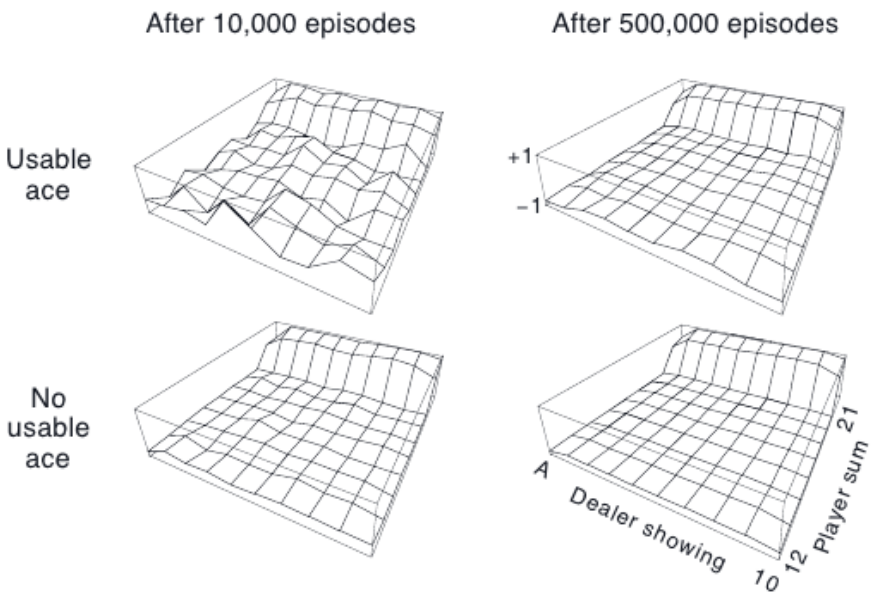
Reinforcement Learning: Tutorial 4

(week from 22. 2. 2016)

1. Is the reinforcement learning framework adequate to usefully represent all goal-directed learning tasks? Can you think of any clear exceptions?

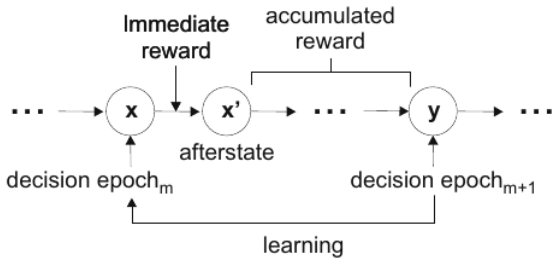
[this and other problems on this sheet are from Sutton & Barto's book]

2. The Figure shows the approximate state-value functions for the blackjack policy that sticks only on 20 or 21, computed by Monte Carlo policy evaluation. Why does the estimated value function jump up for the last two rows in the rear? Why does it drop off for the whole last row on the left? Why are the front-most values higher in the upper diagrams than in the lower?

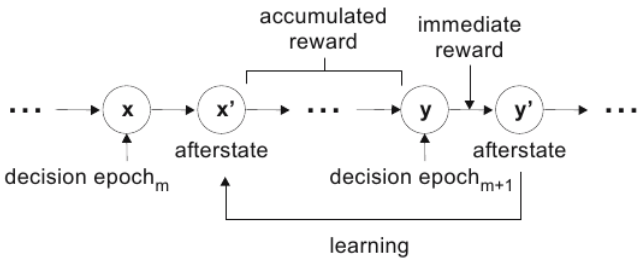


3. Reinforcement learning systems do not need to be “taught” by knowledgeable “teachers”; they learn from their own experience. But teachers of various types can still be helpful. Describe a few different ways in which a teacher might facilitate learning. For each, give a specific example scenario and explain what makes learning more efficient.

4. Consider the game of Tic-Tac-Toe (or any board game with a finite number of moves). How do you define the state space and state value function for such a game? One could define an afterstate in terms of board positions after the agent has made its move. How does this impact on the information required to calculate a value function, and what might be the advantage of such a scheme – explain with a simple example scenario.



(a) Conventional learning.



(b) Afterstates learning.

Describe how the task of Jack's Car Rental (see lecture or Example 4.2 in S&B) could be reformulated in terms of afterstates. Why, in terms of this specific task, would such a reformulation be likely to speed convergence?

5. Suppose, a reinforcement learning algorithm is trained for game playing (e.g. chess), but instead of playing against a random opponent, it played against itself. What do you think would happen in this case? Would it learn a different way of playing?

6. Dynamic programming is said to be "model-based". What does "model" mean in this context? Give an example of a "model-free" reinforcement learning method, explaining in what sense it is "model-free".

7. Suppose you have the task of getting a mine-detection system to plot a safe path through a minefield. The mines are reasonably easily detectable and you have a few small, relatively cheap (and therefore, to some extent, disposable) robots that you can use to aid you. Discuss how you might use reinforcement learning to find a safe path.

8. You find that in an RL problem discrete quantities is not accurate enough; instead, you use continuous quantities. How would you modify your learning system? What is a feature vector and how may one be used in the representation of a state? Why would one choose to use a feature-vector representation of a state?