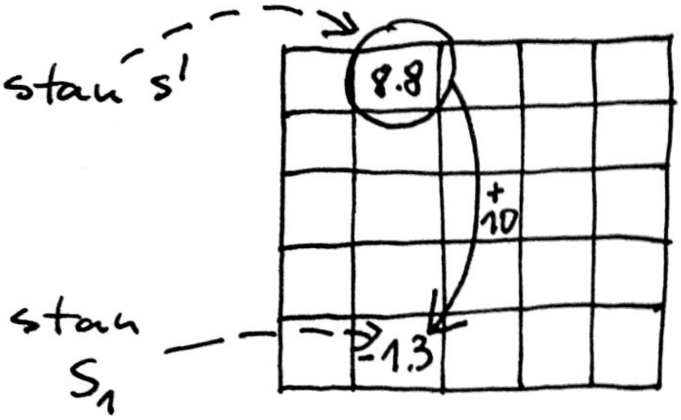
Wskazówka do zadania 9 z LAB3

Pokażę, że równanie Bellmana zachodzi dla stanu oznaczonego poniżej kółkiem:



Wszystkie 4 akcje $(\uparrow, \downarrow, \rightarrow, \leftarrow)$ w tym stanie przenoszą agenta do stanu w ostatnim wierszu dla którego

$$V_\pi(s) = -1.3$$

Stan oznaczony na rysunku kółkiem nazwijmy s'.

Korzystając z równania Bellmana chcemy policzyć

$$V_\pi(s')$$

Z treści zadania wynika, że wartość ta wynosi 8.8.

Zobaczymy czy otrzymamy taką wartość.

Równanie Bellmana:

$$v_\pi(s') = \sum_a \pi(a|s') \sum_{s,r} p(s,r|s',a) \cdot \left[r + \gamma v_\pi(s) \right]$$

W naszym przypadku a przyjmuje cztery wartości:

$$\uparrow, \downarrow, \rightarrow, \leftarrow$$

W przypadku każdej akcji nagroda wynosi +10 i stan do którego przechodzi agent to stan S_1.

Z równania Bellmana otrzymujemy zatem:

$$V_\pi(s') = \pi(\uparrow|s')\sum_{s,r} p(s,r|s',\uparrow)\cdot\left[r + \gamma\cdot V_\pi(s)\right]$$
$$+ \pi(\downarrow|s')\sum_{s,r} p(s,r|s',\downarrow)\cdot\left[r + \gamma\cdot V_\pi(s)\right]$$
$$+ \pi(\rightarrow|s')\sum_{s,r} p(s,r|s',\rightarrow)\cdot\left[r + \gamma\cdot V_\pi(s)\right]$$
$$+ \pi(\leftarrow|s')\sum_{s,r} p(s,r|s',\leftarrow)\left[r + \gamma\cdot V_\pi(s)\right]$$

Z treści zadania wiemy, że:

dla wszystkich akcji a: $\pi(a|s') = \frac{1}{4}$

co więcej dla wszystkich akcji a:

$$p(s_1, 10|s', a) = 1$$

Pozostałe prawdopodobieństwa przejścia wynoszą 0.

A zatem otrzymujemy:

