APELLIDOS: PL:

NOMBRE: DNI:

ESCUELA DE INGENIERÍA INGORMÁTICA

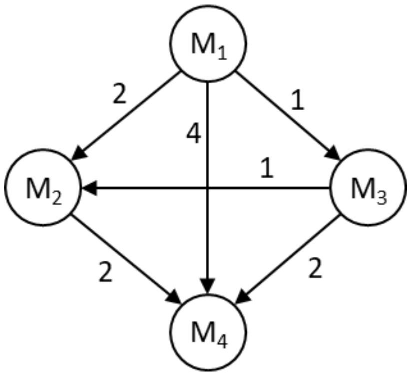
SISTEMAS INTELIGENTES

Convocatoria Extraordinaria. Miércoles 26 de junio de 2019.

I.- Búsqueda (3,5 puntos)

1.- *Sobre búsqueda en espacios de estados*

Consideremos el siguiente espacio de búsqueda abstracto, el algoritmo A* y los heurísticos cuyos valores se indican por extensión en la tabla



		h_1	h_2	h_3	h^*
(inic.)	M_1	2	2	2	
	M_2	0	1	2	
	M_3	2	2	2	
(obj.)	M_4	0	0	0	

a) **[0,25 puntos]** ¿Cuál es el valor de C*, y el de h* para cada uno de los nodos?

RESP: C*=3, h*(M1)=3, , h*(M2)=2 , h*(M3)=2, h*(M4)=0.

b) **[1,5 puntos]** ¿Qué propiedades tienen los heurísticos h1, h2 y h3? ¿Se puede establecer algún tipo de comparación entre ellos?

RESP: A la vista de los valores de la tabla y de los valores de h*, es evidente que se tiene la siguiente relación: Para todo n, h1(n)≤ h2(n)≤ h3(n)≤h*(n), pero en ningún caso se da la relación con la desigualdad estricta.

Con lo anterior, está claro que los tres heurísticos son admisibles. h1 no es monótono, ya que para M3 y M2 no se cumple la condición h1(M3) ≤h1(M2)+c(M3,M2). Sin embargo, h2 y h3 sí son monótonos ya que para estos dos heurísticos se cumple la condición h(n1) ≤h(n2)+c(n1,n2) para todos los nodos n1 y n2 que están conectado por una regla.

En consecuencia, la única relación formal que se puede establecer es la de "dominancia amplia" entre h2 y h3, concretamente h3 domina ampliamente a h2, que tiene como consecuencia que si h3 expande un nodo y h2 no lo expande (en el caso de que se utilicen los dos para guiar al algoritmo A*) entonces ese nodo debe cumplir que h3(n)=h2(n)=C*-g*(n). El único nodo que cumple esta condición es M4 (pero M4 es el único objetivo, con lo cual cualquier se elegirá para expandir con cualquier heurístico).

c) **[0,75 puntos]** De acuerdo con las propiedades anteriores, ¿Qué se puede decir con respecto a la expansión del nodo M_2 con cada uno de los tres heurísticos?

RESP: Dado que el heurístico h1 es admisible pero no monótono, las condiciones necesaria y suficiente de expansión de un nodo n se expresan en función de que exista o no un camino desde el inicial a n que sea C*-acotado (necesaria) estrictamente C*-acotado (suficiente).

La primera condición se enuncia como:

$$\forall n' \in P_{inic-n}\ g_P(n') + h(n') \leq C^*$$

La segunda es igual, pero con el menor estricto. gP(n') es el coste del inicial a n' a través de P.

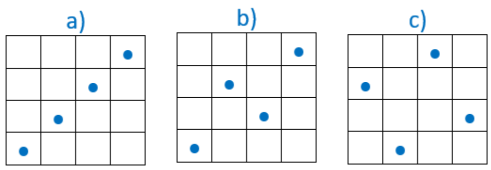
Desde el inicial (M1) a M2 hay dos caminos, el primero P1 es M1 -2->M2, para los nodos M1 y M2 tenemos que:
$g_{P1}(M1)+h1(M1)=0+2 < C^*=3$
$g_{p1}(M2)+h1(M2)=2+0 < C^*=3$
Luego P1 es estrictamente C*-acotado y en consecuencia el nodo M2 cumple la condición suficiente de expansión y por lo tanto si ejecutamos el algoritmo A* con el heurístico h1 sobre el espacio de búsqueda anterior, el nodo M2 se expandirá.

