Inteligência Artificial – 2022/23

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

10/Dezembro/2022 ─ 9:00h ─ Duração: 2:00h

2º Teste (Modelo A)
– Sem consulta–

I) [4,5val] Quatro adeptos estão no Qatar para assistir ao Campeonato do Mundo de Futebol. Eles estão num bar a beber e conversar sobre os seus jogadores preferidos. Pretendemos descobrir qual é o melhor jogador na opinião de cada um, a partir de algumas pistas que temos à nossa disposição. Após análise do problema decidiu-se utilizar a programação por conjuntos de resposta, no dialeto do CLINGO, para resolver o problema. Para facilitar a sua resolução, o seguinte conhecimento já se encontra devidamente modelado:

```
país(portugal).      jogador(ronaldo).   bebida(café).   adepto(1..4).
país(brasil).        jogador(messi).     bebida(água).
país(frança).        jogador(mbappé).    bebida(sumo).
país(argentina).     jogador(neymar).    bebida(chá).

1{nacionalidade(T,P): país(P)}1 :- adepto(T).
1{preferido(T,J): jogador(J)}1 :- adepto(T).
1{bebe(T,B): bebida(B)} :- adepto(T).
```

Para responder às seguintes questões, pode utilizar predicados auxiliares, caso entenda necessário.

a) Quantos modelos estáveis (conjuntos de resposta) são obtidos com o programa anterior? Indique, sumariamente, como obteve esse valor.
b) Sabe-se que cada jogador é preferido por exatamente um dos adeptos. Indique que regra(s) acima garante(m) esta restrição ou, caso ela não esteja já garantida pelas regras acima, especifique as regras/restrições necessárias para a garantir.
c) Especifique as regras/restrições necessárias para garantir que os conjuntos de resposta estão de acordo com cada uma das seguintes afirmações:
   i. Não há mais do que dois adeptos a beber a mesma bebida.
   ii. O jogador preferido do adepto português não é o Ronaldo.
   iii. O adepto que prefere o Neymar bebe sumo ou café.
   iv. O adepto português não bebe a mesma bebida que o adepto que prefere o Messi.
   v. O adepto que prefere o Mbappé bebe a mesma bebida que o adepto argentino.

II) [4,5val] Considere os seguintes atributos e respectivos valores possíveis:

$$x_1 \in \{F, T\} \quad x_2 \in \{A, B, C\}$$

e o seguinte conjunto de 7 exemplos a ser usados na construção de uma árvore de decisão usando o algoritmo DTL.

	x_1	x_2	Classificação
D_1	T	A	+
D_2	T	B	+
D_3	T	C	+
D_4	F	A	−

	x_1	x_2	Classificação
D_5	F	A	−
D_6	F	A	+
D_7	F	C	−

a) Qual o ganho de informação (IG) de cada um dos atributos? Apresente os cálculos.
   Os seguintes valores de entropia poderão ajudar:

x	y	$H\left(\frac{x}{y}, 1-\frac{x}{y}\right)$
1	1	0,00
1	2	1,00
1	3	0,92

x	y	$H\left(\frac{x}{y}, 1-\frac{x}{y}\right)$
1	4	0,81
1	5	0,72
1	6	0,65

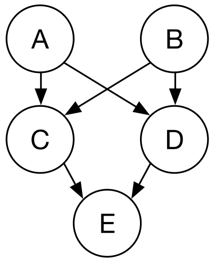
x	y	$H\left(\frac{x}{y}, 1-\frac{x}{y}\right)$
1	7	0,59
1	8	0,54
2	5	0,97

x	y	$H\left(\frac{x}{y}, 1-\frac{x}{y}\right)$
2	7	0,86
3	7	0,99
3	8	0,95

b) Qual o atributo a ser escolhido como raiz da árvore? Justifique.
   Se necessário, desempate a favor do atributo de menor índice (e.g. x_1 vence sobre x_2).
c) Apresente a árvore de decisão induzida pelo algoritmo DTL. Justifique e apresente os cálculos efetuados.
d) O algoritmo DTL pode produzir árvores desnecessariamente grandes. Proponha uma ou várias alterações ao algoritmo que reduzam este problema, sem que afetem a classificação efetuada pela árvore produzida nem tenham um custo computacional substancial. Seja claro, preciso e conciso.

