강화학습

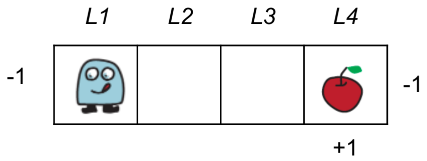
벨만 방정식 - Quiz

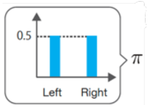
학과 : 산업인공지능학과
학번 : 2024254022
이름 : 정현일

2025.03.23.

충북대학교 산업인공지능학과

(Q1) 4-Grid World 의 각 state 에 대한 state value 를 구하라





$\gamma = 0.9$

$$v_\pi(s) = \sum_\alpha \pi\ (a|s)\{r(s,a,s') + \gamma v_\pi(s')\} \qquad \gamma = 0.9$$

$$\begin{aligned}
v_\pi(L1) &= 0.5\{-1 + 0.9v_\pi(L1)\} + 0.5\{0 + 0.9v_\pi(L2)\} \\
&\rightarrow -0.5 + 0.45v_\pi(L1) + 0.45v_\pi(L2) \\
&\rightarrow -0.55v_\pi(L1) + 0.45v_\pi(L2) = \mathbf{0.5}
\end{aligned}$$

$$\begin{aligned}
v_\pi(L2) &= 0.5\{0 + 0.9v_\pi(L1)\} + 0.5\{0 + 0.9v_\pi(L3)\} \\
&\rightarrow 0.45v_\pi(L1) - v_\pi(L2) + 0.45v_\pi(L3) = \mathbf{0}
\end{aligned}$$

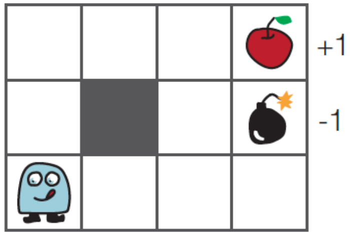
$$\begin{aligned}
v_\pi(L3) &= 0.5\{0 + 0.9v_\pi(L2)\} + 0.5\{1 + 0.9v_\pi(L4)\} \\
&= 0.45v_\pi(L2) + 0.5 + 0.45v_\pi(L4) \\
&\rightarrow 0.45v_\pi(L2) - v_\pi\ (L3) + 0.45v_\pi(L4) = -\ \mathbf{0.5}
\end{aligned}$$

$$\begin{aligned}
v_\pi(L4) &= 0.5\{0 + 0.9v_\pi(L3)\} + 0.5\{-1 + 0.9v_\pi(L4)\} \\
&= 0.45v_\pi(L3) - 0.5 + 0.45v_\pi(L4) \\
&\rightarrow -v_\pi(L4) + 0.45v_\pi(L3) + 0.45v_\pi(L4) = 0.5 \\
&\rightarrow 0.45v_\pi(L3) - 0.55v_\pi(L4) = \mathbf{0.5}
\end{aligned}$$

$$\left[\begin{array}{l}
-0.55v_\pi(L1) + 0.45v_\pi(L2) = \mathbf{0.5} \\
0.45v_\pi(L1) - v_\pi(L2) + 0.45v_\pi(L3) = \mathbf{0} \\
0.45v_\pi(L2) - v_\pi\ (L3) + 0.45v_\pi(L4) = -\ \mathbf{0.5} \\
0.45v_\pi(L3) - 0.55v_\pi(L4) = \mathbf{0.5}
\end{array}\right.$$

$$\left[\begin{array}{l}
v_\pi(L1) = -1.81407563025210 \\
v_\pi(L2) = -1.10609243697479 \\
v_\pi(L3) = -0.643907563025210 \\
v_\pi(L4) = -1.43592436974790
\end{array}\right.$$

(Q2) 3x4 grid world 의 각 state 에 대한 state value 를 구하라



$$v_\pi(s) = \sum_{\alpha\in\{up,down,left,right\}} \pi(a|s)\{r(s,a,s') + \gamma v_\pi(s')\}$$