Como los heurísticos h2 y h3 son monótonos, las condiciones de expansión de un nodo n se expresan simplemente como: $g^*(n)+h(n) \leq C^*$ (suficiente) y $g^*(n)+h(n) < C^*$ (necesaria). Así, tenemos que:
$g^*(M2)+h2(M2) = 2+1 = 3 = C^*$
$g^*(M2)+h3(M2) = 2+2 = 4 > C^*=3$
En consecuencia, con h3 no se expandirá el nodo M3, ya que no se cumple la condición suficiente de expansión, y con h2 puede expandirse o no, ya que se cumple la condición necesaria, pero no la suficiente. Esto es coherente con el hecho de que los heurrísticos mejor informados expanden menos nodos en los experimentos.

2.- *Sobre algoritmos evolutivos.*
Consideremos un algoritmo genético que resuelve el problema de las N-reinas y que utiliza un esquema de codificación basado en permutaciones de [1 … N].

a) **[0,75 puntos]** Indicar cómo se pueden decodificar los cromosomas; es decir, dado un cromosoma indicar qué solución candidata representa para el problema (tened en cuenta que una solución candidata no tiene que ser necesariamente una solución real del problema). Si N=4, indicar la solución candidata que representan los cromosomas [1 2 3 4], [1 3 2 4] y [2 4 1 3], de acuerdo con el esquema de codificación propuesto.

RESP: Una interpretación natural de esta codificación es que el número en la posición i de la permutación representa la fila j en la que está la reina de la columna i. De este modo ya se garantiza una de las restricciones que tienen que cumplir las soluciones: que no haya dos reinas en la misma fila ni en la misma columna. De forma que si hay ataques siempre son en diagonal. Con esta interpretación, los cromosomas anteriores representarán las soluciones siguientes:

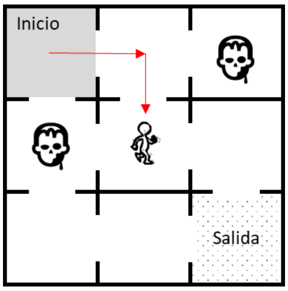


Consideremos ahora las dos formas siguientes de evaluar una solución candidata
i.- Contar el número de reinas que no son atacadas
ii.- Contar el número de pares de reinas que no se atacan entre sí.

b) **[0,75 puntos]** ¿Qué valores producen estas evaluaciones para las soluciones que representan los 3 cromosomas anteriores? ¿Son i. y ii. formas adecuadas de evaluar la calidad (fitness) de una solución candidata? ¿Cuál de las dos será mejor?

II.- Representación (3 puntos)

3.- **[1,5 puntos]** Se desea implementar un sistema de reglas (un bot) para ganar en un sencillo juego. El jugador controla a un policía que puede moverse entre distintas casillas de un tablero, siempre y cuando estén comunicadas por una puerta. En el tablero también hay zombies, de tal manera que si el jugador entra a una casilla con un zombie, muere. No obstante, el jugador puede disparar y matar a los zombies si éstos están a dos casillas de distancia. El objetivo del juego es llegar vivo a la casilla llamada "salida".
El dibujo muestra una posible situación de juego (NOTA: El juego debe funcionar para cualquier tablero, este es solo un ejemplo visual).



