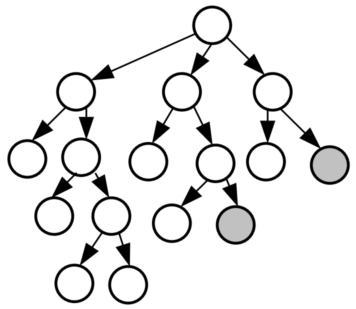
Sistemas Inteligentes – Examen Final (Bloque 1), 23 enero 2020
Test (1,75 puntos) puntuación: max (0, (aciertos – errores/3) * 1,75/6)

Apellidos:		**Nombre**:
Grupo:	A B C D E F G	

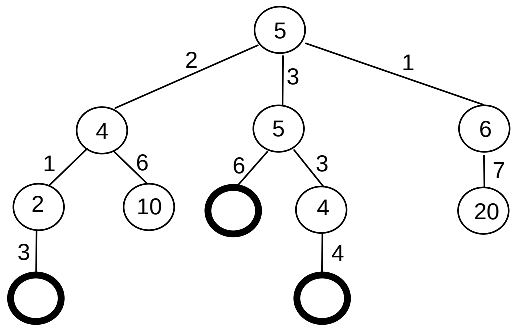
1) Considerando el siguiente árbol de búsqueda, ¿cuántos nodos como máximo se almacenan en memoria, aplicando un procedimiento de búsqueda en profundidad iterativa? (Asúmase que a igual profundad se elige el nodo más a la izquierda)



A. 6
B. 8
C. 10
D. Ninguna de las respuestas anteriores es correcta.

---

2) Sea el árbol de la figura donde los nodos de trazo grueso son nodos meta, el valor dentro del nodo es el valor de la función heurística aplicada a cada nodo y el valor de los arcos es el coste del operador correspondiente. Indica la respuesta CORRECTA:



A. La heurística es admisible y consistente
B. La heurística no es admisible ni consistente
C. Aplicando un algoritmo de tipo A se encuentra la solución óptima
D. Ninguna de las opciones anteriores es correcta

---

3) Sean dos funciones de evaluación f1(n)=g(n)+h1(n) y f2(n)=g(n)+h2(n), tales que h1(n) es admisible y h2(n) no lo es, indica la respuesta CORRECTA:

A. El uso de ambas funciones en un algoritmo de tipo A garantiza, en cada caso, encontrar la solución óptima
B. Se garantiza que f2(n) generará un menor espacio de búsqueda que f1(n)
C. Sólo si h1(n) es una heurística consistente, f1(n) generará un menor espacio de búsqueda que f2(n)
D. Existe algún nodo n para el que h2(n)>h*(n)

---

4) En una búsqueda GRAPH-SEARCH que aplica un algoritmo de tipo A (f(n)=g(n)+h(n)), se tiene un nodo n en la lista CLOSED y un nodo n' en la lista OPEN tal que n'=n. Indica la respuesta CORRECTA:

A. Si la heurística es admisible, se cumple siempre h(n) < h(n').
B. Si la heurística es consistente, se cumple siempre g(n) <= g(n').
C. Independientemente de si la heurística es consistente o no, se cumple siempre f(n) <= f(n').
D. Ninguna de las respuestas anteriores es correcta.