III) [7val] Considere a seguinte Rede de Bayes, onde as variáveis aleatórias A, B, C, D e E são todas booleanas:

P(a)	P(¬a)
0,5	0,5



P(b)	P(¬b)
0,4	0,6

A	B	P(c|A,B)	P(¬c|A,B)
a	b	0,1	0,9
a	¬b	0,5	0,5
¬a	b	0,6	0,4
¬a	¬b	0,2	0,8

A	B	P(d|A,B)	P(¬d|A,B)
a	b	0,4	0,6
a	¬b	0,1	0,9
¬a	b	0,7	0,3
¬a	¬b	0,8	0,2

C	D	P(e|C,D)	P(¬e|C,D)
c	d	0,6	0,4
c	¬d	0,3	0,7
¬c	d	0,5	0,5
¬c	¬d	0,2	0,8

a) Determine a probabilidade do evento $(\neg a, b, c, d, e)$?

b) Para determinar a probabilidade $P(a, b|\neg c, d)$, quais das tabelas de distribuição de probabilidade da Rede de Bayes – **P**(A), **P**(B), **P**(C|A,B), **P**(D|A,B), **P**(E|C,D) – são necessárias?

c) Determine o valor de $P(c|\neg a)$?

d) Qual ou quais das seguintes expressões permitem calcular o valor de $P(\neg c|d, \neg a)$?

   a. $\frac{\sum_B P(\neg c|\neg a,B)P(B)P(d|\neg a,B)P(B)}{P(d|\neg a)}$
   b. $\frac{\sum_B P(\neg c|\neg a,B)P(\neg a)P(d|\neg a,B)P(B)}{P(d)}$
   c. $\frac{\sum_B P(\neg c|\neg a,B)P(B)P(d|\neg a,B)P(B)}{\sum_B P(d|\neg a,B)P(\neg a)}$
   d. $\frac{\sum_B P(\neg c|\neg a,B)P(d|\neg a,B)P(B)}{P(d|\neg a)}$
   e. Nenhuma das anteriores

e) Qual o valor de $P(\neg c|d, \neg a)$?

f) Qual a probabilidade de obtermos a amostra $(\neg a, b, \neg c, d, \neg e)$ usando a amostragem a partir de uma rede vazia (PRIOR-SAMPLE)? [Indique a alínea correta]

   i. $0{,}5 \times 0{,}4 \times 0{,}6 \times 0{,}3 \times 0{,}2$
   ii. $0{,}4 \times 0{,}4 \times 0{,}9 \times 0{,}2 \times 0{,}8$
   iii. $0{,}6 \times 0{,}1 \times 0{,}7 \times 0{,}1 \times 0{,}2$
   iv. $0{,}5 \times 0{,}4 \times 0{,}4 \times 0{,}7 \times 0{,}5$
   v. $0{,}25 \times 0{,}24 \times 0{,}9 \times 0{,}1 \times 0{,}5$
   vi. $0{,}4 \times 0{,}5 \times 0{,}24 \times 0{,}21 \times 0{,}25$
   vii. Nenhuma das anteriores

g) Com o objetivo de estimar a probabilidade $P(\neg c|d, \neg a, e)$, seria a amostra $(\neg a, b, \neg c, d, \neg e)$ descartada pelo algoritmo de amostragem por rejeição (REJECTION-SAMPLING)? [Sim/Não]

h) A amostra $(\neg a, b, \neg c, d, e)$ foi obtida pelo algoritmo de pesagem por verosimilhança (LIKELIHOOD-WEIGHTING), com o objetivo de estimar a probabilidade $P(\neg c|d, \neg a, e)$. Qual é o seu peso? [Indique a alínea correta]

