Nota: 5/10

+50/1/6+

MAC0425 Inteligência Artificial — Segunda Avaliação (P2)

NOME: Vinicius Moreno da Silva — NUSP: 10297776

ASSINATURA: Vinicius Moreno

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 reespota 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 $\Pr(X_0 = 1|E_t = 0, D_t = 1, A_0 = \rightarrow)$?

0/2

☐z ☐p ☐t

$Pr(X_0=1) = 1/3$; $P(X_0=2) = 1/3$; $P(X_0=3) = \frac{1}{3}$

$Pr(E_t=0) = \frac{\overbrace{\{2,3\}}^{\text{espaço possível}}}{\underbrace{\{1,2,3\}}_{\text{espaço total}}} = \frac{2}{3}$ X Sensor não é perfeito

$Pr(D_t=1) = \frac{\{3\}}{\{1,2,3\}} = \frac{1}{3}$

Segundo as propriedades de Markov, ~~[illegible]~~ $Pr(X_0)$ é independente em relação à $Pr(E_t)$ e $P(D_t)$. Portanto,
Isso não é verdade, e não tem a ver com a propriedade de Markov

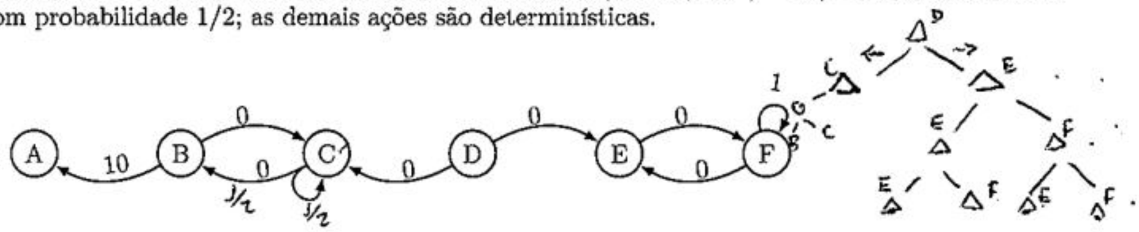
$Pr(X_0=1 | E_t=0, D_t=0, A_0=\rightarrow) = Pr(X_0=1 | A_0=\rightarrow)$

Como no instante $t=0$, $A_0 = \rightarrow$, logo X_0 é uma ~~[illegible]~~ região do corredor em que esta ação é possível, portanto não pode $X_0 = 3$.

Visto que só restam outras duas regiões e a distribuição de probabilidade para X_0 é uniforme, $Pr(X_0=1|A_0=\rightarrow) = \frac{\{1\}}{\{1,2\}} = \frac{1}{2}$