Diseñar un sistema de reglas en CLIPS tal que:

- El jugador se mueva de una casilla a otra siempre que haya una puerta entre ambas.
- El jugador no vuelva nunca a la casilla inmediatamente anterior.
- Si el jugador tiene la opción de desplazarse a varias casillas, escoja una en la que no haya zombies, siempre y cuando se siga respetando el punto anterior *(en el ejemplo, el jugador debe desplazarse a la derecha, ya que la casilla de encima es la inmediatamente anterior y la de la izquierda tiene un zombie)*.
- Si hay un zombie en una casilla que pueda ser alcanzada en dos pasos, el policía mate al zombie (éste desaparece del tablero). Esta acción debe tener prioridad sobre todas las anteriores *(en el ejemplo, el jugador debería matar al zombie de la esquina, ya que está a dos pasos de distancia)*.
- Si el jugador está en la misma casilla que un zombie, el jugador muere y se deberá mostrar un mensaje para indicarlo. Se deberá hacer lo necesario para asegurarse de que en este punto, el sistema de reglas se detiene (ninguna regla se activa)
- Si el jugador alcanza la casilla de "Salida", entonces gana y se mostrará un mensaje para indicarlo.

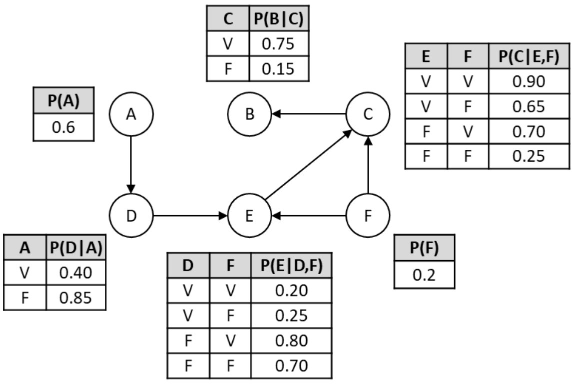
Para ello pueden utilizarse las siguientes plantillas de hechos en CLIPS:

```
(deftemplate jugador
        (slot casilla) ; Nombre de la casilla en la que se encuentra actualmente
        (slot ultimaC) ; Nombre de la última casilla visitada
)

(deftemplate puerta
        (slot origen) ; Casilla origen de la puerta
        (slot destino) ; Casilla a la que se llega a través de la puerta
        )

(deftemplate zombie
        (slot casilla) ; Nombre de la casilla en la que se encuentra este zombie
)
```

Solución:

```
(defrule mover
	?j <- (jugador (casilla ?c) (ultimaC ?u))
	(puerta (origen ?c) (destino ?d))
	(test (neq ?d ?u))
	=>
	(assert (jugador (casilla ?d) (ultimaC ?c)))
	(retract ?j)
)

(defrule escapar
	(declare (salience 5))
	?j <- (jugador (casilla ?c) (ultimaC ?u))
	(puerta(origen ?c) (destino ?d))
	(not (exists (zombie (casilla ?d))))
	(test (neq ?u ?d))
	=>
	(assert (jugador (casilla ?d) (ultimaC ?c)))
	(retract ?j)
)

(defrule matar
	(declare (salience 10))
	(jugador (casilla ?c))
	(puerta (origen ?c) (destino ?d))
	(puerta (origen ?d) (destino ?d2))
	?z <- (zombie (casilla ?d2))
	(test (neq ?c ?d2))
	=>
	(printout t "Has matado al zombie en " ?d2)
	(retract ?z)
)

(defrule perder
	?j <- (jugador (casilla ?c))
	(zombie (casilla ?c))
	=>
	(retract ?j)
	(printout t "Has perdido")
)

(defrule ganar
	?j <- (jugador (casilla "salida"))
	=>
	(retract ?j)
	(printout t "Has ganado")
)
```

4.- Dada la siguiente red bayesiana, responder razonadamente a las siguientes preguntas:



P(A)
0.6

C	P(B|C)
V	0.75
F	0.15

E	F	P(C|E,F)
V	V	0.90
V	F	0.65
F	V	0.70
F	F	0.25

A	P(D|A)
V	0.40
F	0.85

D	F	P(E|D,F)
V	V	0.20
V	F	0.25
F	V	0.80
F	F	0.70

P(F)
0.2

a) **[0,5 puntos]** Sabiendo que la variable E es falsa, ¿podemos afirmar que las variables A y B son independientes? Justifica la respuesta utilizando el criterio de D-separación, indicando todos los posibles caminos y por qué están, o no están, bloqueados.

