen unos hechos del tipo (especial ?id-cromo) para indicar que el cromo identificado por ?id-cromo es un cromo especial. Escribe una única regla que calcule el número de niños que tienen al menos 2 cromos especiales y distintos entre sí. El resultado de la ejecución de la regla será un hecho con formato: (lista-especial [?n]^{m}) donde ?n es el identificador de un niño con al menos dos cromos especiales y diferentes. El identificador de cada niño solo debe aparecer una vez en la lista (aunque tenga varios cromos especiales). Asumir que en la BH existen varios hechos del tipo (especial ?id-cromo) y el hecho (lista-especial).

```
(defrule especiales
     (colecciones $? niño ?n1 $? ?a $? ?b $? fniño ?n1 $?y)
     (especial ?a)
     (especial ?b)
     (test (neq ?a ?b))
 ?r <- (lista-especial $?z)
     (test (not (member$ ?n1 $?z)))
  =>
    (retract ?r)
    (assert (lista-especial $?z ?n1)))
```

Examen Final de Sistemas Inteligentes: Bloque 2

ETSINF, Universitat Politècnica de València, 26 de enero de 2017

Apellidos: ____________________ **Nombre:** ____________

Grupo: ☐ **3A** ☐ **3B** ☐ **3C** ☐ **3D** ☐ **3E** ☐ **3F** ☐ **3FLIP**

Cuestiones (2 puntos; tiempo estimado: 30 minutos)

Marca cada recuadro con una única opción de entre las dadas.

1 $\boxed{C}$ Sean X y Y dos variables aleatorias, y sean $P(X, Y)$, $P(X \mid Y)$, $P(Y \mid X)$, $P(X)$ y $P(Y)$ las probabilidades conjunta, condicionales e incondicionales de esas variables. Indica cuál de las siguientes afirmaciones es *incorrecta*.

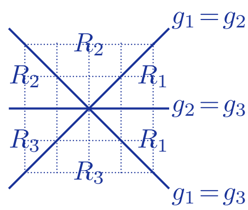
A) Tanto $P(X)$ como $P(Y)$ se pueden derivar a partir de $P(X, Y)$.

B) Tanto $P(X \mid Y)$ como $P(Y \mid X)$ se pueden derivar a partir de $P(X, Y)$.

C) Se puede obtener $P(Y \mid X)$ a partir de $P(X \mid Y)$ y $P(X)$, sin necesidad de conocer previamente $P(Y)$.

D) Se puede obtener $P(Y \mid X)$ a partir de $P(X \mid Y)$ y $P(Y)$, sin necesidad de conocer previamente $P(X)$.

2 $\boxed{C}$ Sean un problema de clasificación de tres clases en $\mathbb{R}^2$ y un clasificador definido por tres funciones discriminantes lineales: $g_1(\mathbf{x}) = x_1$, $g_2(\mathbf{x}) = x_2$ y $g_3(\mathbf{x}) = -x_2$. Indica cuál de las afirmaciones sobre dicho clasificador *no* es correcta.



A) Define tres fronteras de decisión que intersectan en el origen de coordenadas.

B) La región de decisión de la clase 1 se define como $R_1 = \{\mathbf{x} \in \mathbb{R}^2 : x_1 > 0 \wedge x_1 > |x_2|\}$.

C) En la región de decisión R_2, x_2 es menor que cero y en R_3, x_2 es mayor que cero.

D) En la región de decisión R_2, x_2 es mayor que cero y en R_3, x_2 es menor que cero.

3 $\boxed{B}$ Sea $\hat{p}$ la probabilidad de error de un clasificador estimada a partir de un conjunto de test cuya talla es N y sea $I = [\hat{p} \pm \epsilon]$ el intervalo de confianza de esta estimación. Indica la respuesta *correcta*.

A) Si $N = 160$ y el clasificador produce al menos un error, ϵ será menor que $1\,\%$.

