.

CEFET/RJ
Inteligência Artificial (2017.2)
Professor: Eduardo Bezerra (ebezerra@cefet-rj.br)
Lista de exercícios 03

Créditos: essa lista de exercícios contém a tradução dos exercícios disponibilizados na disciplina CS188 - Artificial Intelligence, da Universidade de Berkeley. Os exercícios originais podem ser encontrados em `http://ai.berkeley.edu/section_handouts.html`.

1. No jogo denominado micro-blackjack, você repetidamente tira uma carta (com substituição) que é igualmente provável que seja um 2, 3 ou 4. Você pode ou Comprar ou Parar se a pontuação total dos cartões já comprados for menor do que 6. Caso contrário, você deve parar. Quando você parar, sua utilidade é igual a sua pontuação total (até 5), ou zero se você obter um total de 6 ou superior. Cada vez que você Compra, você não recebe utilidade alguma. Não há desconto (i.e., $\gamma = 1$).
   (a) Quais são os estados e as ações para este MDP?
   (b) Quais são a função de transição e a função de recompensa para este MDP?
   (c) Dê a política ideal para este MDP.
   (d) Qual é o menor número de rodadas (k) de iteração de valor para as quais este MDP terá seus valores exatos (se você acha que a iteração de valor nunca irá convergir exatamente, declare isso).

2. O Pacman está preso no seguinte labirinto 2 por 2 com um fantasma com fome!

   Quando é sua vez de se mover, o Pacman deve mover um passo horizontal ou verticalmente para um quadrado vizinho. Quando é a vez do fantasma, ele também deve mover um passo horizontalmente ou verticalmente. O fantasma e o Pacman alternam movimentos. Depois de cada movimento (pelo Fantasma ou pelo Pacman) se o Pacman e o fantasma ocupam o mesmo quadrado, o Pacman é comido e recebe a utilidade -100. Caso contrário, ele recebe uma utilidade igual a 1. O fantasma tenta minimizar a utilidade que o Pacman recebe. Suponha que o fantasma faz o primeiro movimento. Por exemplo, com um fator de desconto de $\gamma = 1{,}0$, se o fantasma se move para baixo, então o Pacman se move para a esquerda, ganha uma recompensa de 1 após o movimento do fantasma e -100 após a sua movimentação, com uma utilidade total de -99. Note que não há garantias de que jogo termine.

(a) Considere um fator de desconto $\gamma = 0{,}5$, onde o fator de desconto é aplicado uma vez a cada vez que o Pacman ou o fantasma se move. Qual é o valor minimax do jogo truncado após 2 movimentos do fantasma e 2 movimentos do Pacman (Dica: você não precisa construir a árvore minimax)

(b) Considere um fator de desconto $\gamma = 0{,}5$. Qual é o valor minimax do jogo completo (infinito)? (Dica: Você não precisa construir a árvore minimax)

(c) Por que o algoritmo `Iteração de Valor` (*Value Iteration*) é superior ao minimax para resolver este jogo?

(d) Este jogo é semelhante a um MDP porque as recompensas são obtidas em cada passo de tempo. No entanto, é também um jogo adversarial envolvendo decisões de dois agentes.

Seja s o estado (por exemplo, a posição de Pacman e do fantasma), e considere que $A_P(s)$ seja o espaço de ações disponível para o Pacman no estado s (e da mesma forma considere que $A_G(s)$ é o espaço de ações disponíveis para o fantasma). Seja $N(s, a) = s'$ a função sucessora (dado um estado inicial s, esta função retorna o estado s' que resulta depois de tomar a ação a). Finalmente, considere que $R(s)$ denotar a utilidade recebida depois de passar para o estado s.

Escreva uma expressão para $P^*(s)$, o valor do jogo para Pacman em função do estado corrente s (análogo às equações de Bellman). Use um fator de desconto de $\gamma = 1{,}0$. Dica: sua resposta deve incluir $P^*(s)$ no lado direito.

3. Considere um PDM sem descontos e composto por três estados, $(1, 2, 3)$, cujas recompensas são -1, -2 e 0, respectivamente. O estado 3 é o único estado terminal. Nos estados 1 e 2 há duas ações possíveis: a e b. O modelo de transição é o seguinte:

- No estado 1, a ação a move o agente para o estado 2 com probabilidade 0,8 e faz com que o agente fique parado com probabilidade 0,2.
- No estado 2, a ação a move o agente para o estado 1 com probabilidade 0,8 e faz com que o agente fique parado com probabilidade 0,2.
- Tanto no estado 1 quanto no estado 2, ação b move o agente para o estado 3 com probabilidade 0,1 e faz com que o agente fique parado com probabilidade 0,9.

Lembre-se de que o valor ótimo para um estado s, $V^*(s)$, é obtido pela equação de Bellman:

$$V^*(s) = \max_a \sum_{s'} T(s,a,s') \left(R(s, a, s') + \gamma\, V^*(s')\right)$$

Lembre-se também de que essa equação é adaptada pelo algoritmo `Iteração de Valor` como uma equação de atualização para computar $V^*(s)$:

$$V_{k+1}(s) \leftarrow \max_a \sum_{s'} T(s,a,s') \left(R(s, a, s') + \gamma\, V_k(s')\right)$$

(a) O que pode ser determinado qualitativamente sobre a política ótima nos estados 1 e 2? Ou seja, que ações um agente racional deve tomar nesses estados? Justifique sua resposta.

(b) Aplique o algoritmo `Iteração de Valor` para determinar as utilidades dos estados 1 e 2 na primeira iteração, isto é, determine $V_1(1)$ e $V_1(2)$. Lembre-se de que a iniciação desse algoritmo é fazer com que $V_0(s) = 0$, $\forall s$.