Solución: Existen dos caminos distintos entre las variables A y B.

Camino 1: A -> D -> E -> C -> B

Camino 2: A -> D -> E <- F -> C -> B

En el primer camino existe un bloqueo en el nodo E, ya que **no** es un nodo "cabeza-cabeza", y está observado. En el segundo camino no hay ningún bloqueo, ya que aunque el nodo E es cabeza-cabeza, al estar observado no produce bloqueo. Por tanto, dado que existe un camino no bloqueado, **A y B <u>no</u> son independientes.**

b) **[1 punto]** Sabiendo que A y F se cumplen, pero B y D no, ¿cuál es la probabilidad de que se cumpla C?.

Solución:

$$P(C|A,\neg B,\neg D,F)=\frac{P(C,A,\neg B,\neg D,F)}{P(A,\neg B,\neg D,F)}=\frac{P(C,A,\neg B,\neg D,F)}{P(C,A,\neg B,\neg D,F)+P(\neg C,A,\neg B,\neg D,F)}$$

$$P(C,A,\neg B,\neg D,F)=P(C,A,\neg B,\neg D,F,E)+P(C,A,\neg B,\neg D,F,\neg E)$$
$$P(\neg C,A,\neg B,\neg D,F)=P(\neg C,A,\neg B,\neg D,F,E)+P(\neg C,A,\neg B,\neg D,F,\neg E)$$

$$P(C,A,\neg B,\neg D,F,E)=P(C|F,E)\cdot P(A)\cdot P(\neg B|C)\cdot P(\neg D|A)\cdot P(F)\cdot P(E|\neg D,F)$$
$$=0.9\cdot 0.6\cdot 0.25\cdot 0.6\cdot 0.2\cdot 0.8=0.01296$$

$$P(C,A,\neg B,\neg D,F,\neg E)=P(C|F,\neg E)\cdot P(A)\cdot P(\neg B|C)\cdot P(\neg D|A)\cdot P(F)\cdot P(\neg E|\neg D,F)$$
$$=0.7\cdot 0.6\cdot 0.25\cdot 0.6\cdot 0.2\cdot 0.2=0.00252$$

$$P(\neg C,A,\neg B,\neg D,F,E)=P(\neg C|F,E)\cdot P(A)\cdot P(\neg B|\neg C)\cdot P(\neg D|A)\cdot P(F)\cdot P(E|\neg D,F)$$
$$=0.1\cdot 0.6\cdot 0.85\cdot 0.6\cdot 0.2\cdot 0.8=0.004896$$

$$P(\neg C,A,\neg B,\neg D,F,\neg E)=P(\neg C|F,\neg E)\cdot P(A)\cdot P(\neg B|\neg C)\cdot P(\neg D|A)\cdot P(F)\cdot P(\neg E|\neg D,F)$$
$$=0.3\cdot 0.6\cdot 0.85\cdot 0.6\cdot 0.2\cdot 0.2=0.003672$$

$$P(C|A,\neg B,\neg D,F)=\frac{0.01296+0.00252}{0.01296+0.00252+0.004896+0.003672}=\mathbf{0.6437}$$

III.- Aprendizaje (3,5 puntos)

5.- **[0,5 puntos]** Explica en que consiste la validación cruzada y sus ventajas y/o desventajas.

Solución: Es una de las opciones para evaluar la capacidad de generalización de un modelo y consiste en dividir el conjunto de instancias en *k* grupos de forma que el aprendizaje se hará *k* veces, cada vez se usarán *k*-1 de los grupos para generar el modelo y el grupo restante para validar. Este sistema permite adaptarse a distintos tamaños de instancias simplemente usando un valor de *k* apropiado. Sin embargo, al tener que repetir *k* veces el ajuste del modelo el coste computacional será alto, especialmente con conjuntos de datos muy grandes.

6.- **[1 puntos]** Dado el siguiente conjunto de datos donde Play es la variable a predecir, indica razonadamente que variable se seleccionaría como nodo raíz de un árbol de decisión usando el algoritmo ID3.