X

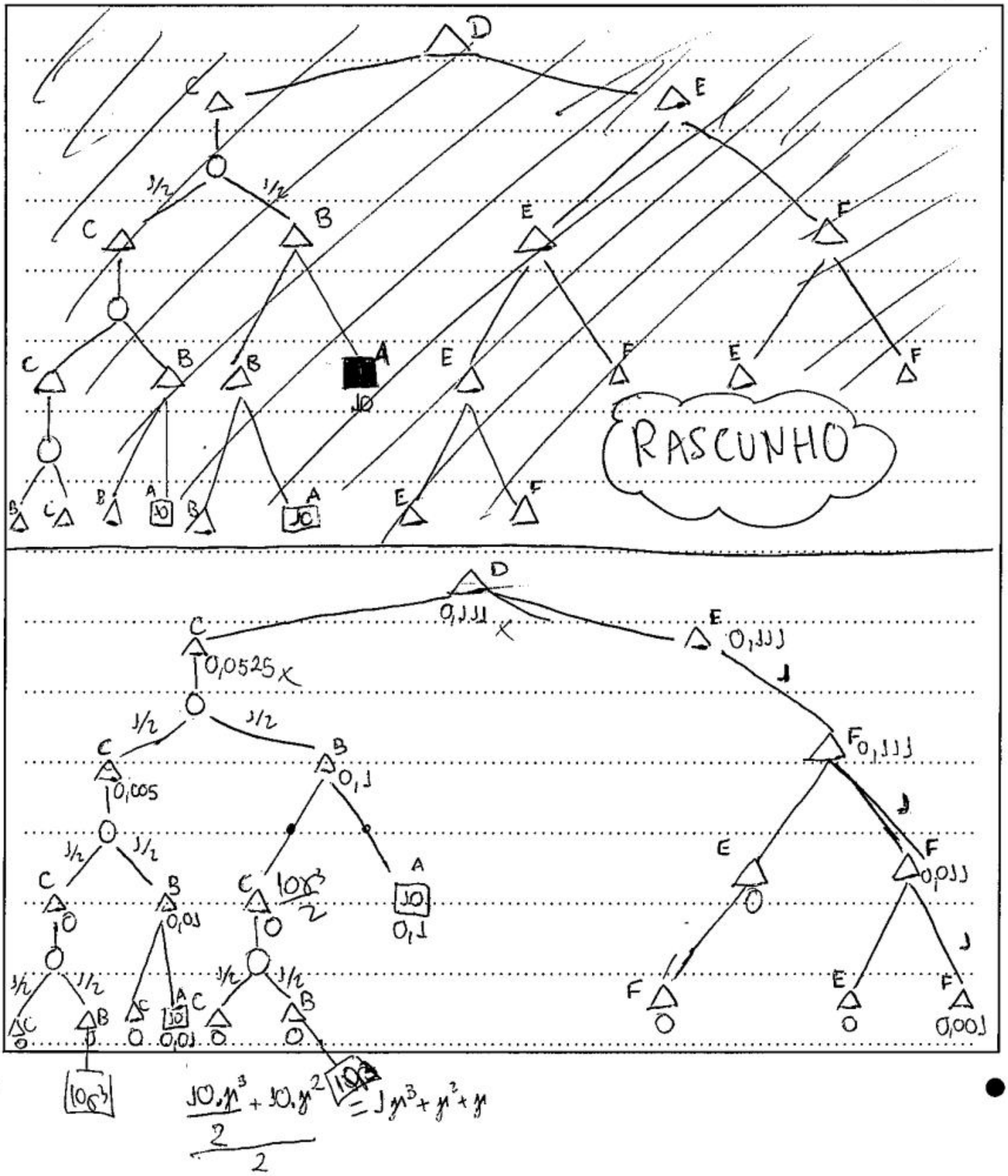
Questão 2. O diagrama abaixo representa um processo de decisão Markoviano (MDP). Os valores associados às transições indicam a recompensa obtida ao transitar para o próximo estado (ex.: $R(B, \leftarrow) = 10$). A ação de transitar de C para B "falha" com probabilidade 1/2; as demais ações são determinísticas.



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.

1/2

☐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.

0/2

z p t

$$Q^*(D, \leftarrow) = \frac{\frac{10\gamma^3}{2} + 10\gamma^2}{2} = \frac{10\gamma^3}{4} + \frac{10\gamma^2}{2}$$

recompensa total de cada lado da árvore

$$Q^*(D, \rightarrow) = \gamma^3 + \gamma^2 + \gamma + \gamma^4 + \gamma^5 + \dots$$

implica → horizonte infinito

$$Q^*(D, \leftarrow) = Q^*(D, \rightarrow) \Leftrightarrow \frac{10\gamma^3}{4} + \frac{10\gamma^2}{2} = \gamma^3 + \gamma^2 + \gamma + \gamma^4 + \dots \Rightarrow$$

$$\Rightarrow \frac{6\gamma^3}{4} + 4\gamma^2 - \gamma = 0 \Rightarrow \gamma\left(\frac{3\gamma^2}{2} + 4\gamma - 1\right) = 0$$

Aplicando baskhera:

$$\frac{-4 \pm \sqrt{16 + 6}}{3}$$

$$\gamma' = \frac{\sqrt{22} - 4}{3}$$

$$\gamma'' = \frac{-\sqrt{22} - 4}{3}$$

$$\gamma''' = 0$$ → precisa ser maior que 0

$$\gamma = \frac{\sqrt{22} - 4}{3}$$

$$\frac{-4 \pm \sqrt{-4}}{3}$$ $16 - 4ac$

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**.

