Examen-Mayo-2021.pdf

Anónimo

Sistemas Inteligentes

4º Grado en Ingeniería Informática del Software

Escuela de Ingeniería Informática
Universidad de Oviedo

APELLIDOS:	PL:
NOMBRE:	DNI:

ESCUELA DE INGENIERÍA INFORMÁTICA

SISTEMAS INTELIGENTES

Examen Final de Teoría. Lunes 24 de mayo de 2021.

I.- Búsqueda

1.- [2 puntos] Consideremos la solución del problema de las N-reinas con A* y el espacio de búsqueda completo. Dado un estado *n*, sea *X* el máximo de reinas que atacan a otra reina en el estado *n* y un heurístico *h* definido como:

$$h(n) = \begin{cases} 0 & si\ X = 0 \\ max(1, X-1) & si\ X > 0 \end{cases}$$

Responder de forma razonada a las siguientes cuestiones considerando un estado inicial en el que todas las reinas están en la primera fila.

a) ¿Está *h* bien definido?
RESP: Sí lo está ya que en cualquier objetivo X=0 y por lo tanto h(objetivo)=0, y en cualquier nodo n no objetivo h(n)>=1>0.

b) ¿Cuáles son los valores máximos que pueden tomar *h* y *h** para un nodo del espacio de búsqueda?
RESP: Una reina puede estar atacada por las otras N-1 como máximo, por lo tanto, h(n)<=(N-1)-1 para todo nodo n. Por otra parte, h*(n)<=N ya que puede haber algún estado en el que sea preciso mover todas las reinas para llegar a una solución, por ejemplo, para N=4 si todas las reinas están en una diagonal.

c) ¿Cuáles son los valores de *h*(inicial) y de *C**?
RESP: C*=h*(inicial) = N-1 ya que al menos una de las reinas debe permanecer en la primera fila. h(inicial) = (N-1)-1 = N-2 ya que cada reina está atacada por las N-1 restantes.

d) Para N=4, indicar dos estados *n1* y *n2* distintos del nodo inicial y calcular los valores *de h* y de *h** para estos nodos.
RESP:

n1 =

	*		
*			*
		*	

n2=

	*		
*			
			*
		*	

h(n1) = 3-1=2, h*(n1)=1 h(n2)=2-1=1, h*(n2)=2

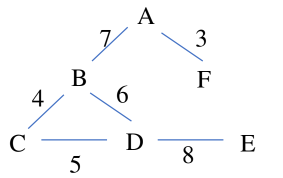
e) Para N= 4, ¿hay algún estado *n* tal que *h*(*n*)>*h**(*n*)?
RESP: Sí lo hay, por ejemplo el estado n1 anterior.

f) ¿Es *h* consistente?
RESP: Dado que h(n1)=2 > h*(n1) =1, h no es admisible y por lo tanto tampoco es consistente.

g) ¿Se puede hacer algún tipo de comparación de *h* con el heurístico que calcula el número de pares de reinas que se atacan en un estado *n*?
RESP: Tanto h como el heurístico que calcula el número de pares de reinas que se atacan son no admisibles, por lo tanto, no se puede establecer una comparación estricta entre ellos. Sin embargo, se puede observar que h(n) es menor que el número de pares de reinas que se atacan, y en general bastante menor. En consecuencia, h estará más próximo a h* y así se podría esperar un comportamiento mejor en cuanto

a la calidad de la solución. Pero eso debe ser confirmado o descartado experimentalmente.

