15/1/36

MAC0425 Inteligência Artificial

Segunda Avaliação (P2)

nome: [redacted] nº USP: [redacted]

assinatura: GABARITO

INSTRUÇÕES:

- *Leia atentamente os enunciados.*
- *Este exame é composto por questões dissertativas; todas as respostas devem ser justificadas.*
- *Escreva suas resposta nos locais indicados, utilizando caneta esferográfica azul ou preta.*
- *Use os versos das folhas como rascunho; escreva sua resposta apenas quando tiver certeza.*
- *Você pode consultar apenas uma folha A4 (frente e verso) individual.*
- *O uso de equipamentos eletrônicos (calculadoras, celulares, computadores) não é permitido.*
- ***Duração:** 120 minutos.*

Questão 1. Considere um agente que se situa num corredor dividido em 3 regiões: | 1 | 2 | 3 |
Denote a variável aleatória representando a posição (região) do agente no instante t por X_t, a ação tomada no instante t por A_t, a observação do sensor esquerdo no mesmo instante por E_t (1 significa presença de parede, 0 ausência de parede) e por D_t a observação do sensor direito. Considerando que o agente inicialmente ($t = 0$) desconhece sua posição e atribui uma distribuição de probabilidades uniforme para X_0, qual o valor de $\Pr(X_0 = 1 | E_1 = 0, D_1 = 1, A_0 = \rightarrow)$?

Ações determinísticas

☐z ☐p ☐t

$$P(X_0 = 1 | E_1 = 0, D_1 = 1, A_0 = \rightarrow) =$$

$$\frac{P(E_1 = 0, D_1 = 1 | X_0 = 1, A_0 = \rightarrow)\cancel{P(X_0 = 1)}}{\sum_{x_0} P(E_1 = 0, D_1 = 1 | X_0 = x_0, A_0 = \rightarrow)\cancel{P(X_0 = x_0)}}$$

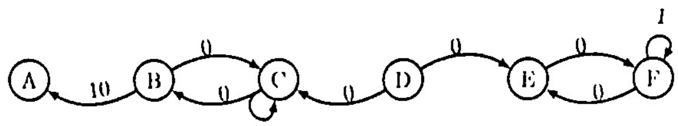
$$= \frac{\sum_{x_1} P(E_1 = 0, D_1 = 1 | X_1 = x_1) P(X_1 = x_1 | X_0 = 1, A_0 = \rightarrow)\cancel{P(X_0 = 1)}}{\sum_{x_0} \cancel{P(X_0 = x_0)} \sum_{x_1} P(E_1 = 0, D_1 = 1 | X_1 = x_1) P(X_1 = x_1 | X_0 = x_0, A_0 = \rightarrow)}$$

$$= \frac{\sum_{x_1} P(E_1 = 0, D_1 = 1 | X_1 = x_1) P(X_1 = x_1 | X_0 = 1, A_0 = \rightarrow)}{\sum_{x_0} \sum_{x_1} P(\underbrace{E_1 = 0, D_1 = 1}_{\text{sensor}} | X_1 = x_1) P(X_1 = x_1 | X_0 = x_0, A_0 = \rightarrow)}$$

$$= \frac{P(E_1 = 0, D_1 = 1 | X_1 = 2)}{P(\text{sensor} | X_1 = 2) + P(\text{sensor} | X_1 = 3) + P(\text{sensor} | X_1 = 3)}$$

