RL Tutorial-2

February 2023

Q2: Suppose that you want to travel from a start point S to destination point D in minimum average time. There are two options: 1. There are direct route that requires a time units. 2. Take a potential shortcut that requires b time units to go to an intermediate point I. From I you can go to the destination D in c time units or return to the start (this will take additional b time units). You will find out the value of c one you reach the intermediate point I. What you know a priori is that c has one of the m values $c_1, c_2, \ldots, c_m$ with probabilities $p_1, p_2, \ldots, p_m$. Consider two cases: (i) The value of c is constant over time, and (ii) The value c changes each time you return to start state independently of the value at the previous time periods.
Formulate the problem as an SSP problem. Draw the state-transition diagram. Write Bellman's equation and characterize the optimal stationary policies as best as you can in terms of the given problem data. Solve the problem for the case $a = 2, b = 1, c_1 = 0, c_2 = 5p_1 = 0.5, p_2 = 0.5$.

Q2: A gambler engages in a game of successive coin flipping over an infinite horizon. He wins \$1 each time heads comes up, and loses $\$m > 0$ each time two successive tails come up (so the sequence TTTT loses $\$3m$). the gambler at each time period either flips a fair coin or else cheats by flipping a two-headed coin. In the later case, however, he gets caught with probability $p > 0$ before he flips the coin, the game terminates, and the gambler keeps his earnings thus far. The gambler wishes to maximize his expected earnings.
(a) View this as SSP problem. Draw the state-transition diagram. Identify all proper and all improper policies.
(b) Identify a critical value $\bar{m}$, and show that if $m > \bar{m}$, then all improper policies give an infinite cost for some initial state.
(c) Assume that $m > \bar{m}$, and show that it is then optimal to try to cheat if the last flip was tail and to fair otherwise.

Q3: Consider the SSP problem, and assume that $g(i, u) \leq 0, \forall i, u \in U(i)$. Show that either the optimal cost is $-\infty$ for some initial state, or else, under every policy, the system eventually enters with probability 1 a set of cost-free states and never leaves that set thereafter.