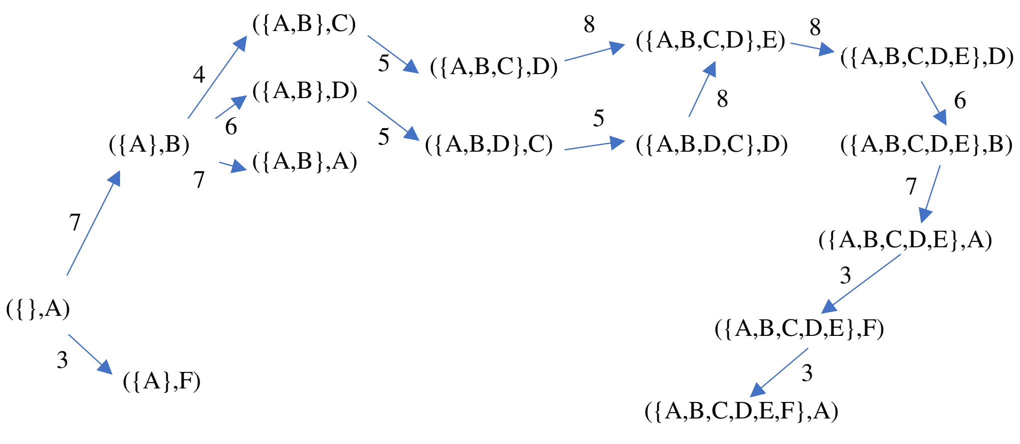
2.- En este ejercicio se pide resolver una variante del problema del viajante de comercio con búsqueda en espacios de estados. El problema consiste en que tenemos un conjunto de ciudades que no están conectadas directamente todas con todas y se trata de encontrar una ruta de coste mínimo que parta de una ciudad y regrese a la misma ciudad después de visitar cada una de las demás ciudades al menos una vez. Un ejemplo de conexiones es el siguiente:



Concretamente, se pide:

a) **[0,75 puntos]** Definir el espacio de búsqueda para un problema genérico.
RESP: La representación que utilizamos para el problema TSP original, considerando el grafo de ciudades totalmente conectado, se puede extender a esta versión del problema teniendo en cuenta que ahora se puede pasar varias veces por la misma ciudad, y en concreto que desde la ciudad actual se puede ir a una ciudad ya visitada, si hay una conexión, por supuesto.
Con esta idea, un estado se puede representar por un par (C, X), siendo C un conjunto de ciudades ya visitadas y X la ciudad actual. Las reglas aplicables a un estado (C,X) vienen dadas por todas las conexiones desde X a otra ciudad cualquiera. Para cada ciudad Y conectada con X, tendremos un sucesor (C ∪ {X}, Y) y el coste de la regla es el coste del arco (X,Y) en el grafo de conexiones.

b) **[0,75 puntos]** Dibujar una parte significativa del espacio de búsqueda para la instancia del problema definida por el grafo anterior, suponiendo que la ciudad de partida y llegada es A.
RESP: Una parte representativa del espacio de búsqueda para la instancia definida por el grafo de conexiones anterior, que incluye una solución y dos caminos desde el inicial hasta un nodo intermedio, es la siguiente:



ENGINEERING THE FUTURE
indra
En Indra sentimos
#OrgulloIngeniero
Tú también estás a un paso de transformar el futuro.
Pero... ¿sabes qué opina la sociedad de la personalidad de los ingenieros?
Que no tenemos creatividad, llevamos la lógica al extremo, somos indescifrables o que solo pensamos en números son algunos de los estereotipos.
¿A que no estás de acuerdo?
¡Responde a estos estereotipos y derriba estos mitos!
Comparte los motivos por los que sientes #OrgulloIngeniero y ayúdanos a poner en valor a la ingeniería.
orgullo-ingeniero.indra.es
10:16
Para ti
Siguiendo
CONFÍA EN MI
SOY INGENIERO

II.- Representación

3.- [0.75 punto] Dado el siguiente sistema conjunto de reglas

R1: X_7, X_3 -> X_{10} R2: X_4, X_8 -> X_{10} R3: X_9 -> X_1 R4: X_8-> X_7 R5: X_1, X_8-> X_{10} R6: X_9, X_7, X_6 -> X_1 R7: X_1, X_9 -> X_7 R8: X_1 -> X_5 R9: X_{10},X_1 -> X_2	• **[0.375 puntos]** Aplicando encadenamiento hacia delante, describe el proceso de inferencia resultante suponiendo más prioritaria la regla con más condiciones en su antecedente (en caso de igualdad, se aplicará el principio de prioridad, aplicándose primero la regla con menor subíndice). La Base de Hechos inicial es BH: { X_8 X_9} y la meta es X_{10} • **[0.375 puntos]** Aplicando encadenamiento hacia detrás, describir el proceso de inferencia resultante suponiendo que la regla más prioritaria es la de subíndice mayor, que la meta a obtener es X_2 y que la base de hechos inicial es BH: { X_3, X_9}