2 / 2

☐ z ☐ p ■ t

$V_0 = \{0; 0; 0; 0; 0; 0\}$

$V(s) = \max_a \left[R(s,a) + \gamma \sum T(s,a,s') V(s') \right]$

$V_1(A) = 0$

$V_1(B) = 10 + 0,\ 0 + 0 = 10$

$V_1(C) = 0 + \gamma\left(\frac{1}{2} \cdot V(B) + \frac{1}{2} V(C)\right) = 5\gamma = 1$

$V_1(D) = 0 + \gamma V(C),\ 0 + \cancel{\gamma V(E)}^{\,0} = 1 \cdot \gamma = 0,2$

$V_1(E) = 0 + \gamma V(F) = 0$

$V_1(F) = 0 + \gamma V(E),\ 1 + \gamma V(F) = 1$

$V_1 = \{0; 10; 1; 0,2; 0; 1\}$

$V_2(A) = 0$

$V_2(B) = 10 + 0,\ 0 + \gamma \overbrace{V(C)}^{0.2} = 10$

$V_2(C) = 0 + \gamma\left(\frac{1}{2} V(B) + \frac{1}{2} \cdot V(C)\right) = \gamma(5 + 0,5) = 1,1$

$V_2(D) = 0 + \gamma V(C),\ 0 + \gamma V(E)^{\to 0} = 1,1 \cdot \gamma = 0,22$

$V_2(E) = 0 + \gamma V(F) = \gamma = 0,2$

$V_2(F) = 0 + \gamma V(E),\ 1 + \gamma V(F) = 0,04,\ 1 + 0,2 = 1 + 0,2 = 1,2$

$V_2 = \{0; 10; 1,1; 0,22; 0,2; 1,2\}$

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 $k \neq 0$.

1/1

□z □p ■t

Falso. Prova por contra-exemplo: (mesmo molde da questão 2)

Assumindo o MDP a seguir $\rightarrow$ $A \xleftarrow{2} B \xrightarrow{0} C \xrightarrow{0} D \xrightarrow{1} E$,

a política ótima a partir de B seria $T(B,\leftarrow)$, com recompensa total sendo $R(B,\leftarrow) = 2$. No entanto, somando-se $K=2$ em todas recompensas das transições desse MDP, ele fica assim:

$A \xleftarrow{2+2=4} B \xrightarrow{0+2=2} C \xrightarrow{0+2=2} D \xrightarrow{1+2=3} E$, portanto, a política ótima deverá ser trocada, pois, a partir de B, $R(B,\leftarrow) = 4 < R(B,\rightarrow) + R(C,\rightarrow) + R(D,\rightarrow) = 7$

0,81
0,9
0,729

3,71
0,729
4,439

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.

1/1

□z □p ■t

Falso. Prova por contra-exemplo:

Assumindo $\gamma = 0.9$ e a seguinte MDP: $A \xleftarrow{2} B \xrightarrow{1} C \xrightarrow{0} D \xrightarrow{0} E \xrightarrow{0} F$, a política ótima seria, a partir de B, $T(B,\leftarrow)$, com recompensa $R(B,\leftarrow) = 2$. Somando-se $K=1$ nas recompensas de todas transições desse MDP, ele ficará assim:

$A \xleftarrow{2+1=3} B \xrightarrow{1+1=2} C \xrightarrow{0+1\gamma=0,9} D \xrightarrow{0+1\gamma^2=0,81} E \xrightarrow{0+1\gamma^3=0,729} F$, portanto, a política ótima deverá ser alterada, pois, a partir de B, $R(B,\leftarrow) = 3 < R(B,\rightarrow) + R(C,\rightarrow) + R(D,\rightarrow) + R(E,\rightarrow) = 4,439$