B) Si $N > 150$ y se obtiene $\hat{p} = 0.1$, ϵ será menor que $5\,\%$.

C) Si N_e es el número de errores del clasificador, entonces $\hat{p} = N/N_e$ y ϵ es inversamente proporcional a N.

D) No es posible determinar ϵ si $\hat{p} = 0$.

4 $\boxed{D}$ Supongamos que hemos aplicado el algoritmo C-medias a un conjunto de puntos bi-dimensionales para obtener un agrupamiento (partición) en dos *clústers*. Tras una serie de iteraciones del algoritmo C-medias tenemos el agrupamiento: $\{\ \{(0,1)^{\mathbf{t}}, (0,2)^{\mathbf{t}}\},\ \{(0,3)^{\mathbf{t}}, (0,5)^{\mathbf{t}}, (0,6)^{\mathbf{t}}, (0,7)^{\mathbf{t}}, (1,6)^{\mathbf{t}}, (-1,6)^{\mathbf{t}}\}\ \}$. Indica la respuesta *correcta*.

A) La suma de errores cuadráticos (SEC) es 15 y puede llegar a ser 8.

B) La SEC es 15 y cuando el algoritmo converja será 12.

C) La SEC es 12 y cuando el algoritmo converja será 10.

D) La SEC es 12 y cuando el algoritmo converja será 6.

5 $\boxed{D}$ Sean M un modelo de Markov oculto y x una cadena de longitud T a la que M asigna una probabilidad positiva. Sean q un estado no final de M y t un instante de tiempo no mayor que T. Considera el valor $\alpha(q,t)$ calculado por el algoritmo *Forward* y el valor $V(q,t)$ calculado por el algoritmo de Viterbi. Supón que tanto $\alpha(q,t)$ como $V(q,t)$ son positivos. Indica la respuesta *correcta* en relación con $\alpha(q,t)$ y $V(q,t)$:

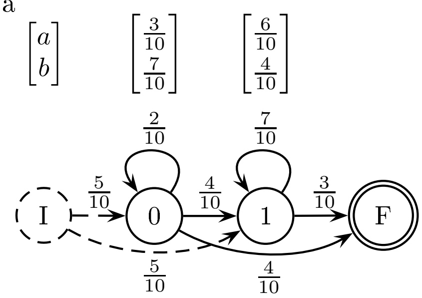
A) Siempre coinciden si $t > 1$.

B) Nunca coinciden si $t > 1$.

C) Nunca coinciden si $t = 1$.

D) Siempre coinciden si $t = 1$.

6 $\boxed{A}$ Dado el modelo de Markov oculto M a la derecha, la aproximación de Viterbi a la probabilidad exacta que M asigna a la cadena "`bab`", $\tilde{P}_M(bab)$, es:



A) $0.000 \leq \tilde{P}_M(bab) < 0.010$ $\qquad \tilde{P}_M(bab, 011F)$

B) $0.010 \leq \tilde{P}_M(bab) < 0.015$ $\qquad = \tilde{P}_M(bab, 111F)$

C) $0.015 \leq \tilde{P}_M(bab) < 0.020$ $\qquad = \frac{70560}{10^7} = 0.007056$

D) $0.020 \leq \tilde{P}_M(bab)$

Examen Final de Sistemas Inteligentes: Bloque 2

ETSINF, Universitat Politècnica de València, 26 de enero de 2017

Apellidos: ☐ **Nombre:** ☐

Grupo: ☐ **3A** ☐ **3B** ☐ **3C** ☐ **3D** ☐ **3E** ☐ **3F** ☐ **3FLIP**

Problema (3 puntos; tiempo estimado: 45 minutos)

Se tiene un problema de clasificación en dos clases, 0 y 1, para objetos representados en $\{0,1\}^2$, esto es, vectores de bits de la forma $\mathbf{x} = (x_1, x_2)^t$ con $x_1, x_2 \in \{0,1\}$. Asimismo, disponemos de cuatro muestras de entrenamiento:

$\mathbf{x}_n$	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
x_{n1}	0	0	1	1
x_{n2}	0	1	0	1
c_n	0	1	1	0

Se pide:

1 (0.75 puntos) Aplica una iteración del algoritmo Perceptrón con pesos iniciales nulos, constante de aprendizaje $\alpha = 1$ y margen $b = 0.1$. ¿Qué pesos se obtienen al finalizar la iteración aplicada?