RESP a)
BH0: {G, H}, CC0= {R4, R7} . Aplicamos R4
BH1: {A, G, H}, CC1= {R7, R8, R9}. Aplicamos R8
BH2: {I, A, G, H}, CC2= {R6, R7,R9}

RESP b)
BH0: {C, H}, Aplicando R6, B es cierto si I y A lo son.

Submeta I
CC={R2,R3,R8}. Aplicamos R8. I es cierto si A y G lo son

Submeta G: No se puede verificar

CC={R2,R3}. Aplicamos R3. I es cierto si D y G lo son

Submeta G: No se puede verificar

CC={R2}. Aplicamos R2. I es cierto si F y C lo son. C está en la base de hechos. Intentamos verificar F.

Submeta F: CC={R9}. F es cierto si A y H lo son. H está en la base de hechos.
Submeta A. CC={R4}

4.- Dada la red bayesiana de la figura, utilizando el criterio de D-separación o la condición de Markov

a) **[0.75 puntos]** Indica, utilizando el criterio de D-separación, el manto de Markov o la condición de Markov, si son ciertas o falsas las siguientes igualdades:
- $P(Vacuna, Mascarilla) = P(Vacuna) \cdot P(Mascarilla)$
- $P(Mascarilla|Trabajo, Viaje, Enfermo) = P(Mascarilla|Trabajo, Viaje)$
- $P(Vacuna|Estrés) = P(Vacuna)$

RESP:
1) Esta igualdad implica que VACUNA y MASCARILLA son independientes y lo son porque todos los caminos entre los dos nodos están bloqueados:
VACUNA -> TRABAJO <- MASCARILLA bloqueado en TRABAJO
VACUNA -> TRABAJO -> VIAJE -> ESTRÉS <- MASCARILLA bloqueado en ESTRÉS

VACUNA -> TRABAJO -> ENFERMO <- VIAJE -> ESTRÉS <- MASCARILLA bloqueado en ENFERMO y ESTRÉS

2) Esta igualdad implica que MASCARILLA y ENFERMO son condicionalmente independientes dados TRABAJO y VIAJE. El manto de Markov de ENFERMO son precisamente TRABAJO y VIAJE por lo que ENFERMO es condicionalmente independiente de MASCARILLA (también de VACUNA y ESTRÉS) dados TRABAJO y VIAJE.

3) Esta igualdad implica que VACUNA y ESTRÉS son marginalmente independientes. En este caso la igualdad no es cierta porque existe el camino VACUNA -> TRABAJO -> VIAJE -> ESTRÉS que no está bloqueado.

b) **[0.75 puntos]** Calcula $P(\neg Trabajo = X3, Viaje = X4, Estrés = X6 | Vacuna = X1, \neg Mascarilla = X2)$

RESP: $P(\neg X3, X4, X6 | X1, \neg X2) = \frac{P(X1, \neg X2, \neg X3, X4, X6)}{P(X1, \neg X2)}$

$$P(X1, \neg X2, \neg X3, X4, X6) = \sum_{X5} P(X1, \neg X2, \neg X3, X4, X5, X6)$$

$$= \sum_{X5} P(X1) \cdot P(\neg X2) \cdot P(\neg X3 | X1, \neg X2) \cdot P(X4 | \neg X3) \cdot P(X5 | \neg X3, X4) \cdot P(X6 | \neg X2, X4)$$

$$= (0.8 \cdot 0.9 \cdot 0.3 \cdot 0.7 \cdot 0.3 \cdot 0.4) + (0.8 \cdot 0.9 \cdot 0.3 \cdot 0.7 \cdot 0.7 \cdot 0.4)$$

$$= 0.018144 + 0.042336 = 0.06048$$

Dado que VACUNA y MASCARILLA son independientes $P(X1, \neg X2) = P(X1) \cdot P(\neg X2) = 0.8 \cdot 0.9 = 0.72$

