Reinforcement Learning (SS18) - Exercise 2

Daniel Hennes

09.05.2018

5. **Calculate the optimal value function v_* for the cleaning robot problem.**

We assume that the optimal (or only) action for state $s = 0$ is $a = 1$ and for state $s = 5$ it is $a = -1$.

From the definition of v_* we get the following system of non-linear equations:

$$
\begin{aligned}
v_*(0) &= \gamma v_*(1) \\
v_*(1) &= \max\{1 + \gamma v_*(0), \gamma v_*(2)\} \\
v_*(2) &= \max\{\gamma v_*(1), \gamma v_*(3)\} \\
v_*(3) &= \max\{\gamma v_*(2), \gamma v_*(4)\} \\
v_*(4) &= \max\{\gamma v_*(3), 5 + \gamma v_*(5)\} \\
v_*(5) &= \gamma v_*(4)
\end{aligned}
$$

Independent of the value of γ we know that for states 3 and 4 the optimal action is 1. For state 4 the argument is trivial due to the immediate reward; in state 3 we need two transitions to receive reward when going *right* and three transitions when going *left*.

$$
\begin{aligned}
v_*(3) &= \gamma v_*(4) \\
v_*(4) &= 5 + \gamma v_*(5)
\end{aligned}
$$

With $v_*(5) = \gamma v_*(4)$: $v_*(4) = 5 + \gamma^2 v_*(4) \Rightarrow v_*(4) = \frac{5}{1-\gamma^2}$ and $v_*(3) = v_*(5) = \frac{5\gamma}{1-\gamma^2}$.

We want to solve $\gamma v_*(1) \stackrel{!}{=} \gamma v_*(3)$, thus making the agent indifferent in state 2. If the agent is indifferent in state 2, going *left* is optimal, it follows that in state 1 going *left* must be optimal otherwise the agent would alternate between 1 and 2 and not accumulate reward.

$$
v_*(1) = 1 + \gamma v_*(0)
$$

With $v_*(0) = \gamma v_*(1)$: $v_*(1) = 1 + \gamma^2 v_*(1) \Rightarrow v_*(1) = \frac{1}{1-\gamma^2}$ and $v_*(0) = \frac{\gamma}{1-\gamma^2}$.

Thus,

$$
\gamma v_*(1) \stackrel{!}{=} \gamma v_*(3) \Leftrightarrow v_*(1) = \frac{1}{1-\gamma^2} \stackrel{!}{=} \frac{5\gamma}{1-\gamma^2} = v_*(3) \Rightarrow \gamma = \frac{1}{5}
$$

For $\gamma = \frac{1}{5}$ the agent is indifferent in state 2 and we have $v_*(2) = \gamma v_*(1) = \gamma v_*(3) = \frac{5}{24}$.

For $\gamma < \frac{1}{5}$ the optimal action is -1 (*left*) and we have $v_*(2) = \gamma v_*(1) = \frac{\gamma}{1-\gamma^2}$.

For $\gamma > \frac{1}{5}$ the optimal action is 1 (*right*) and we have $v_*(2) = \gamma v_*(3) = \frac{5\gamma^2}{1-\gamma^2}$.

For the last case we look at the value in state 1:

$$
v_*(1) = \max\{1 + \gamma v_*(0), \gamma v_*(2)\}
$$

If the agent is indifferent in state 1, action -1 must be optimal and thus we have from above: $v_*(1) = \frac{1}{1-\gamma^2}$.

Thus[1],

$$v_*(1) = \frac{1}{1-\gamma^2} \overset{!}{=} \frac{5\gamma^3}{1-\gamma^2} = \gamma v_*(2) \Leftrightarrow \frac{1}{5} = \gamma^3 \Rightarrow \gamma = \sqrt[3]{\frac{1}{5}}$$

For $\gamma = \sqrt[3]{\frac{1}{5}}$ the agent is indifferent in state 1 and we have $v_*(1) = 1 + \gamma v_*(0) = \gamma v_*(2) = \frac{1}{1-\frac{1}{5^{2/3}}}$.

For $\gamma < \sqrt[3]{\frac{1}{5}}$ the optimal action is -1 (*left*) and we have $v_*(1) = \frac{1}{1-\gamma^2}$.