2 (0.50 puntos) A partir de la inicialización dada en el apartado anterior, ¿convergerá el algoritmo Perceptrón a una solución sin datos de entrenamiento mal clasificados?
Indica si sí o no y después comenta brevemente la respuesta.

3 (0.25 puntos) A partir de alguna inicialización con pesos no nulos, $\alpha > 0$ y $b = 0.1$, ¿convergerá el algoritmo Perceptrón a una solución sin datos de entrenamiento mal clasificados?
Indica si sí o no y después comenta brevemente la respuesta.

4 (0.75 puntos) Aplica el algoritmo de aprendizaje de árboles de clasificación visto en clase al problema dado. Con el fin de medir el grado de impureza de un nodo, utiliza la entropía de la distribución empírica de las probabilidades a posteriori de las clases en dicho nodo. Por otro lado, para decidir si un nodo es terminal o no, emplea el criterio de parada visto en clase con umbral de (decremento de) impureza $\epsilon = 0.1$. Asimismo, con el fin de explorar posibles particiones ("splits") de un nodo, considera únicamente el umbral de corte $r = 0.5$.

5 (0.50 puntos) Repite el apartado anterior con el criterio de parada "mínimamente estricto", esto es, permitiendo la partición ("split") de un nodo siempre que no dé lugar a nodos hijo vacíos.

6 (0.25 puntos) Supongamos que los diferentes objetos de nuestro problema se presentan con igual probabilidad, es decir, $P(\mathbf{x}_1) = P(\mathbf{x}_2) = P(\mathbf{x}_3) = P(\mathbf{x}_4) = 0.25$. De entre los clasificadores obtenidos en los apartados anteriores, ¿existe alguno de menor error de clasificación (teórico) que el resto? Justifica brevemente la respuesta.

