6CCS3ML1 (Machine Learning)

Tutorial 08

(Version 1.2)

1. Figure 1 shows a world in which a reinforcement learning agent operates. Write down the Bellman equation (slide 7, lecture Reinforcement Learning 2) for state $(1,1)$

2. Question 4 in Tutorial 7 provides data from some runs that the agent undertakes in the world of Figure 1. Using the utility and probability estimates that can be computed from this data, carry out one round of value iteration for the states $(1,1)$ and $(3,3)$. Take γ to be 1.

   Note that doing this is one of the steps in applying Adaptive Dynamic Programming for the world of Figure 1.

3. Again using the data provided by Question 4 in Tutorial 7, carry out a Temporal Difference update for states $(1,1)$ and $(3,3)$.

   Take $\gamma = 1$ and $\alpha = 0.1$.

4. The update rule in Q-learning is:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left(R(s) + \gamma \max_{a'} Q(s',a') - Q(s,a) \right)$$

   Explain the intuition behind the rule.

5. (a) Given the following run:

$$(1,1)_{-0.04} \overset{Up}{\rightarrow} (1,2)_{-0.04} \overset{Up}{\rightarrow} (1,3)_{-0.04} \overset{Right}{\rightarrow} (1,3)_{-0.04} \overset{Right}{\rightarrow} (2,3)_{-0.04}$$
$$\overset{Right}{\rightarrow} (2,3)_{-0.04} \overset{Right}{\rightarrow} (2,3)_{-0.04} \overset{Right}{\rightarrow} (3,3)_{-0.04} \overset{Right}{\rightarrow} (4,3)_{1}$$

   write down the state/action representation for $(1,3)$ and $(3,3)$, and the Q-values for these state/action representations.

   (b) Given the additional run:

$$(1,1)_{-0.04} \overset{Up}{\rightarrow} (1,2)_{-0.04} \overset{Up}{\rightarrow} (1,2)_{-0.04} \overset{Up}{\rightarrow} (1,3)_{-0.04}$$
$$\overset{Right}{\rightarrow} (2,3)_{-0.04} \overset{Right}{\rightarrow} (3,3)_{-0.04} \overset{Right}{\rightarrow} (4,3)_{1}$$

   compute the updated Q-values relating to $(1,3)$ and $(3,3)$.

6. How do updates in SARSA differ from those in Q-learning?

7. What would SARSA compute as the Q-values relating to $(1,3)$ and $(3,3)$ following the second run in Question 5?

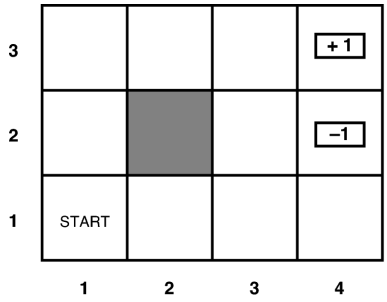

Figure 1: A familiar scenario

Version list

- Version 1.0, March 16th 2019.
- Version 1.1, March 24th 2019.
- Version 1.2, February 9th 2021.