TP2 learning for robotics

Sarah Curtit Caroline Pascal

September 2019

1 Question 1

$$\Pi(S_0) = \left\{ \begin{array}{l} a_1 : P(S_1) = 1 \\ a_2 : P(S_2) = 1 \end{array} \right.$$
$$\Pi(S_1) = \left\{ \; a_0 : P(S_1) = 1 - x, P(S_3) = x \right.$$
$$\Pi(S_2) = \left\{ \; a_0 : P(S_0) = 1 - y, P(S_3) = y \right.$$
$$\Pi(S_3) = \left\{ \; a_0 : P(S_0) = 1 \right.$$

2 Question 2

$$\begin{aligned} V^*(S_0) &= R(S_0) + \max_a \gamma \sum_{S'} T(S, a, S')V^*(S') \\ &= \max \gamma (V^*(S_1'), V^*(S_2')) \end{aligned}$$

$$\begin{aligned} V^*(S_1) &= R(S_1) + \max_a \gamma \sum_{S'} T(S, a, S')V^*(S') \\ &= \gamma \Big((1 - x) * V^*(S_1') + x * V^*(S_3') \Big) \end{aligned}$$

$$\begin{aligned} V^*(S_2) &= R(S_2) + \max_a \gamma \sum_{S'} T(S, a, S')V^*(S') \\ &= 1 + \gamma \Big((1 - y) * V^*(S_0') + y * V^*(S_3') \Big) \end{aligned}$$

$$\begin{aligned} V^*(S_3) &= R(S_3) + \max_a \gamma \sum_{S'} T(S, a, S')V^*(S') \\ &= 10 + \gamma V^*(S_0') \end{aligned}$$

2.1 Question 3

$$\begin{aligned} V_{t+1}(S_2) &= \gamma(1 - y)V_t(S_0) + \gamma y V_t(S_3) + 1 \\ &= \gamma(1 - y)V_t(S_0) + 10\gamma y + \gamma^2 y V_t - 1)(S_0) + 1 \\ V_{t+1}(S_2) &> \gamma(1 - y)V_t(S_0) + \gamma^2 y (V_{t-1}(S_0) \end{aligned}$$

we have

$$V_{t+1}(S_0) \geq \gamma V_t(S_1)$$

so

$$V_{t+1}(S_2) > \gamma^2(1 - y)V_{t-1}(S_1) + \gamma^3 y V_{t-2}(S_1)$$

For x = 0

$$V_{t+1}(S_1) = \gamma V_t(S_1)$$

We can therefore re-write the inequality

$$
\begin{aligned}
V_{t+1}(S_2) &> (1-y)V_{t+1}(S_1) + yV_{t+1}(S_1) \\
V_{t+1}(S_2) &> V_{t+1}(S_1) \\
V^*(S_2) &> V^*(S_1)
\end{aligned}
$$

2.2 Question 4

We know that the optimum value functions for each state converge
Therefore $\lim_{t\to\infty} V_t = V^*$ and we can consider than once it converged, we can approximate

$$V_t(S) = V_{t+1}(S)$$

Let's suppose $V(S_1) = V(S_2)$ We have

$$
\begin{aligned}
V(S_1) &= \frac{10\gamma x}{1-\gamma(1-x)-\gamma^3 x} \\
V(S_2) &= \gamma^2(1-y)V(S_1) + 10\gamma y + \gamma^3 y V(S1) + 1
\end{aligned}
$$

$$
\begin{aligned}
V(S_1) &= V(S_2) \\
V(S_1) &> \gamma^2(1-y)V(S_1) + 10\gamma y + \gamma^3 y V(S1) + 1 \\
V(S_1)\Big(1-\gamma^2(1-y)-\gamma^3 y\Big) &= 1 + 10\gamma y
\end{aligned}
$$

Let's take any tuple of values for which $1-\gamma(1-x)-\gamma^3 x > 0$
For example $x = 0.1\gamma = 0.1$ The inequation becomes

$$
\begin{aligned}
10\gamma x\Big(1-\gamma^2(1-y)-\gamma^3 y\Big) &= \Big(1+10\gamma y\Big)\Big(1-\gamma(1-x)-\gamma^3 x\Big) \\
y &< \frac{9\gamma^2 x + 9\gamma x - 1}{10\gamma(\gamma x + 1)}
\end{aligned}
$$

However for this tupple of values,

$$\frac{9\gamma^2 x + 9\gamma x - 1}{10\gamma(\gamma x + 1)} \simeq -0.89$$

which is impossible for $y \in [0,1]$
Therefore, there's no value of y for which $\Pi(S_0) = a_1 \forall x > 0, \gamma \in [0,1]$

2.3 Question 5

Running the value iteration, we obtain

$$
\begin{aligned}
V^*(S_0) &\simeq 14.19 \\
V^*(S_1) &\simeq 15.76 \\
V^*(S_1) &\simeq 15.70 \\
V^*(S_1) &\simeq 22.77
\end{aligned}
$$

$$
\begin{aligned}
\Pi^*(S_0) &= a_1 \\
\Pi^*(S_1) &= a_0 \\
\Pi^*(S_1) &= a_0 \\
\Pi^*(S_1) &= a_0
\end{aligned}
$$