Outlook	Temp.	Humidity	Wind	Play
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Solución:

$$H(Play) = -\frac{4}{10}\cdot\log_2\frac{4}{10} - \frac{6}{10}\cdot\log_2\frac{6}{10} = 0.97$$

$$\begin{aligned}IG(Play, Outlook) &= 0.97 - H(Play|Outlook)\\ &= 0.97 - \left(\frac{4}{10}\cdot 1 + \frac{3}{10}\cdot 0 + \frac{3}{10}\cdot\left(-\frac{2}{3}\cdot\log_2\frac{2}{3} - \frac{1}{3}\cdot\log_2\frac{1}{3}\right)\right)\\ &= 0.97 - (0.4 + 0 + 0.28) = 0.29\end{aligned}$$

$$\begin{aligned}IG(Play, Temp) &= 0.97 - H(Play|Temp)\\ &= 0.97\\ &- \left(\frac{3}{10}\cdot\left(-\frac{2}{3}\cdot\log_2\frac{2}{3} - \frac{1}{3}\cdot\log_2\frac{1}{3}\right) + \frac{4}{10}\cdot 1 + \frac{3}{10}\right.\\ &\left.\cdot\left(-\frac{2}{3}\cdot\log_2\frac{2}{3} - \frac{1}{3}\cdot\log_2\frac{1}{3}\right)\right) = 0.97 - (0.28 + 0.4 + 0.28) = 0.01\end{aligned}$$

$$\begin{aligned}IG(Play, Humidity) &= 0.97 - H(Play|Humidity)\\ &= 0.97 - \left(\frac{6}{10}\cdot\left(-\frac{1}{6}\cdot\log_2\frac{1}{6} - \frac{5}{6}\cdot\log_2\frac{5}{6}\right) + \frac{4}{10}\cdot\left(-\frac{3}{4}\cdot\log_2\frac{3}{4} - \frac{1}{4}\cdot\log_2\frac{1}{4}\right)\right)\\ &= 0.97 - (0.39 + 0.32) = 0.26\end{aligned}$$

$$\begin{aligned}IG(Play, Wind) &= 0.97 - H(Play|Wind)\\ &= 0.97 - \left(\frac{5}{10}\cdot\left(-\frac{3}{5}\cdot\log_2\frac{3}{5} - \frac{2}{5}\cdot\log_2\frac{2}{5}\right) + \frac{5}{10}\cdot\left(-\frac{1}{5}\cdot\log_2\frac{1}{5} - \frac{4}{5}\cdot\log_2\frac{4}{5}\right)\right)\\ &= 0.97 - (0.49 + 0.36) = 0.12\end{aligned}$$

Por tanto, el nodo raíz del árbol será el correspondiente a la variable *Outlook*.
En realidad, la entropía de *Play* no hace falta calcularla porque es suficiente con encontrar la entropía condicional de *Play* con otra variable que sea menor. Tampoco es necesario calcular

la entropía condicional con *Temp* si se observa que los valores de *Play* están muy uniformemente distribuidos por cada uno de *Temp*.

7.- **[1 punto]** Considera una red neuronal con **función de activación lineal**. Tiene dos entradas (a, b), una neurona oculta (c) y una unidad de salida (d). Habrá un total de cinco pesos $(w_{ca}, w_{cb}, w_{c0}, w_{dc}, w_{d0})$ que vamos a inicializar todos a 0.1 y con una tasa de aprendizaje $\eta = 0.3$. Calcula el valor de los pesos utilizando el algoritmo de propagación hacía atrás para la siguiente instancia:

a	b	d
1	1	0

Recuerda que en este caso: $\delta_k = (y_k - o_k)$ y $\delta_h = \sum_k \delta_k w_{kh}$