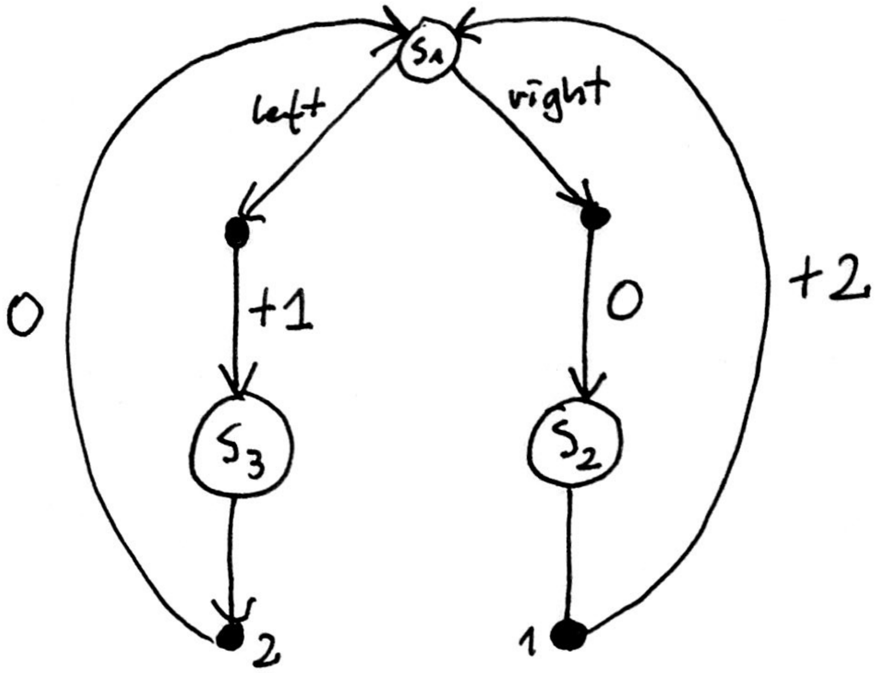
$$V_\pi(s') = \frac{1}{4}\cdot 1\cdot\left[10 + \gamma V_\pi(s_1)\right] + \frac{1}{4}\cdot 1\cdot\left[10 + \gamma V_\pi(s_1)\right]$$
$$+ \frac{1}{4}\cdot 1\cdot\left[10 + \gamma V_\pi(s_1)\right] + \frac{1}{4}\cdot 1\cdot\left[10 + \gamma\cdot V_\pi(s_1)\right] =$$

$$= 4\cdot\frac{1}{4}\left[10 + \gamma\cdot \underbrace{V_\pi(s_1)}_{=\,-1.3}\right] = 10 + \gamma\cdot(-1.3)$$

Dla $\gamma = 0.9$:

$$V_\pi(s') = 10 + 0.9\cdot(-1.3) = 8.83 \approx 8.8$$

Wskazówka do zadania 1 z LAB 4



Wykorzystamy równanie optymalizacyjne Bellmana:

$$V_*(s) = \max_a \sum_{s',r} p(s',r|s,a)\left[r + \gamma \cdot V_*(s')\right]$$

Zatem:

$$V_*(s_1) = \max\left\{\sum_{s',r} p(s',r|s_1,\text{left})\left[r + \gamma \cdot V_*(s')\right], \sum_{s',r} p(s',r|s_1,\text{right})\left[r + \gamma \cdot V_*(s')\right]\right\} \quad (\Delta)$$

wiemy, że:

$$p(s_3,1|s_1,\text{left}) = 1$$
$$p(s_2,0|s_1,\text{right}) = 1 \quad (\square)$$

Pozostałe prawdopodobieństwa przejścia ze stanu s_1 wynoszą 0.

Podstawiamy (□) do (△):

$$V_*(s_1) = \max\left\{1\cdot\left[1+\gamma\cdot V_*(s_3)\right],\ 1\cdot\left[0+\gamma\cdot V_*(s_2)\right]\right\}$$

Po uproszczeniu:

$$V_*(s_1) = \max\left\{1+\gamma\cdot V_*(s_3),\ \gamma\cdot V_*(s_2)\right\}$$

Zastosujmy teraz równanie optymalizacyjne Bellmana do stanu s_2:

$$V_*(s_2) = \sum_{s',r} p(s',r\mid s_2,①)\cdot\left[r+\gamma\cdot V_*(s')\right]$$

(pomijamy $\max\limits_a$ bo w stanie s_2 możliwa jest tylko jedna akcja oznaczona ①)

Oczywiście:

$$p(s_1,2\mid s_2,①) = 1$$

Zatem:

$$V_*(s_2) = 2+\gamma\cdot V_*(s_1)$$

Pozostało nam do wyliczenia (w analogiczny sposób):

$$V_*(s_3) - ?$$

Otrzymamy wówczas:

$$\begin{cases} V_*(s_1) = \max\left\{1+\gamma\cdot V_*(s_3),\ \gamma\cdot V_*(s_2)\right\} \\ V_*(s_2) = 2+\gamma\cdot V_*(s_1) \\ V_*(s_3) = ? \end{cases} \qquad (*)$$

Są to 3 równania z trzema niewiadomymi:

$$V_*(s_1),\ V_*(s_2),\ V_*(s_3)$$

Układ równań (*) należy rozwiązać trzykrotnie, dla 3 różnych wartości γ:

$$\gamma = 0,\ \gamma = 0.9,\ \gamma = 0.5$$