For $\gamma > \sqrt[3]{\frac{1}{5}}$ the optimal action is 1 (*right*) and we have $v_*(1) = \gamma v_*(2) = \frac{5\gamma^3}{1-\gamma^2}$.

In summary:

$$\begin{array}{rcccccc}
0 \leq \gamma \leq \frac{1}{5}: & v_*(0) = \frac{\gamma}{1-\gamma^2} & v_*(1) = \frac{1}{1-\gamma^2} & v_*(2) = \frac{\gamma}{1-\gamma^2} & v_*(3) = \frac{5\gamma}{1-\gamma^2} & v_*(4) = \frac{5}{1-\gamma^2} & v_*(5) = \frac{5\gamma}{1-\gamma^2} \\
\frac{1}{5} \leq \gamma \leq \sqrt[3]{\frac{1}{5}}: & v_*(0) = \frac{\gamma}{1-\gamma^2} & v_*(1) = \frac{1}{1-\gamma^2} & v_*(2) = \frac{5\gamma^2}{1-\gamma^2} & v_*(3) = \frac{5\gamma}{1-\gamma^2} & v_*(4) = \frac{5}{1-\gamma^2} & v_*(5) = \frac{5\gamma}{1-\gamma^2} \\
\sqrt[3]{\frac{1}{5}} \leq \gamma < 1: & v_*(0) = \frac{5\gamma^4}{1-\gamma^2} & v_*(1) = \frac{5\gamma^3}{1-\gamma^2} & v_*(2) = \frac{5\gamma^2}{1-\gamma^2} & v_*(3) = \frac{5\gamma}{1-\gamma^2} & v_*(4) = \frac{5}{1-\gamma^2} & v_*(5) = \frac{5\gamma}{1-\gamma^2}
\end{array}$$

In terms of optimal policies we have:

$$\begin{array}{rcccccc}
\gamma = 0: & \overset{0}{\leftrightarrow} & \overset{1}{\leftarrow} & \overset{2}{\leftrightarrow} & \overset{3}{\leftrightarrow} & \overset{4}{\rightarrow} & \overset{5}{\leftrightarrow} \\
0 < \gamma < \frac{1}{5}: & \overset{0}{\rightarrow} & \overset{1}{\leftarrow} & \overset{2}{\leftarrow} & \overset{3}{\rightarrow} & \overset{4}{\rightarrow} & \overset{5}{\leftarrow} \\
\gamma = \frac{1}{5}: & \overset{0}{\rightarrow} & \overset{1}{\leftarrow} & \overset{2}{\leftrightarrow} & \overset{3}{\rightarrow} & \overset{4}{\rightarrow} & \overset{5}{\leftarrow} \\
\frac{1}{5} < \gamma < \sqrt[3]{\frac{1}{5}}: & \overset{0}{\rightarrow} & \overset{1}{\leftarrow} & \overset{2}{\rightarrow} & \overset{3}{\rightarrow} & \overset{4}{\rightarrow} & \overset{5}{\leftarrow} \\
\gamma = \sqrt[3]{\frac{1}{5}}: & \overset{0}{\rightarrow} & \overset{1}{\leftrightarrow} & \overset{2}{\rightarrow} & \overset{3}{\rightarrow} & \overset{4}{\rightarrow} & \overset{5}{\leftarrow} \\
\sqrt[3]{\frac{1}{5}} < \gamma < 1: & \overset{0}{\rightarrow} & \overset{1}{\rightarrow} & \overset{2}{\rightarrow} & \overset{3}{\rightarrow} & \overset{4}{\rightarrow} & \overset{5}{\leftarrow}
\end{array}$$

[1]Alternatively, you can solve: $1 + \gamma v_*(0) = 1 + \frac{\gamma^2}{1-\gamma^2} \overset{!}{=} \frac{5\gamma^3}{1-\gamma^2} = \gamma v_*(2) \Leftrightarrow 1 - \gamma^2 + \gamma^2 = 1 \overset{!}{=} 5\gamma^3 \Rightarrow \gamma = \sqrt[3]{\frac{1}{5}}$