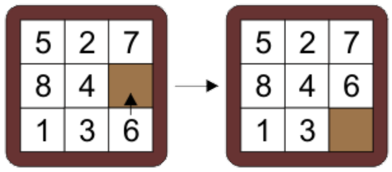
   i. $0{,}5 \times 0{,}5 \times 0{,}5$
   ii. $0{,}4 \times 0{,}9 \times 0{,}6$
   iii. $0{,}4 \times 0{,}24 \times 0{,}6$
   iv. $0{,}5 \times 0{,}7 \times 0{,}5$
   v. $0{,}5 \times 0{,}3 \times 0{,}5$
   vi. $0{,}6 \times 0{,}3 \times 0{,}6$
   vii. Nenhum das anteriores

i) A amostragem usando o algoritmo de pesagem por verosimilhança (LIKELIHOOD-WEIGHTING) sobrestima sistematicamente a probabilidade da variável condicionada num dos seus antecessores? [Indique a alínea correta]

   i. Sim, porque a pesagem por verosimilhança não faz a amostragem de todas as variáveis, criando assim uma tendência (*bias*).
   ii. Sim, mas não pela razão anterior.
   iii. Não, porque a pesagem por verosimilhança é consistente i.e., não introduz qualquer tendência (*bias*).
   iv. Não, mas não pela razão anterior.

j) Com o objetivo de aproximar a probabilidade $P(\neg c|d, \neg a, e)$, obtivemos a seguinte sequência de amostras: $(\neg a, b, \neg c, d, e)$; $(\neg a, b, \neg c, d, e)$; $(\neg a, b, c, d, e)$; $(\neg a, b, c, d, e)$; $(\neg a, b, \neg c, d, e)$.
Poderia esta sequência ter sido obtida através do Markov Chain Monte Carlo? Justifique sucintamente.

IV) [2val] Considere a charada-8 (ver figura) onde cada estado é representado usando os seguintes predicados: **`p(N)`** onde **`N`** é um número inteiro de 1 a 8, indicando as peças existentes; **`pos(X)`** onde **`X`** é um inteiro de 1 a 3 indicando as colunas/linhas existentes; **`at(N,X,Y)`** indicando que a peça **`N`** está localizada na célula com coordenadas **`X,Y`**; e **`b(X,Y)`** indicando que a casa vazia está localizada na célula com coordenadas **`X,Y`**. Assuma que pode usar operações aritméticas e igualdade.

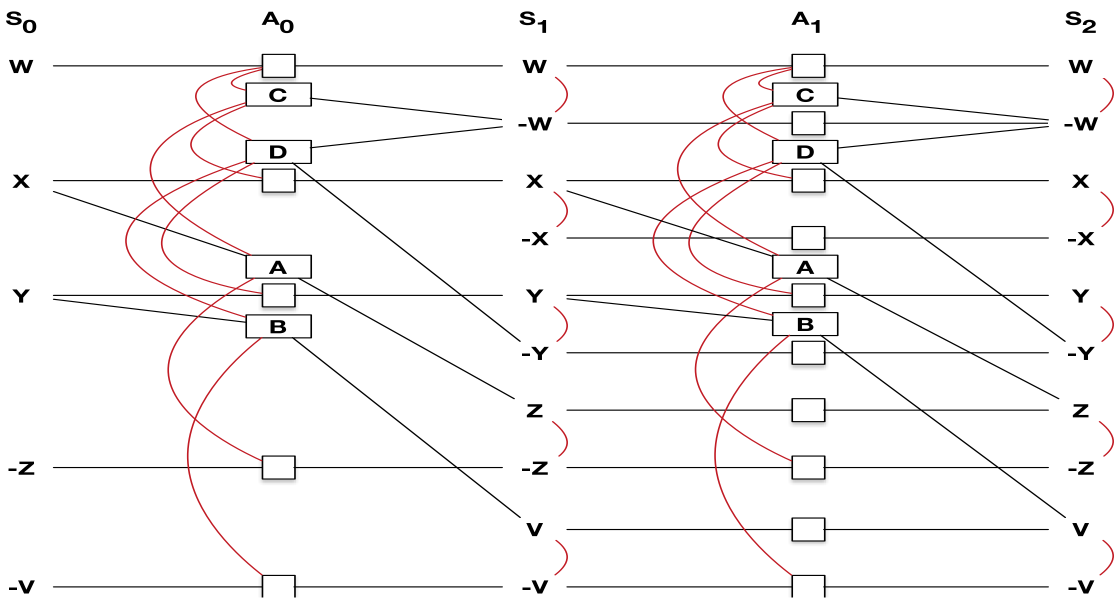


a) Represente o estado correspondente à situação inicial (figura do lado esquerdo).
b) Modele na linguagem PDDL (ou STRIPS) a ação de mover uma peça para cima como ilustrado na figura.