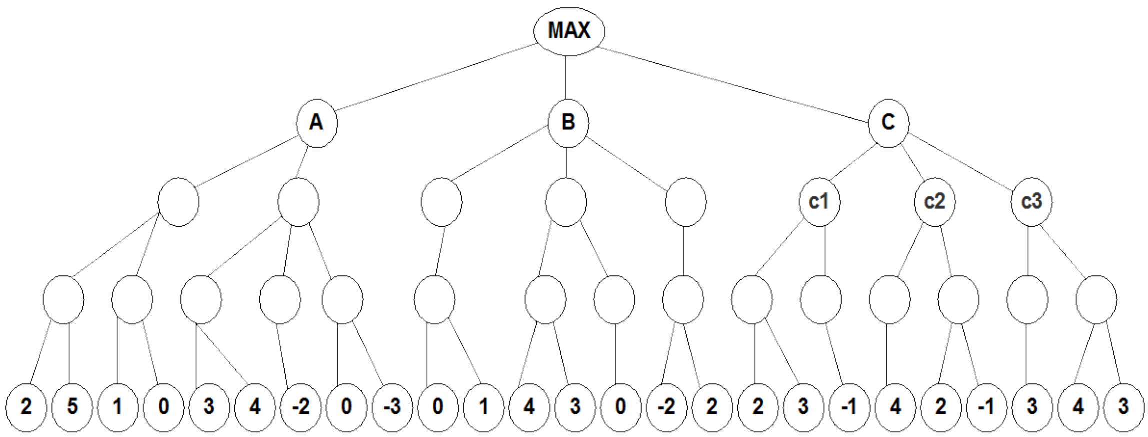
---

5) Sea n1 y n2 los dos únicos nodos hijo de un nodo n el cual es un nodo MAX en un árbol de juego. Asumimos que se explora primero el nodo n1 y luego n2. Indica la respuesta CORRECTA:

A. El valor definitivo del nodo n será el máximo entre el valor definitivo de n1 y n2 solo cuando n1 y n2 son nodos terminales.
B. Cuando se vuelca el valor de n1 al nodo padre n, este puede tener asociado un valor volcado anteriormente.
C. Cuando se vuelva el valor de n1 al nodo padre n, se puede producir un corte beta en n.
D. Ninguna de las respuestas anteriores es correcta.

---

6) ¿Cuál es la mejor jugada para el nodo raíz MAX si aplicamos un alfa-beta al árbol de juego?



A. La rama A
B. La rama B
C. La rama C
D. La rama A ó B

Sistemas Inteligentes – Examen Final (Bloque 1), 23 enero 2019
Problema: 2 puntos

Se desea formar dos grupos de personas, uno de personas que hablen ruso y otro grupo de personas que hablen chino. Para ellos se presentan varias personas que acreditan dominio de uno o los dos idiomas. El nivel de clasificación del dominio de la lengua es de 1 a 5, siendo 1 el menor nivel y 5 el nivel máximo.

- P1 acredita chino con nivel 3 y ruso con nivel 1
- P2 acredita ruso con nivel 4
- P3 acredita ruso con nivel 1 y chino con nivel 2
- P4 acredita chino con nivel 3
- P5 acredita ruso con nivel 3
- P6 acredita chino con nivel 2 y ruso con nivel 5
- P7 acredita chino con nivel 4
- P8 acredita ruso con nivel 3 y chino con nivel 2

El patrón para la formación de los dos grupos es el siguiente:

(grupos ruso p^m chino q^m) donde $p,q \in$ {P1,P2,P3,P4,P5,P6,P7,P8}

1) (0,5 puntos) Escribe la Base de Hechos correspondiente a la situación inicial que se muestra arriba asumiendo que los grupos inicialmente están vacíos. Incluye los patrones adicionales que necesites para representar la información estática del problema, así como los hechos asociados a dichos patrones.
2) (0,8 puntos) Escribe una única regla para añadir una persona al grupo de chino o ruso, comprobando que la persona acredita el idioma correspondiente con un nivel mínimo de 2 y que dicha persona no está ya apuntada a ningún grupo.
3) (0,7 puntos) Escribe una regla que muestre un mensaje por pantalla indicando el número de personas en cada grupo cuando se hayan conseguido al menos tres personas en cada uno de ellos.