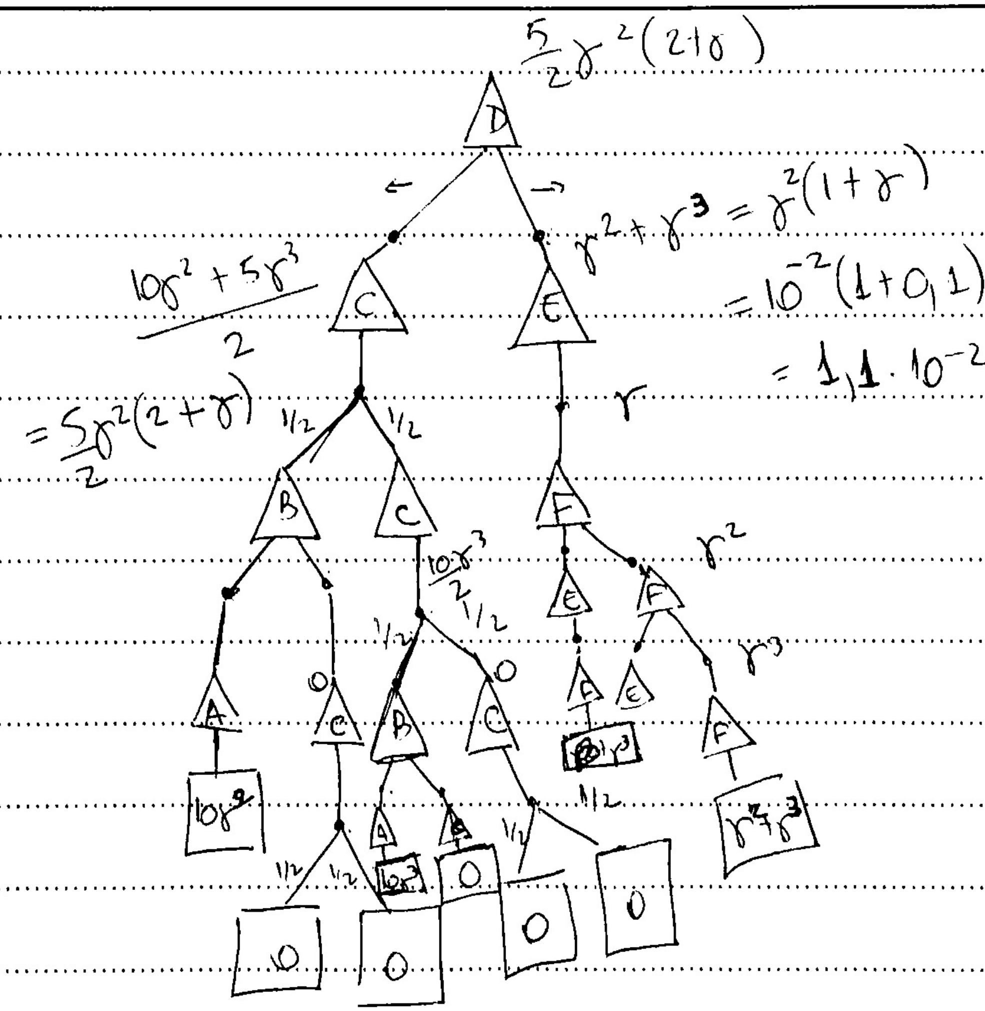
$$= \left(1 + 2\frac{P(\text{sensor} | X = 3)}{P(\text{sensor} | X = 2)}\right)^{-1}$$

Questão 2. O diagrama abaixo representa um processo de decisão Markoviano (MDP). Os valores associados às transições indicam a recompensa obtida ao transitar para o próximo estado (ex.: $R(B, \leftarrow) = 10$). A ação de transitar de C para B "falha" com probabilidade 1/2; as demais ações são determinísticas.



Desenhe a árvore de busca EXPECTIMAX a partir do estado D com horizonte 4 e fator de desconto $\gamma = 0,1$. Indique os valores de cada nó.

☐z ☐p ☐t



Questão 3. Considere o MDP da questão anterior. Encontre o valor do fator de desconto γ para o qual $Q^*(D, \leftarrow) = Q^*(D, \rightarrow)$, ou seja, para o qual qualquer ação em D é ótima.

