

Capítulo 3 Exercícios - Sutton

Paulo Henrique Albuquerque

2023-04-25

Exercício 3.1

Senhoras Courier: Um coletivo de mulheres

Exercício 3.2

Exercício 3.3

Exercício 3.4

Exercício 3.5

Exercício 3.6

Exercício 3.7 Não tem

Exercício 3.8 Referindo ao diagrama 3.4b, obtemos,

$$q_{\pi}(s, a) = \sum_{s', r} p(s', r | s, a) [r + \gamma \sum_{a' \in \mathcal{A}(s')} \pi(a' | s') q_{\pi}(s', a')].$$

A sequência de passos é semelhante àquela da seção da Demonstração da equação de Bellman do arquivo notex.tex. A diferença é que não condicionamos pelas ação tomadas, por isso a diferença no formato da equação.

Exercício 3.9

$$v_{meio} = \frac{1}{4} 0.9 (2.3 + 0.4 - 0.4 + 0.7) = 0.675.$$

Exercício 3.10 Dado um histórico $S_0 A_0 R_0 S_1 A_1 \dots$:

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$
$$\overline{G}_t = \sum_{k=0}^{\infty} \gamma^k (R_{t+k+1} + c) = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} + \sum_{k=0}^{\infty} c \gamma^k = G_t + \frac{c}{1-\gamma}.$$

Logo, $v_c = \frac{c}{1-\gamma}$.

Exercício 3.11 De forma semelhante ao que foi feito na questão acima, obtemos

$$\overline{G}_t = G_t + \frac{c}{1-\gamma} (1 - \gamma^{T-2t-1}).$$

Então, a adição de uma constante às recompensas afeta o problema. Em particular, uma política ruim que faz o agente perder uma partida de xadrez mas demora muito gera recompensas possivelmente maiores do que uma política que ganha rapidamente uma partida.

Exercício 3.12 Considerando o diagrama abaixo,

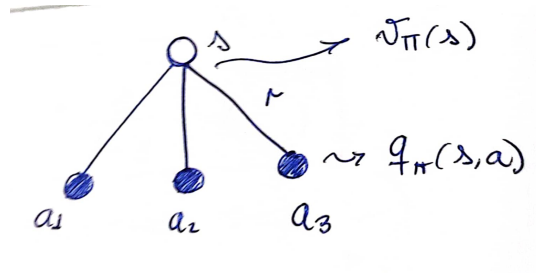


Figure 1: Diagrama backup

obtemos,

$$v_{\pi}(s) = \sum_{a \in \mathcal{A}(s)} \pi(a|s) q_{\pi}(s, a) = \mathbb{E}_{\pi}[q_{\pi}(s, S_{t+1}) | S_t = s].$$

Exercício 3.13

$$q_{\pi}(s, a) = \mathbb{E}[R_{t+1} | S_t = s, A_t = a] + \gamma \mathbb{E}_{\pi}[v_{\pi}(S_{t+1}) | S_t = s, A_t = a] = \sum_{s', r} p(s', r | s, a) [r + \gamma v_{\pi}(s')].$$

Exercício 3.18 Considerando o diagrama abaixo,

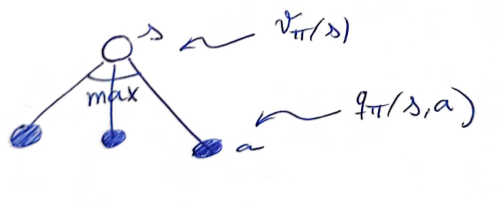


Figure 2: Diagrama backup

obtemos,

$$v_{\star}(s) = \max_{a \in \mathcal{A}(s)} \{q_{\star}(s, a)\}.$$

Exercício 3.19

$$q_{\star}(s, a) = \sum_{s', r} p(s', r | s, a) [r + \gamma v_{\star}(s')].$$

Exercício 3.18