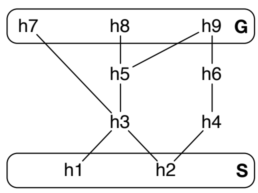
V) [2val] Considere que o predicado **`A(x,y)`** significa que o aluno **`x`** obteve aprovação na unidade curricular **`y`**. Nesta pergunta, pode assumir que o primeiro argumento é um aluno e que o segundo argumento é uma unidade curricular. Para cada uma das seguintes frases, escreva a fórmula na linguagem da lógica de primeira ordem que melhor a representa. [Cada resposta errada acarreta um desconto substancial. A pergunta tem uma cotação mínima de 0 valores.].

i. Existe um aluno que obteve aprovação a pelo menos uma unidade curricular.
ii. Nenhum aluno obteve aprovação a qualquer unidade curricular.
iii. Existe um aluno que não obteve aprovação a qualquer unidade curricular.
iv. Todos os alunos obtiveram aprovação a pelo menos uma unidade curricular.
v. Existe uma unidade curricular à qual nenhum aluno obteve aprovação.
vi. Nenhum aluno obtive aprovação a todas as unidades curriculares.
vii. Todos os alunos obtiveram aprovação a todas as unidades curriculares.
viii. Existe um aluno que obtive aprovação a todas as unidades curriculares.
ix. Existe uma unidade curricular à qual todos os alunos obtiveram aprovação.

VI) [Bónus: até 1val] Considere o seguinte Grafo de Planeamento construído para se encontrar um plano para o objetivo {V, -W, Z}. Indique um plano (linearizado) que pudesse ser obtido pelo algoritmo GRAPHPLAN.



VII) [Bónus: até 1val] Sejam $\boldsymbol{S} = \{h1, h2\}$ e $\boldsymbol{G} = \{h7, h8, h9\}$ a fronteira mais específica e a fronteira mais geral, respetivamente, numa iteração do algoritmo de eliminação de candidatos. A ordenação parcial entre as hipóteses remanescentes é ilustrada na figura à direita. Considere um novo exemplo de treino negativo d consistente com as hipóteses $h1, h3, h5, h6$ e $h7$, sendo inconsistente com as restantes. Indique as novas fronteiras $\boldsymbol{S}$ e $\boldsymbol{G}$ após o tratamento do exemplo d pelo algoritmo de eliminação de candidatos.



10/Dez/2022 – 9:00 **Inteligência Artificial – Folha de Respostas** **DI/FCT/UNL, 2º Teste, 2022/23, Duração: 2h** **Modelo:** A e B

I.a) Resposta: 3317760000 modelos estáveis.
Justificação: 3317760000 = 4^4. 4^4. 15^4 dado que há 4^4=256 combinações de nacionalidades diferentes; 4^4=256 combinações de jogadores preferidos diferentes; e 15^4 = 50625 combinações de bebidas diferentes sendo $15 = \sum_{i=1}^{4} {}^4C_i = 4 + 6 + 4 + 1$ onde ${}^nC_i$ é o número de combinações de n, i a i elementos.