Solución:

```
(deffacts datos
  (persona P1 chino 3)
  (persona P1 ruso 1)
  (persona P2 ruso 4)
  (persona P3 ruso 1)
  (persona P3 chino 2)
  (persona P4 chino 3)
  (persona P5 ruso 3)
  (persona P6 chino 2)
  (persona P6 ruso 5)
  (persona P7 chino 4)
  (persona P8 ruso 3)
  (persona P8 chino 2)
```

```
  (grupos ruso chino))


(defrule R1
   (grupos $?x ?idioma $?y)
   (test (or (eq ?idioma ruso)(eq ?idioma chino)))
   (persona ?per ?idioma ?nivel)
   (test (>= ?nivel 2))
   (test (not (member ?per $?x)))
   (test (not (member ?per $?y)))
 =>
    (assert (grupos $?x ?idioma ?per $?y)))


(defrule R2
   (declare (salience 100))
     (grupos ruso $?rus chino $?chi)
     (test (>= (length$ $?rus) 3))
     (test (>= (length$ $?chi) 3))
  =>
     (printout t "Se han conseguido "(length$ $?rus) " personas en el grupo de ruso y " (length$ $?chi) " personas en el grupo de chino " crlf)
     (halt))
```

Examen final de SIN: bloque 2

ETSINF, Universitat Politècnica de València, 23 de enero de 2020

Apellidos: ______ **Nombre:** ______

Grupo: ☐ **3A** ☐ **3B** ☐ **3C** ☐ **3D** ☐ **3E** ☐ **3F** ☐ **3G** ☐ **4IA**

Test (1,75 puntos)

Marca cada recuadro con una única opción. Puntuación: máx(0, (aciertos − errores / 3) · 1,75 / 6).

1 $\boxed{\text{A}}$ Sean C, L, S variables aleatorias que toman valores en $\{\text{DES},\text{NUB},\text{LLU}\}$, $\{\text{DIA}, \text{NOC}\}$, y $\{\text{SEG},\text{ACC}\}$, respectivamente. Su probabilidad conjunta viene dada en la siguiente tabla:

s l c	SEG DIA DES	SEG DIA NUB	SEG DIA LLU	SEG NOC DES	SEG NOC NUB	SEG NOC LLU	ACC DIA DES	ACC DIA NUB	ACC DIA LLU	ACC NOC DES	ACC NOC NUB	ACC NOC LLU
$P(s,l,c)$	0.30	0.20	0.07	0.13	0.10	0.06	0.01	0.01	0.03	0.02	0.02	0.05

¿Cuál es la probabilidad condicional $P(C = \text{LLU} | S = \text{ACC}, L = \text{DIA})$?:

A) 0.60. $P(C = \text{LLU} \mid S = \text{ACC}, L = \text{DIA}) = P(C = \text{LLU}, S = \text{ACC}, L = \text{DIA})/P(S = \text{ACC}, L = \text{DIA})$
B) 0.03. $P(S = \text{ACC}, L = \text{DIA}) = \sum_c P(S = \text{ACC}, L = \text{DIA}, c) = 0.05$
C) 0.05. $P(C = \text{LLU} \mid S = \text{ACC}, L = \text{DIA}) = 0.03/0.05 = 0.60$
D) 0.02.

2 $\boxed{\text{D}}$ La expresión $\hat{c} = \arg\max_{1\leq c\leq C} P(c \mid \mathbf{y})$, donde $\mathbf{y} \in \mathbb{R}^d$ es un dato a clasificar, corresponde a un clasificador de mínimo riesgo de error o de Bayes en C clases. Con algunas asunciones, este clasificador coincide con un clasificador basado en C *Funciones Discriminantes*, definido por $\hat{c} = \arg\max_{1\leq c\leq C} g_c(\mathbf{y})$. Indica cuál de las siguientes asunciones *no* sería generalmente correcta:

A) $g_c(\mathbf{y}) = P(c \mid \mathbf{y})$.
B) $g_c(\mathbf{y}) = \log P(c \mid \mathbf{y})$.
C) $g_c(\mathbf{y}) = 0.5 \cdot P(c \mid \mathbf{y}) + 0.5$.
D) $g_c(\mathbf{y}) = \sum_{j=1}^{d} a_j P(c \mid y_j) + a_0$, donde $a_j, 0 \leq j \leq d$, son coeficientes reales no nulos cualesquiera.