Solución: $o_c = 0.1 \cdot 1 + 0.1 \cdot 1 + 0.1 \cdot 1 = 0.3$; $o_d = 0.1 \cdot 1 + 0.1 \cdot 0.3 = 0.13$
$\delta_d = 0 - 0.13 = -0.13$; $\delta_c = -0.13 \cdot 0.1 = -0.013$

$$w_{d0} = 0.1 + 0.3 \cdot (-0.13) \cdot 1 = 0.061$$
$$w_{dc} = 0.1 + 0.3 \cdot (-0.13) \cdot 0.3 = 0.0883$$
$$w_{c0} = 0.1 + 0.3 \cdot (-0.013) \cdot 1 = 0.0961$$
$$w_{ca} = 0.1 + 0.3 \cdot (-0.013) \cdot 1 = 0.0961$$
$$w_{cb} = 0.1 + 0.3 \cdot (-0.013) \cdot 1 = 0.0961$$

8.- **[1 punto]** Dado el siguiente conjunto de entrenamiento {(1,+),(2,+),(4,-),(5,-),(6,+)} al que se la he aplicado SVM con kernel $K(x,y) = (xy+1)^2$, obteniendo $\alpha_1 = 0$, $\alpha_2 = 2.499$, $\alpha_3 = 0$, $\alpha_4 = 7.331$, $\alpha_5 = 4.832$. Calcula la predicción de este modelo para 6.5.

Solución: En primer lugar, es necesario calcular el valor del b, usamos el primer vector soporte para este cálculo
$b = 1 - (2.499 \cdot 1 \cdot (2 \cdot 2 + 1)^2 + 7.331 \cdot (-1) \cdot (5 \cdot 2 + 1)^2 + 4.832 \cdot 1 \cdot (6 \cdot 2 + 1)^2 \approx 9$

$$g(6.5) = 2.499 \cdot 1 \cdot (2 \cdot 6.5 + 1)^2 + 7.331 \cdot (-1) \cdot (5 \cdot 6.5 + 1)^2 + 4.832 \cdot 1 \cdot (6 \cdot 6.5 + 1)^2 + 9 = 489.804 - 8227.215 + 7731.2 + 9 = 2.789$$

Según estos resultados la entrada 6.5 se clasifica como '+'.

$\forall n' \in P_{inic-n}\ g_P(n') + h(n') \leq C^*$		$g^*(n) + h(n) \leq C^*$
$h_1(n) = h_2(n) = C^* - g^*(n)$		$h(n_1) \leq h(n_2) + c(n_1, n_2)$
$E(\boldsymbol{w}) = \frac{1}{2} \sum_d \sum_k (y_{kd} - o_{kd})^2$	$K(\boldsymbol{x}, \boldsymbol{y}) = e^{-\lVert x-y \rVert^2/(2\sigma^2)}$	$H(X) = -\sum_{i=1}^{k} p_i \log_2 p_i$
$\Delta w_i = \eta(y_d - \boldsymbol{w}\boldsymbol{x}_d)x_{di}$	$b = \frac{1}{\lvert VS \rvert} \sum_{i \in VS} y_i - \boldsymbol{w}\boldsymbol{x}_i$	$IG(X,Y) = H(X) - H(X \mid Y)$
$g(\cdot) = \begin{cases} 1 \text{ si } \boldsymbol{wx} > 0 \\ -1 \text{ en otro caso} \end{cases}$	$d = \frac{2}{\lVert \boldsymbol{w} \rVert}$	$g(\boldsymbol{x}) = \sum_{i \in VS} \alpha_i y_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b$
$K(x_i, x_j) = \phi(x_i)\phi(x_j)$	$dist(\vec{x}, \vec{y}) = \left(\sum_i \lvert x_i - y_i \rvert^p \right)^{1/p}$	$\boldsymbol{w} = \sum_{i \in VS} \alpha_i y_i \boldsymbol{x}_i$
$K(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{xy} + 1)^d$	$g(x) = \frac{1}{1 - e^{-x}}$	$y(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{x} + b$
$g(\boldsymbol{x}) = \boldsymbol{wx} + b$	$H(X \mid Y) = \sum_{y \in Y} p(y) H(X \mid Y = y)$	$w_{ji} \leftarrow w_{ji} + \eta \delta_j x'_{ji}$