I.b)
```
1{preferido(T,J):adepto(T)}1 :- jogador(J).
        Ou
:- preferido(T1,J), preferido(T2,J), T1!=T2.
```

I.c) i)
```
:- bebida(X), 3{bebe(T,X)}.
        Ou
{bebe(T,B):adepto(A)}2 :- bebida(B).
        Ou
:- bebe(A1,X), bebe(A2,X), bebe(A3,X), A1!=A2, A2!=A3, A1!=A3.
```

ii)
```
:- nacionalidade(T,portugal), preferido(T,ronaldo).
        Ou
ok2 :- nacionalidade(T,portugal), not preferido(T,ronaldo).
:- not ok2
```

iii)
```
1{bebe(T,sumo), bebe(T,café)}:- preferido(T,neymar).
        Ou
ok3 :- bebe(T,cafe), preferido(T,neymar).
ok3 :- bebe(T,sumo), preferido(T,neymar).
:- not ok3.
        Ou
:- preferido(T,neymar), not bebe(T,cafe), not bebe(T,sumo).
```

iv)
```
:- nacionalidade(X,portugal), bebe(X,B), preferido(Y,messi), bebe(Y,B).
```

v)
```
ok5 :- preferido(X,mbappé), bebe(X,B), nacionalidade(Y,argentina), bebe(Y,B).
:- not ok5.
```

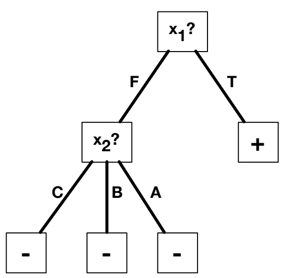
II.a) $IG(x_1) = 0{,}527$ $IG(x_2) = 0{,}133$
Cálculos:

$$IG(x_1) = H\left(\frac{4}{7},\frac{3}{7}\right) - \left[\frac{3}{7}\cdot H(1{,}0) + \frac{4}{7}\cdot H\left(\frac{3}{4},\frac{1}{4}\right)\right] = 0{,}99 - \left(0 + \frac{4}{7}\cdot 0.81\right) = 0.527$$

$$IG(x_2) = H\left(\frac{4}{7},\frac{3}{7}\right) - \left[\frac{4}{7}\cdot H\left(\frac{1}{2},\frac{1}{2}\right) + \frac{1}{7}\cdot H(1{,}0) + \frac{2}{7}\cdot H\left(\frac{1}{2},\frac{1}{2}\right)\right] = 0{,}99 - \left(\frac{4}{7} + 0 + \frac{2}{7}\right) = 0.133$$

II.b) Atributo: x_1
Justificação: Porque tem o maior ganho de informação.

II.c) Árvore:



Justificação/cálculos:
No ramo x_1=T, bem como no ramo correspondente ao valor C de x_2, todas as instâncias têm o mesmo valor (+ no primeiro caso e – no segundo), justificando assim o valor das respetivas folhas. No ramo correspondente ao valor B de x_2, como não há instâncias, adota-se o valor da moda do pai, i.e. o valor negativo. Por último, no ramo correspondente ao valor A de x_2, como nem todas as instâncias têm o mesmo valor – e não há mais variáveis para as discriminar (significa que estes exemplos não foram gerados por uma função) – adotamos o valor da moda, i.e., o valor negativo.

IId)
O algoritmo seleciona atributos com ganho de informação nulo (não ilustrado por este exemplo). O algoritmo poderia ser melhorado impedindo desde logo a seleção de atributos com ganho de informação igual a zero.
Mesmo com a melhoria anterior, o algoritmo constrói árvores onde podem existir subárvores nas quais todas as folhas têm o mesmo resultado, logo sem influência na classificação feita pela árvore (e.g. a subárvore de x_1=F é desnecessária pois basta o teste x_1=F para determinar a classificação final). O algoritmo poderia ser melhorado introduzindo um passo final que recursivamente eliminasse subárvores cujas folhas tivessem todas a mesma classificação.