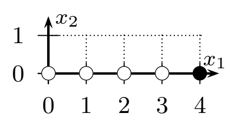
3 $\boxed{\text{B}}$ Sea S un conjunto de N pares de entrenamiento y C el número de clases. Considera una iteración cualquiera, que no sea la última, del algoritmo Perceptrón aplicado a S. En esa iteración se modifican k vectores de pesos. ¿Cuál de las siguientes afirmaciones es *incorrecta*?

A) $2 \leq k \leq C \cdot N$.
B) $2 \leq k \leq (C - 1) \cdot N$.
C) $2 \leq k \leq C' \cdot N$, donde C' está acotada según $1 \leq C' \leq C$.
D) $2 \leq k \leq \sum_{n=1}^{N} k_n$, donde $k_n, 1 \leq k_n \leq C$, es el número de vectores modificados para el dato n-ésimo.

4 $\boxed{\text{C}}$ Se está aplicando el algoritmo de aprendizaje de árboles de clasificación para un problema de C clases. Durante la ejecución de dicho algoritmo se ha alcanzado un nodo t que incluye $N(t)$ datos. Se pretende evaluar el criterio de parada de particionamiento en dicho nodo y se sabe que su conjunto de posibles *"splits"* toma valores de decremento de impureza (medida en términos de entropía) en el intervalo $(0.5, 1.0]$. ¿Cuál de los siguientes rangos de valores de la constante ϵ garantiza la detención del particionamiento?

A) $0.0 < \epsilon \leq 0.5$.
B) $0.5 < \epsilon \leq 1.0$.
C) $1.0 < \epsilon \leq 2.0$.
D) Ninguno de los anteriores.

5 $\boxed{\text{C}}$ La figura a la derecha muestra una partición de 5 puntos bidimensionales en 2 clústers (representados mediante los símbolos • y ∘). Si, durante la ejecución del algoritmo C-medias, el punto $(3, 0)$ se cambia del clúster ∘ al •, entonces (indica cuál de la siguientes afirmaciones es cierta):



A) Las medias de clúster no cambian.
B) La suma de errores cuadráticos crece.
C) La suma de errores cuadráticos decrece.
D) Solo cambia la suma de errores cuadráticos de uno de los clústers.

6 $\boxed{\text{C}}$ El algoritmo de re-estimación por Viterbi es un método para aprender los parámetros de un modelo de Markov a partir de un conjunto de secuencias de símbolos. En este contexto, indica qué afirmación es falsa:

A) En el algoritmo de re-estimación por Viterbi se cuenta el número de veces que se ha utilizado cada transición entre estados, a partir de las secuencias de estados halladas mediante el algoritmo de Viterbi. Posteriormente, se normalizan los contadores obtenidos.

B) En el algoritmo de re-estimación por Viterbi se cuenta el número de veces que cada símbolo ha sido emitido en cada estado, a partir de las secuencias de estados halladas mediante el algoritmo de Viterbi. Posteriormente, se normalizan los contadores obtenidos.

C) El algoritmo de re-estimación por Viterbi consiste en aplicar únicamente el algoritmo de Viterbi y calcular la probabilidad de que el modelo de Markov genere cada secuencia de símbolos de entrenamiento.

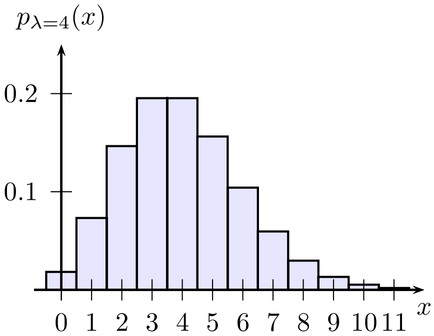
D) En el algoritmo de re-estimación por Viterbi es importante la inicialización de los parámetros del modelo.

