Artificial Intelligence Lab 10: Value Iteration

Department Of Computer Science and Engineering, IIT Palakkad

Q1) [Chain MDP] The state is given by $s = (x, y)$. There are 2 actions from each state namely $A = \{left, right\}$. Each action is successful with probability p, and the other action is made with probability $1 - p$. There are two terminal states T_1 and T_2 (once in terminal state, the agent is stuck there forever). The reward in the L state is -1 and R state is $+1$, and every other state it is 0.

1. Generate the chain environment. It should take the following inputs: length of chain and output the model i.e., the reward and transition probabilities (for a given state and action). [25 Marks]

2. Implement the Bellman operator. It takes input as V and outputs TV. [15 Marks]

3. Perform value iteration and output the optimal value function and optimal policy. Start with various values V_0 and plot $||V_t - V_*||_\infty$. [10 Marks]

Q2) [Grid MDP] The state is given by $s = (x, y)$. There are 4 actions from each state namely $A = \{up, down, left, right\}$. Each action is successful with probability p, and with probability $\frac{1-p}{3}$ other 3 actions are chosen.

1. Generate a grid environment. It should take the following inputs: x-size, y-size, goal state, blocked states, and outputs the model, i.e., the reward and transition probabilities (for a given state and action). [20 Marks]

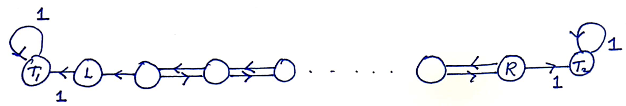
2. Perform value iteration and output the optimal value function and optimal policy. [10 Marks]

Q3) [Mountain Car: Deterministic Control] There is an under-powered car stuck in the bottom of a 1-dim valley. It needs to find its way to the top. The car has three actions namely A=-1,0,+1 which means accelerate backward, no acceleration and accelerate forward respectively. The ranges for position and velocity are [-1.2,0.5] and [-0.07,0.07] respectively. The car is needs to reach the top on the right, i.e., position of 0.5. The dynamics is according to the equations:

$$
\begin{aligned}
v_{t+1} &= v_t + 0.001a_t - 0.0025cos(3p_t) \\
p_{t+1} &= p_t + v_t
\end{aligned}
\tag{1}
$$

1. Perform value iteration and output the optimal value function and optimal policy. [20 Marks] (Hint: Discretise the state space into 100×100 grid (i.e., divide the position and velocity co-ordinates into 100 intervals each.)

CHAIN MDP



MOUNTAIN CAR