Assignment9: Learning

<xiangru.chen@fau.de> yd08ucoz
<yamei.zhao@fau.de> mo02buqo
<ekaterina.bobrova@fau.de> il71ywod

July 24, 2022

Exercise 9.3 (Passive Reinforcement Learning)

1. Give the transition model to the extent that it can be learned from these trials.

   Given policy trials were:

   - $up|up \rightarrow up|up \rightarrow down|right \rightarrow up|up \rightarrow right|right \rightarrow right|right \rightarrow right|right$
   - $up|up \rightarrow up|up \rightarrow right|right \rightarrow right|right \rightarrow down|right \rightarrow up|up \rightarrow right|right$
   - $right|up \rightarrow right|left \rightarrow up|left \rightarrow right|up$

   Hence, from this trials we can calculate following transition model:

   $$up|up = \frac{\#up|up}{\sum_{i\in\{up,left,right\}} \#i|up} = \frac{6}{8},$$
   $$right|up = \frac{\#right|up}{\sum_{i\in\{up,left,right\}} \#i|up} = \frac{2}{8}$$
   $$up|left = \frac{\#up|left}{\sum_{i\in\{up,left,right\}} \#i|left} = \frac{1}{2},$$
   $$right|left = \frac{\#right|left}{\sum_{i\in\{up,left,right\}} \#i|left} = \frac{1}{2}$$
   $$down|right = \frac{\#right|right}{\sum_{i\in\{up,left,down,right\}} \#i|right} = \frac{2}{8}$$
   $$right|right = \frac{\#right|right}{\sum_{i\in\{up,left,down,right\}} \#i|right} = \frac{6}{8}$$

2. How could we learn the entire model?

   Using direct utility estimation algorithm. For each step in trial we calculate the utility of state using following equation: $U^{\pi}(s) = E[\sum_{t=0}^{\infty} \gamma^t R(S_t)]$. Then,we calculate average utility for the same states over all trials. Given utility we can compute optimal policy.

3. How would we proceed to learn the utilities of the states? Using direct utility estimation algorithm and corresponding equaition of utility of state or using adaptive dynamic

programming and Bellman equation to calculate utility of state.