III.a) **Resposta**: 0,0504

Cálculos:

$$P(\neg a, b, c, d, e) = P(\neg a).P(b).P(c|\neg a, b).P(d|\neg a, b).P(e|c, d) = 0{,}5.0{,}4.0{,}6.0{,}7.0{,}6 = 0{,}0504$$

b) [Sim/Não] **P**(A): **Sim** **P**(B): **Sim** **P**(C|A,B): **Sim** **P**(D|A,B): **Sim** **P**(E|C,D): **Não**

c) **Resposta:** 0,3600

Cálculos:

Versão mais simples:

$$P(c|\neg a) = \sum_B P(c|\neg a, B)P(B) = 0{,}6.0{,}4 + 0{,}2.0{,}6 = 0{,}3600$$

Versão mais trabalhosa:

$$P(c|\neg a) = \frac{P(c, \neg a)}{P(\neg a)} = \frac{\sum_B P(c, \neg a, B)}{P(\neg a)} = \frac{\sum_B P(\neg a)P(c|\neg a, B)P(B)}{P(\neg a)} = \sum_B P(c|\neg a, B)P(B) = 0{,}3600$$

Versão ainda mais trabalhosa:

$$P(c|\neg a) = \alpha P(c, \neg a) = \alpha \sum_B P(c, \neg a, B)$$

$$\sum_B \mathbf{P}(\mathbf{C}, \neg a, B) = \sum_B P(\neg a)\mathbf{P}(\mathbf{C}|\neg a, B)P(B) = P(\neg a)\sum_B \mathbf{P}(\mathbf{C}|\neg a, B)P(B)$$

$$\mathbf{C} = true: \quad P(\neg a)\sum_B P(c|\neg a, B)P(B) = 0{,}5(0{,}6.0{,}4 + 0{,}2.0{,}6) = 0{,}18$$

$$\mathbf{C} = false: \quad P(\neg a)\sum_B P(\neg c|\neg a, B)P(B) = 0{,}5(0{,}4.0{,}4 + 0{,}8.0{,}6) = 0{,}32$$

$$P(c|\neg a) = \alpha P(c, \neg a) = \frac{0{,}18}{(0{,}18 + 0{,}32)} = 0{,}3600$$

Versão desnecessariamente trabalhosa (não reproduzida aqui): marginalizar também para as variáveis D e E, que, no entanto, acabariam por ser eliminadas, por simplificação, por serem irrelevantes, como poderia ser determinado logo à partida dado não fazerem parte do conjunto $Ancestors(\{C\} \cup \{A\}) = \{A, B, C\}$.

d) a. ☐ b. ☐ c. ☐ d. ☒ e. ☐

e) 0,6526

f)	g)	h)	i)
iv	Sim	iv	iii

j) **Resposta**: Sim

Justificação: No MCMC, cada amostra é obtida a partir da anterior, escolhendo uma das variáveis que não são evidência na consulta (no caso, escolhendo **b** ou **c**) e determinando o seu novo valor, amostrando-a dado o valor das restantes variáveis na amostra anterior. Assim, qualquer amostra pode variar, no máximo, no valor de uma das variáveis que não as de evidência, o que é o caso da sequência apresentada. Mais, o novo valor da variável escolhida apenas pode tomar um valor para o qual a sua probabilidade condicionada nas restantes seja diferente de 0, o que acontece em cada uma das amostras.

IV.a)

```
p(1), p(2), p(3), p(4), p(5), p(6), p(7), p(8),
pos(1), pos(2), pos(3),
at(1,1,1),at(3,2,1),at(6,3,1),at(8,1,2),at(4,2,2),
at(5,1,3),at(2,2,3),at(7,3,3), b(3,2).
```

IV.b)

```
Acção: mover_para_cima(N,X,Y)
Precondições: p(N), pos(X), pos(Y), pos(Y+1), b(X,Y+1), at(N,X,Y)
Efeitos: ~b(X,Y+1), ~at(N,X,Y), b(X,Y), at(N,X,Y+1)
```

V. i. $\exists x \exists y\, A(x, y)$	**iv.** $\forall x \exists y\, A(x, y)$	**vii.** $\forall x \forall y\, A(x, y)$
ii. $\neg\exists x \exists y\, A(x, y)$	**v.** $\exists y \neg\exists x\, A(x, y)$	**viii.** $\exists x \forall y\, A(x, y)$
iii. $\exists x \neg\exists y\, A(x, y)$	**vi.** $\neg\exists x \forall y\, A(x, y)$	**ix.** $\exists y \forall x\, A(x, y)$

VI	VII
[A, C, B] (há outras alternativas)	**S = {h_1}** **G = {h_5, h_7}**