$$P(\neg X3, X4, X6 | X1, \neg X2) = \frac{0.06048}{0.72} = 0.084$$

c) **[0.75 puntos]** Utilizando los números aleatorios siguientes: 0.04, 0.54, 0.65, 0.17, 0.32, 0.83, 0.97, 0.58 0.09, 0.25; indica paso a paso cuál sería la muestra generada por el método de muestreo estocástico para calcular la probabilidad $P(Viaje, \neg Enfermo | \neg Vacuna, Estrés)$.

Solución: Primero hay que ordenar las variables en orden topológico. Uno posible es VACUNA, MASCARILLA, TRABAJO, VIAJE, ENFERMO y ESTRÉS.

Con VACUNA, como la probabilidad de true es 0.8 y el número aleatorio es 0.04 se queda VACUNA=true.

Con MASCARILLA, como la probabilidad de true es 0.1 y el número aleatorio es 0.54 se queda MASCARILLA=false.

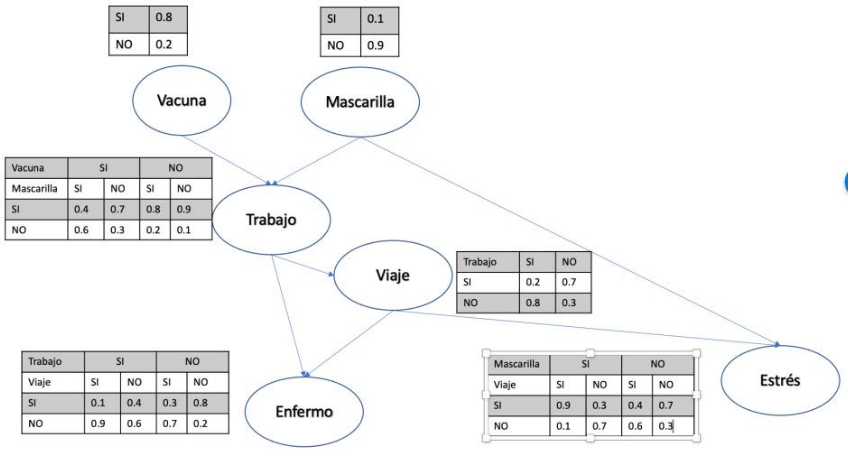
Con TRABAJO, como la probabilidad de true (dado VACUNA=true, MASCARILLA=false) es 0.7 y el número aleatorio es 0.65 se queda TRABAJO=true.

Con VIAJE, como la probabilidad de true (dado TRABAJO=true) es 0.2 y el número aleatorio es 0.17 se queda VIAJE=true.

Con ENFERMO, como la probabilidad de true (dado TRABAJO=true, VIAJE=true) es 0.1 y el número aleatorio es 0.32 se queda ENFERMO=false.

Con ESTRÉS, como la probabilidad de true (dado MASCARILLA=false, VIAJE=true) es 0.4 y el número aleatorio es 0.83 se queda ESTRÉS=false.

Por tanto, la muestra generada es $\{Vacuna, \neg Mascarilla, Trabajo, Viaje, \neg Enfermo, \neg Estrés\}$



III.- Aprendizaje

5.- [0.5 puntos] Explica si el error en entrenamiento de un clasificador es una estimación optimista o pesimista.

RESP: Es optimista, puesto que calcula el error con los mismos ejemplos usados para construir el modelo

6.- [0.5 puntos] Juan odia ver mensajes de spam en su bandeja de entrada. Sin embargo, no le importa comprobar periódicamente la carpeta *Correos no deseados* en busca de correos marcados incorrectamente como spam.
Patricia nunca consulta la carpeta *Correos no deseados* Prefiere ver correos electrónicos no deseados en su bandeja de entrada que perder correos auténticos sin saberlo.

Si construimos un clasificador de SPAM cuya precisión es 80% y la cobertura es del 60%, ¿quién estará más satisfecho con el clasificador? ¿Por qué?

