[CSED342] 2023-Spring Artificial Intelligence Assignment 4

20210716 최채현 chaehyeonchoi@postech.ac.kr

Problem 1: MDP Warm-up

Consider an MDP problem. There are four states $\{S_A, S_B, S_C, S_D\}$, at each of which two actions $\{+, -\}$ are available, and the state transition and reward have no randomness. All the (action, reward) pairs are described in Figure 1. Assume all the episodes have length 3 (e.g. $S_A \xrightarrow{+} S_B \xrightarrow{-} S_A \xrightarrow{-} S_A$).

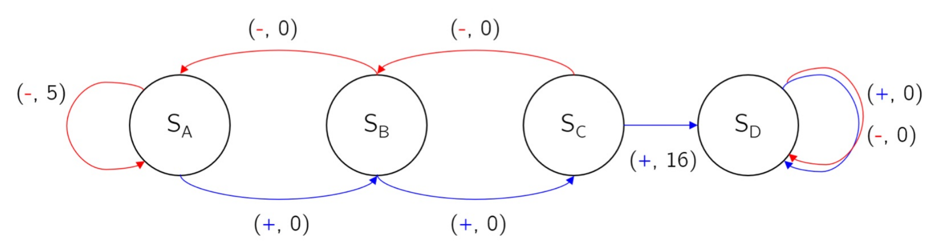

Figure 1: MDP problem with (action, reward) pairs.

Problem 1a [2 points] ✎

Find the optimal policy at the initial state S_A with discount factor $\gamma = 0.001$. Justify your answer.

Problem 1b [2 points] ✎

Find the optimal policy at the initial state S_A with discount factor $\gamma = 0.999$. Justify your answer.

Problem 1c [2 points] ✎

What is the optimal policy at the initial state S_B? Explain your answer in terms of discount factor $\gamma \in (0, 1)$.

1a, 1b $\pi(S_A)$는 $+$와 $-$이 될 수 있다. $T(s, \pi(s), s') = 1$이다. (No randomness)

① $\pi(S_A) = +$ ② $\pi(S_A) = -$

let action sequence = e (episode)

$e = (+,+,+) \rightarrow R_e = 16\gamma^2$ $e = (-,-,-) \rightarrow R_e = 5(1+\gamma+\gamma^2)$

$e = (+,+,-) \rightarrow R_e = 0$ $e = (-,-,+) \rightarrow R_e = 5(1+\gamma)$

$e = (+,-,+) \rightarrow R_e = 0$ $e = (-,+,-) \rightarrow R_e = 5$

$e = (+,-,-) \rightarrow R_e = 5\gamma^2$ $e = (-,+,+) \rightarrow R_e = 5$

각 episode의 probability $= \frac{1}{4}$

$\pi(S_A) = +$이고 3회의 action을 취했을 때 $\pi(S_A) = -$이고 3회의 action을 취했을 때

value $= \frac{21\gamma^2}{4}$ value $= \frac{20+10\gamma+5\gamma^2}{4}$

1a. $\gamma = 0.001$ $\frac{21}{4}(0.001)^2 < \frac{20+10\times(0.001)+5\times(0.001)^2}{4}$ optimal policy $= -$

1b. $\gamma = 0.999$ $\frac{21}{4}(0.999)^2 < \frac{20+10\times(0.999)+5\times(0.999)^2}{4}$ optimal policy $= -$

1c. ① $\pi(S_B) = +$ ② $\pi(S_B) = -$

$e = (+,+,+) \rightarrow R_e = 16\gamma$ $e = (-,-,-) \rightarrow R_e = 5(\gamma+\gamma^2)$

$e = (+,+,-) \rightarrow R_e = 16\gamma$ $e = (-,-,+) \rightarrow R_e = 5\gamma$

$e = (+,-,+) \rightarrow R_e = 0$ $e = (-,+,-) \rightarrow R_e = 0$

$e = (+,-,-) \rightarrow R_e = 0$ $e = (-,+,+) \rightarrow R_e = 0$

각 episode의 probability $= \frac{1}{4}$

$\pi(S_B) = +$이고 3회의 action을 취했을 때 $\pi(S_B) = -$이고 3회의 action을 취했을 때

value $= 8\gamma$ value $= \frac{10\gamma+5\gamma^2}{4}$

$10\gamma + 5\gamma^2$, 32γ 비교

$10\gamma + 5\gamma^2 - 32\gamma = 5\gamma^2 - 22\gamma$

$\pi(S_B) = +$이 더 클조건 : $5\gamma^2 - 22\gamma < 0$

$5\gamma(\gamma - \frac{22}{5}) < 0$

$0 < \gamma < \frac{22}{5}$

$\therefore$ $\gamma \in (0,1)$에서 항상

$\pi(S_B) = +$이 episode의 expected value가 더 크다