Problema (2 puntos)

Sea $\lambda \in \mathbb{R}^+$. Decimos que una variable aleatoria $x \in \{0, 1, 2, \ldots\}$ es Poisson(λ) si su función de masa de probabilidad es:

$$p_\lambda(x) = \frac{\exp(-\lambda)\, \lambda^x}{x!}$$

La distribución de Poisson se emplea para modelizar la probabilidad de que un evento dado ocurra un cierto número de veces en un contexto prefijado. El parámetro λ puede interpretarse como la media de ocurrencias de dicho evento. Por ejemplo, x podría ser el número de llamadas telefónicas que recibimos en un día o el número de ocurrencias de una cierta palabra en un documento dado. La figura a la derecha muestra $p_{\lambda=4}(x)$ para todo $x \in \{0, 1, \ldots, 11\}$.



Sea un problema de clasificación en C clases para objetos representados mediante una característica de tipo contador, $x \in \{0, 1, 2, \ldots\}$. Para toda clase c, suponemos dadas:

- Su probabilidad a priori, $P(c)$.
- Su función de (masa de) probabilidad condicional, $P(x \mid c)$, la cual es Poisson(λ_c) con λ_c conocida.

Se pide:

1. (0.5 puntos) Sea el caso particular: $C = 2$, $P(c = 1) = P(c = 2) = \frac{1}{2}$, $\lambda_1 = 1$, $\lambda_2 = 2$ y $x = 2$. Determina la probabilidad *incondicional* de ocurrencia de $x = 2$, $P(x = 2)$.
2. (0.5 puntos) En el caso particular anterior, halla la probabilidad a posteriori $P(c = 2 \mid x = 2)$, así como la probabilidad de error si $x = 2$ se clasifica en la clase $c = 2$.
3. (0.5 puntos) Más generalmente, para cualquier número de clases C y cualesquiera probabilidades a priori, considera el caso en el que, dado un cierto $\tilde{\lambda} \in \mathbb{R}^+$, $\lambda_c = \tilde{\lambda}$ para todo c. En tal caso, existe una clase que no depende de x, c^*, en la que se puede clasificar todo x con mínima probabilidad de error. Determínala.
4. (0.5 puntos) En el caso general, prueba que el clasificador de Bayes para este problema puede expresarse como un clasificador basado en funciones discriminantes lineales como sigue (ln indica logaritmo natural):

$$c^*(x) = \arg\max_c \; g_c(x) \quad \text{con} \quad g_c(x) = w_c\, x + w_{c0}, \quad w_c = \ln \lambda_c \quad \text{y} \quad w_{c0} = \ln p(c) - \lambda_c$$

Solución:

1. $P(x = 2 \mid c = 1) = \frac{1}{2e} = 0.1839 \qquad P(x = 2 \mid c = 2) = \frac{2}{e^2} = 0.2707.$
   $P(x = 2) = 0.5 \cdot 0.1839 + 0.5 \cdot 0.2707 = 0.2273.$
2. $P(c = 2 \mid x = 2) = \frac{P(c=2) \cdot P(x=2 \mid c=2)}{P(x=2)} = \frac{0.5 \cdot 0.2707}{0.2273} = 0.5955.$
   $P(c \neq 2 \mid x = 2) = 1 - P(c = 2 \mid x = 2) = 0.4045.$
3. $c^*(x) = \arg\max_c P(c)\, P(x \mid c) = \arg\max_c P(c)\ \text{Poisson}(\lambda) = \arg\max_c P(c) \rightarrow c^* = \arg\max_c P(c).$
4. $$\begin{aligned} c^*(x) &= \arg\max_c \; \ln P(c) + \ln P(x \mid c) \\ &= \arg\max_c \; \ln P(c) - \lambda_c + x \ln \lambda_c - \ln x! \\ &= \arg\max_c \; x \ln \lambda_c + (\ln P(c) - \lambda_c) \end{aligned}$$