RESP: Para responder a esta pregunta, pensemos en lo que significa tener alta precisión y baja cobertura con respecto al SPAM y, a la inversa, lo que significa tener alta cobertura y baja precisión con respecto al SPAM.
Alta precisión y baja cobertura con respecto al SPAM: cualquier cosa que el modelo clasifique como SPAM es probablemente SPAM. Sin embargo, muchos correos electrónicos que son realmente SPAM se clasifican erróneamente como NO SPAM. Es probable que el usuario vea algunos mensajes de SPAM en su bandeja de entrada, pero nunca tendrá que ir al directorio de Correos no deseados para buscar mensajes marcados incorrectamente como SPAM.
Alta cobertura y baja precisión con respecto al SPAM: el modelo filtra todos los correos electrónicos SPAM, pero también clasifica incorrectamente algunos correos genuinos como SPAM. El usuario nunca verá correos electrónicos SPAM en su bandeja de entrada, pero tendrá que comprobar periódicamente el directorio Correos no deseados para encontrar correos genuinos marcados incorrectamente como SPAM.

Dado que el clasificador consigue una precisión mayor que el recall, es más probable que Patricia esté satisfecho con el clasificador

7.- [0.5 puntos] Indica razonadamente si incrementar el número nodos en una capa oculta de un perceptrón multicapa siempre repercute en una mejora de la generalización.

RESP: No siempre, puede conducir al sobreajuste

8.- [1 punto] Tenemos un perceptrón con dos entradas x1 y x2 donde el valor de la entrada de sesgo es 1 y los pesos asignados a cada entrada son (w0 = 2, w1 = 1, w2 = −1). La función de activación de la neurona es la función signo, de manera que los ejemplos se clasifican como positivos si el valor en el hiperplano es positivo o cero y como negativo en otro caso.
a) Escribe la ecuación del hiperplano separador e indica cual es la clasificación que daría para los puntos (0, 0) y (−1, 2).
b) Supongamos que el punto (−1, 0) es de la clase negativa. Dados los pesos anteriores aplica la regla del perceptrón para que se clasifique correctamente y da la ecuación del hiperplano resultante usando una tasa de aprendizaje de 0.5.

RESP
a)Ecuación del hiperplano: g(x1,x2)= x1-x2+2; g(0,0)=2, por tanto clase positiva g(-1,2)=-1, por tanto clase negativa.
b)g(x1,x2)=2x1-x2+1

9.- [0.5 puntos] En relación al clasificador SVM
a) Supón que tienes un clasificador binario SVM lineal. Considera un punto que está clasificado correctamente, y que está lejos de la frontera de decisión. Si eliminamos el ejemplo del conjunto de entrenamiento y volvemos a entrenar el clasificador, ¿cambiará la frontera de decisión o permanecerá igual? Explica tu respuesta

RESP: La frontera de decisión de la SVM está definido por los vectores de soporte. Debido a la regularización, si sólo eliminamos un punto del conjunto de entrenamiento, y no era un vector de soporte, es poco probable que el límite cambie.

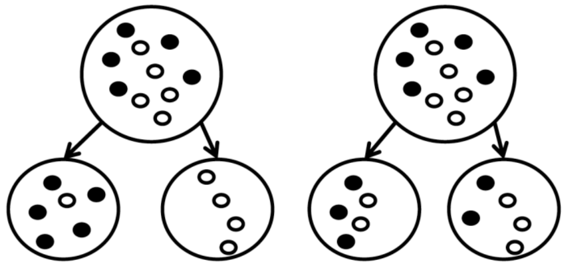
b) Si los ejemplos de entrenamiento son linealmente separables, ¿cuántas fronteras de decisión pueden separar los ejemplos positivos de los negativos en un problema de clasificación binario? ¿Cuántas produce un SVM lineal con margen duro? ¿Qué características tienen, es decir, que propiedad verifican los hiperplanos que produce SVM?

RESP: Hay infinitas fronteras de decisión que separan los puntos de datos positivos de los negativos. El algoritmo SVM encuentra la frontera (hiperplano) con el máximo margen. Esta frontera minimiza el límite superior del error de clasificación.