☐z ☐p ☐t

$$Q^*(D,\leftarrow) = 0 + \gamma Q^*(C,\leftarrow) = 0 + \gamma\underbrace{[0 + \gamma[\tfrac{1}{2}Q^*(B,\leftarrow) + \tfrac{1}{2}Q^*(C,\leftarrow)]]}_{= Q^*(C,\leftarrow)}$$

$$= 0 + 0 + \frac{\gamma^2 \cdot 10}{2} + \frac{\gamma}{2} Q^*(C,\leftarrow)$$

$$\Rightarrow \cancel{\gamma} Q^*(C,\leftarrow) = \frac{10}{2}\gamma^2 + \frac{\gamma^{\cancel{2}}}{2} Q^*(C,\leftarrow)$$

$$\Rightarrow Q^*(C,\leftarrow) = \frac{10\gamma}{2-\gamma} \Rightarrow Q^*(D,\leftarrow) = \frac{10\gamma^2}{2-\gamma}$$

$$Q^*(D,\rightarrow) = 0 + \gamma[0 + \gamma Q^*(F,\rightarrow)]$$

$$Q^*(F,\rightarrow) = 1 + \gamma Q^*(F,\rightarrow) \Rightarrow Q^*(F,\rightarrow) = 1/(1-\gamma)$$

$$\Rightarrow Q^*(D,\rightarrow) = \frac{\gamma^2}{1-\gamma}$$

$$\Rightarrow Q^*(D,\leftarrow) = Q^*(D,\rightarrow) \Leftrightarrow \frac{10\cancel{\gamma^2}}{2-\gamma} = \frac{\cancel{\gamma^2}}{1-\gamma}$$

$$\Leftrightarrow 10 - 10\gamma = 2 - \gamma \Leftrightarrow 9\gamma = 8 \Leftrightarrow \gamma = 8/9$$

Questão 4. Considere ainda o MDP da Questão 2, e assuma fator de desconto $\gamma = 0,2$. Compute as primeiras 2 iterações do algoritmo de iteração de valor **assíncrono**, atualizando a função valor na ordem **lexicográfica**.

☐z ☐p ☐t

$V^0 = 0 \quad V(s) = \max_a R(s,a) + \gamma \sum_{s'} T(s,a,s')V(s')$

$V^1(A) = 0 \quad V^1(B) = 10 + 0 = 10$

$V^1(C) = 0 + 1/2 \cdot 0 \cdot 2/10 + 1/2 \cdot 10 \cdot 2/10 = 1$

$V^1(D) = 2/10 \cdot 1 = 2/10$

$V^1(E) = 0 \qquad V^1(F) = 1$

$V^2(A) = 0 \quad V^2(B) = 10$

$V^2(C) = 1/2 \cdot 2/10 \cdot 1 + 1/2 \cdot 2/10 \cdot 10 = 11/10 = 1,1$

$V^2(D) = 2/10 \cdot 11/10 = 22/100 = 0,22$

$V^2(E) = 2/10 \cdot 1 = 2/10 = 0,2$

$V^2(F) = 1 + \frac{2}{10} \cdot 1 = 1,2$

Responda verdadeiro ou falso, e justifique sua resposta.

Questão 5. Em um processo de decisão Markoviano com fator de desconto $\gamma = 1$, a política ótima não se altera se adicionarmos uma constante a todas as recompensas, ou seja, se modificarmos $R(s, a)$ para $R(s, a) + k$ para todo s e a, para $k \neq 0$.

☐z ☐p ☐t

Falso. Processo determinístico

Para $\gamma = 1$, só existe política ótima se houver estado terminal

$$\begin{tikzcd} & B \arrow[dr, "c+k"] & \\ A \arrow[ur, "c+k"] \arrow[rr, bend right, "C+k"] & & M \end{tikzcd}$$

tal que $C \not> 2c$ mas $C+k < 2c+2k$

Assumindo estado terminal.

Questão 6. Em um processo de decisão Markoviano com fator de desconto $\gamma < 1$, a política ótima não se altera se adicionarmos uma constante a todas as recompensas.

☐z ☐p ☐t

$$V^*(s_0) = E\left[\sum_{t=0}^{\infty} \gamma^t R_t\right]$$

$$R'_t = R_t + k \rightarrow V^*(s_0) = E\left[\sum_{t=0}^{\infty} \gamma^t (R_t + k)\right]$$

$$= E\left[\sum_{t=0}^{\infty} \gamma^t R_t\right] + k \underbrace{\sum_{t=0}^{\infty} \gamma^t}_{= \frac{1}{1-\gamma}}$$

Verdadeiro se não houver estado terminal.