$$= 0.25\sum_{\alpha}\{r(s,a,s') \ + \ \gamma v_\pi(s')\}$$

$$\begin{aligned}
v_\pi(0{,}0) &= 0.25\{0 + 0.9v_\pi(0{,}0) + 0 + 0.9v_\pi(0{,}1) + 0 + 0.9v_\pi(0{,}0) + 0 + 0.9v_\pi(1{,}0)\} \\
&= 0.25\{0.9(2v_\pi(0{,}0) + v_\pi(0{,}1) + v_\pi(1{,}0))\} \\
v_\pi(0{,}1) &= 0.25\{0 + 0.9v_\pi(0{,}0) + 0 + 0.9v_\pi(0{,}2) + 0 + 0.9v_\pi(0{,}1) + 0 + 0.9v_\pi(0{,}1)\} \\
&= 0.25\{0.9(v_\pi(0{,}0) + 2v_\pi(0{,}1) + v_\pi(0{,}2))\} \\
v_\pi(0{,}2) &= 0.25\{0 + 0.9v_\pi(0{,}1) + 1 + 0.9v_\pi(0{,}3) + 0 + 0.9v_\pi(0{,}2) + 0 + 0.9v_\pi(1{,}2)\} \\
&= 0.25\{1 + \ 0.9(v_\pi(0{,}1) + v_\pi(0{,}2) + v_\pi(1{,}2))\} \\
v_\pi(1{,}0) &= 0.25\{0 + 0.9v_\pi(1{,}0) + 0 + 0.9v_\pi(1{,}0) + 0 + 0.9v_\pi(0{,}0) + 0 + 0.9v_\pi(2{,}0)\} \\
&= 0.25\{0.9(2v_\pi(1{,}0) + v_\pi(0{,}0) + v_\pi(2{,}0))\} \\
v_\pi(1{,}2) &= 0.25\{0 + 0.9v_\pi(1{,}2) - 1 + 0.9v_\pi(1{,}3) + 0 + 0.9v_\pi(0{,}2) + 0 + 0.9v_\pi(2{,}2)\} \\
&= 0.25\{-1 + 0.9(v_\pi(1{,}2) + v_\pi(1{,}3) + v_\pi(0{,}2) + v_\pi(2{,}2))\} \\
v_\pi(1{,}3) &= 0.25\{0 + 0.9v_\pi(1{,}2) + 0 + 0.9v_\pi(1{,}3) + 1 + 0.9v_\pi(0{,}3) + 0 + 0.9v_\pi(2{,}3)\} \\
&= 0.25\{1 + 0.9(v_\pi(1{,}2) + v_\pi(1{,}3) + v_\pi(2{,}3))\}
\end{aligned}$$

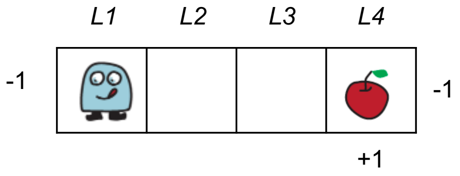
(Q2) 3x4 grid world 의 각 state 에 대한 state value 를 구하라

$$
\begin{aligned}
v_\pi(2{,}0) &= \ 0.25\{0 + 0.9v_\pi(2{,}0) + 0 + 0.9v_\pi(2{,}1) + 0 + 0.9v_\pi(1{,}0) + 0 + 0.9v_\pi(2{,}0)\} \\
&= 0.25\{0.9(2v_\pi(2{,}0) + v_\pi(2{,}1) + v_\pi(1{,}0))\} \\
v_\pi(2{,}1) &= \ 0.25\{0 + 0.9v_\pi(2{,}0) + 0 + 0.9v_\pi(2{,}2) + 0 + 0.9v_\pi(2{,}1) + 0 + 0.9v_\pi(2{,}1)\} \\
&= 0.25\{0.9(v_\pi(2{,}0) + v_\pi(2{,}2) + 2v_\pi(2{,}1))\} \\
v_\pi(2{,}2) &= \ 0.25\{0 + 0.9v_\pi(2{,}1) + 0 + 0.9v_\pi(2{,}3) + 0 + 0.9v_\pi(1{,}2) + 0 + 0.9v_\pi(2{,}2)\} \\
&= 0.25\{0.9(v_\pi(2{,}1) + v_\pi(2{,}3) + v_\pi(1{,}2) + v_\pi(2{,}2))\} \\
v_\pi(2{,}3) &= \ 0.25\{0 + 0.9v_\pi(2{,}2) + 0 + 0.9v_\pi(2{,}3) - 1 + 0.9v_\pi(1{,}3) + 0 + 0.9v_\pi(2{,}3)\} \\
&= 0.25\{-1 + \ 0.9(v_\pi(2{,}2) + 2v_\pi(2{,}3) + v_\pi(1{,}3))\}
\end{aligned}
$$