1 Notación homogénea. $\mathbf{w}_0 = \mathbf{w}_1 = (0\ 0\ 0)^t$, $\alpha = 1$ y $b = 0.1$.

$$\begin{aligned}
&g_0(\mathbf{x}_1) = (0\ 0\ 0)(1\ 0\ 0)^t = 0\\
&g_1(\mathbf{x}_2) = (0\ 0\ 0)(1\ 0\ 0)^t = 0\\
&g_1(\mathbf{x}_1) + b > g_0(\mathbf{x}_2)? \quad \text{Sí}\\
&\ \to \mathbf{w}_1 = \mathbf{w}_1 - \mathbf{x}_1 = (0\ 0\ 0)^t - (1\ 0\ 0)^t = (-1\ 0\ 0)^t\\
&\ \to \mathbf{w}_0 = \mathbf{w}_0 + \mathbf{x}_1 = (0\ 0\ 0)^t + (1\ 0\ 0)^t = (1\ 0\ 0)^t\\
&g_1(\mathbf{x}_2) = (-1\ 0\ 0)(1\ 0\ 1)^t = -1\\
&g_0(\mathbf{x}_2) = (1\ 0\ 0)(1\ 0\ 1)^t = 1\\
&g_0(\mathbf{x}_2) + b > g_1(\mathbf{x}_2)? \quad \text{Sí}\\
&\ \to \mathbf{w}_0 = \mathbf{w}_0 - \mathbf{x}_2 = (1\ 0\ 0)^t - (1\ 0\ 1)^t = (0\ 0\ -1)^t\\
&\ \to \mathbf{w}_1 = \mathbf{w}_1 + \mathbf{x}_2 = (-1\ 0\ 0)^t + (1\ 0\ 1)^t = (0\ 0\ 1)^t\\
&g_1(\mathbf{x}_3) = (0\ 0\ 1)(1\ 1\ 0)^t = 0\\
&g_0(\mathbf{x}_3) = (0\ 0\ -1)(1\ 1\ 0)^t = 0\\
&g_0(\mathbf{x}_3) + b > g_1(\mathbf{x}_3)? \quad \text{Sí}\\
&\ \to \mathbf{w}_0 = \mathbf{w}_0 - \mathbf{x}_3 = (0\ 0\ -1)^t - (1\ 1\ 0)^t = (-1\ -1\ -1)^t\\
&\ \to \mathbf{w}_1 = \mathbf{w}_1 + \mathbf{x}_3 = (0\ 0\ 1)^t + (1\ 1\ 0)^t = (1\ 1\ 1)^t\\
&g_0(\mathbf{x}_4) = (-1\ -1\ -1)(1\ 1\ 1)^t = -3\\
&g_1(\mathbf{x}_4) = (1\ 1\ 1)(1\ 1\ 1)^t = 3\\
&g_1(\mathbf{x}_4) + b > g_0(\mathbf{x}_4)? \quad \text{Sí}\\
&\ \to \mathbf{w}_1 = \mathbf{w}_1 - \mathbf{x}_4 = (1\ 1\ 1)^t - (1\ 1\ 1)^t = (0\ 0\ 0)^t\\
&\ \to \mathbf{w}_0 = \mathbf{w}_0 + \mathbf{x}_4 = (-1\ -1\ -1)^t + (1\ 1\ 1)^t = (0\ 0\ 0)^t
\end{aligned}$$

Se obtienen pesos nulos, esto es, los mismos utilizados como inicialización.

2 No. El conjunto de muestras de entrenamiento *no* es linealmente separable. En el mejor de los casos, podríamos clasificar bien tres de las cuatro muestras de entrenamiento.

3 No, por el mismo motivo dado en el apartado anterior.

4 La impureza del nodo raíz es 1. Existen dos particiones ("splits") posibles del nodo raíz con umbrales de corte nulos: $(x_1, 0)$ y $(x_2, 0)$. En ambos casos se generan nodos hijo con un dato de cada clase, por lo que también tienen impureza 1. Así pues, el máximo decremento de impureza alcanzable es nulo y el algoritmo termina sin dicotomizar el nodo raíz.

5 El algoritmo genera un árbol binario completo de profundidad dos y con un único dato de entrenamiento en cada nodo hoja.

6 Si $P(\mathbf{x}_1) = P(\mathbf{x}_2) = P(\mathbf{x}_3) = P(\mathbf{x}_4) = 0.25$, la probabilidad de error será una simple media aritmética de la probabilidad de error a posteriori:

$$\begin{aligned}
P(error) &= P(error, \mathbf{x}_1) + P(error, \mathbf{x}_2) + P(error, \mathbf{x}_2) + P(error, \mathbf{x}_2)\\
&= P(\mathbf{x}_1)\, P(error \mid \mathbf{x}_1) + P(\mathbf{x}_2)\, P(error \mid \mathbf{x}_2) + P(\mathbf{x}_3)\, P(error \mid \mathbf{x}_3) + P(\mathbf{x}_4)\, P(error \mid \mathbf{x}_4)\\
&= 0.25 \cdot (P(error \mid \mathbf{x}_1) + P(error \mid \mathbf{x}_2) P(error \mid \mathbf{x}_3) + P(\mathbf{x}_4)\, P(error \mid \mathbf{x}_4))
\end{aligned}$$

Tan solo el árbol de clasificación obtenido en el apartado anterior será capaz de clasificar los diferentes objetos sin error, por lo que su probabilidad de error será nula. El resto de clasificadores clasificaría erróneamente uno o más de los cuatro objetos distintos que pueden darse.