10.- [0.5 puntos]. Supongamos que tenemos N ejemplos definidos mediante k variables binarias para predecir si este verano va a hacer calor o no. ¿Cuál es la profundidad un árbol de decisión que construyamos a partir de ese dataset (la profundidad de un árbol es la longitud del camino más largo desde la raíz a cada hoja)?. ¿Y si las k variables son continuas?

RESP: En el caso de las variables continuas no se puede determinar, puesto que hay infinotas combinaciones (variable, punto de corte). En el caso de que las variables sean dsicretas, la profundidad es la longitud del camino más largo desde la raíz a una hoja. Esta situación se da cuando en el camino aparecen los k nodos.

La siguiente figura representa la división del conjunto de entrenamiento que se produce cuando utilizo como raíz la variable A o la B. Los círculos blancos y negros representan la etiqueta de la clase.



A B

¿Cuál es la entropía de cada división? ¿Qué división produce más nodos puros?

RESP: Entropy(A)=6/10*(-5/6*log(5/6,2)-1/6*log(1/6,2))+4/10*(-4/4*log(4/4,2))=0.39
Entropy(B)=5/10*(-3/5*log(3/5,2)-2/5*log(2/5,2))+5/10*(-2/5*log(2/5,2)-3/5*log(3/5,2))=0.97
A produce nodos más puros.

$P(x_1,\dots,x_i)=\sum_{x_{i+1},\dots,x_n} P(x_1,\dots,x_i,x_{i+1},\dots,x_n)=\sum_{x_{i+1}} P(x_1,\dots,x_i,x_{i+1})$		
$P(x\|y) \equiv P(X=x\|Y=y)=\frac{P(x,y)}{P(y)}$	$P(X,Y)=P(X\|Y)P(Y)=P(Y\|X)P(X)$	
$P(Y)=\sum_{i=1}^{m} P(Y\|x_i)P(x_i))$	$P(X_1=x_1,\dots,X_n=x_n)=\prod_{i=1}^{n} P(X_i=x_i\|X_{i+1}=x_{i+1},\dots,X_n=x_n)$	
$P(X\|Y)=P(X)$	$P(X\|Y,K)=P(X\|K)$	$P(X\|Y,K)\neq P(X\|K)$
$P(A\|MB(A),B)=P(A\|MB(A))$	$P(x_1,\dots,x_n)=\prod_{i=1}^{n} P(X_i=x_i\|Padres(X_i))$	
	$P(X\|Padres(X))=P(X\|Padres(X),NoDescendientes(X))$	
$E(\boldsymbol{w})=\frac{1}{2}\sum_{d}\sum_{k}(y_{kd}-o_{kd})^2$	$\delta_k=o_k(1-o_k)(y_k-o_k)$ $\delta_h=o_h(1-o_h)\sum_k w_{kh}\delta_k$	$\Delta w_i=\eta(y_d-\boldsymbol{wx}_d)x_{di}$ $g(x)=\frac{1}{1-e^{-x}}$
$dist(\vec{x},\vec{y})=\left(\sum_i \|x_i-y_i\|^p\right)^{1/p}$	$E(\boldsymbol{w})=\frac{1}{2}(y_d-\boldsymbol{wx}_d)^2$	$g(\cdot)=\begin{cases}1 \text{ si } \boldsymbol{wx}>0\\ -1 \text{ en otro caso}\end{cases}$
$H(X)=-\sum_{i=1}^{k} p_i \log_2 p_i$	$H(X\|Y)=\sum_{y\in Y} p(y)H(X\|Y=y)$	$K(x_i,x_j)=\phi(x_i)\phi(x_j)$ $K(\boldsymbol{x},\boldsymbol{y})=(\boldsymbol{xy}+1)^d$
$b=\frac{1}{\|VS\|}\sum_{i\in VS} y_i-\boldsymbol{wx_i}$	$\boldsymbol{w}=\sum_{i\in VS}\alpha_i y_i \boldsymbol{x_i}$	$d=\frac{2}{\|\|\boldsymbol{w}\|\|}$