$$
\left[
\begin{aligned}
v_\pi(0{,}0) &= 0.25\{0.9(2v_\pi(0{,}0) + v_\pi(0{,}1) + v_\pi(1{,}0))\} \\
v_\pi(0{,}1) &= 0.25\{0.9(v_\pi(0{,}0) + 2v_\pi(0{,}1) + v_\pi(0{,}2))\} \\
v_\pi(0{,}2) &= 0.25\{1 + \ 0.9(\, v_\pi(0{,}1) + v_\pi(0{,}2) + v_\pi(1{,}2))\} \\
v_\pi(1{,}0) &= 0.25\{0.9(2v_\pi(1{,}0) + v_\pi(0{,}0) + v_\pi(2{,}0))\} \\
v_\pi(1{,}2) &= 0.25\{-1 + 0.9(v_\pi(1{,}2) + v_\pi(1{,}3) + v_\pi(0{,}2) + v_\pi(2{,}2))\} \\
v_\pi(1{,}3) &= 0.25\{1 + 0.9(v_\pi(1{,}2) + v_\pi(1{,}3) + v_\pi(2{,}3))\} \\
v_\pi(2{,}0) &= 0.25\{0.9(2v_\pi(2{,}0) + v_\pi(2{,}1) + v_\pi(1{,}0))\} \\
v_\pi(2{,}1) &= 0.25\{0.9(v_\pi(2{,}0) + v_\pi(2{,}2) + 2v_\pi(2{,}1))\} \\
v_\pi(2{,}2) &= 0.25\{0.9(v_\pi(2{,}1) + v_\pi(2{,}3) + v_\pi(1{,}2) + v_\pi(2{,}2))\} \\
v_\pi(2{,}3) &= 0.25\{-1 + \ 0.9(v_\pi(2{,}2) + 2v_\pi(2{,}3) + v_\pi(1{,}3))\}
\end{aligned}
\right.
$$

$$
\left[
\begin{aligned}
v_\pi(0{,}0) &= 0.0541 \\
v_\pi(0{,}1) &= 0.134 \\
v_\pi(0{,}2) &= 0.2733 \\
v_\pi(1{,}0) &= -0.0017 \\
v_\pi(1{,}2) &= \text{-}0.3036 \\
v_\pi(1{,}3) &= 0.0778 \\
v_\pi(2{,}0) &= -0.0582 \\
v_\pi(2{,}1) &= -0.1407 \\
v_\pi(2{,}2) &= -0.2856 \\
v_\pi(2{,}3) &= -0.5396
\end{aligned}
\right.
$$

(Q3) 4-Grid World 의 각 state 에 대한 optimal policy 를 구하기 위한 연립 방정식을 유도하라



$$v_{.}(s) = max_{\alpha} \sum_{s'} p\,(s'|s,a)\{r(s,a,s') + \gamma v_{.}(s')\}$$

$$v_{.}(s) = max_{\alpha}\,\{r(s,a,s') + \gamma v_{.}(s')\}$$

$$v_{.}(L1) = max \begin{Bmatrix} -1\ +\ 0.9v_{.}(L1), \\ 0 + 0.9v_{.}(L2) \end{Bmatrix} = 4.263 \qquad = \begin{Bmatrix} 2.8367, a = Left \\ 4.2633, a = Right \end{Bmatrix} \qquad v_{.}(L1) = Right$$

$$v_{.}(L2) = max \begin{Bmatrix} 0 +\ 0.9v_{.}(L1), \\ 0 + 0.9v_{.}(L3) \end{Bmatrix} = 4.737 \qquad = \begin{Bmatrix} 3.8367, a = Left \\ 4.7367, a = Right \end{Bmatrix} \qquad v_{.}(L2) = Right$$

$$v_{.}(L3) = max \begin{Bmatrix} 0 +\ 0.9v_{.}(L2), \\ 1 + 0.9v_{.}(L4) \end{Bmatrix} = 5.263 \qquad = \begin{Bmatrix} 4.2633, a = Left \\ 5.2633, a = Right \end{Bmatrix} \qquad v_{.}(L3) = Right$$

$$v_{.}(L4) = max \begin{Bmatrix} 0 +\ 0.9v_{.}(L3), \\ -1 + 0.9v_{.}(L4) \end{Bmatrix} = 4.737 \qquad = \begin{Bmatrix} 4.7367, a = Left \\ 3.2633, a = Right \end{Bmatrix} \qquad v_{.}(L4) = Left$$