AI 3000 / CS 5500 : Reinforcement Learning

Exam № 1

01/10/2024, 06.00 PM - 07.00 PM

Easswar Subramanian, IIT Hyderabad 01/10/2024

Problem 1

A driver wants to park his / her car as close as possible to the restaurant. The restaurant has M parking slots numbered 1 to M with parking slot M being the one that is closest to the restaurant. The satisfaction factor is highest when the car gets parked closest to the restaurant. The probability of a parking slot $i \in \{1, \cdots, M\}$ being available for parking is $p(i)$ and the driver cannot see if a parking slot is available unless he / she is in front of that slot. At each available parking slot $i \in \{1, \cdots, M\}$, the driver can choose to park his / her car or move on to the next slot (even if the slot is available). If the driver doesn't park the car anywhere up until slot M, he / she leaves the restaurant. The objective is to maximize the satisfaction index of the driver.

(a) Formulate the above problem as an MDP by suitably defining the state space ($\mathcal{S}$), action space ($\mathcal{A}$), reward function ($\mathcal{R}$), transition dynamics ($\mathcal{P}$) and discount factor γ. (10 Points)

(b) What will be a suitable objective function to maximize (in terms of the reward function formulated) ? Justify. (3 Points)

(c) Is the problem finite / infinite / indefinite horizon ? Does the nature of the horizon have consequence to the choice of the discount factor ? Explain. (2 Points)

(d) Derive an expression for the optimal value function for a parking slot $i \in \{1, \cdots, M\}$ when the slot i available. (5 Points)

Problem 2

Consider the one-dimensional grid world problem as given below with S as the start state and the double edged states as exit states.

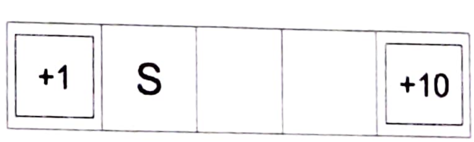

Figure 1: One dimensional grid world

At any non-exit state, the agent can choose **Left** or **Right** actions which results in the agent moving to the intended square with no rewards. At exit states the agent has only action called **Exit** which gives the listed reward pertaining to that state and the game ends thereafter. For now, assume that the discount factor $\gamma = 1$. We start the value iteration algorithm with $V_0(s) = 0$ for all states s of the MDP.

(a) What is the optimal value of $V^*(S)$?

(1 Point)

(b) What is the smallest value of k for which $V_k(S)$ would be non-zero ? What will be $V_k(S)$ for that k ?

(1 Point)

(c) At what k, will $V_k(S) = V^*(S)$?

(1 Point)

(d) What will be the optimal policy for each non-exit state s of the MDP when value iteration converges ?

(1 Point)

(e) Suppose we perform policy iteration for this MDP. Would the policy iteration algorithm converge to the same optimal policy and same optimal value function ? Explain with reasoning. (2 Points)

(f) Suppose if $\gamma = 0.5$, what will be $V^*(S)$? Will the optimal policy remain the same (compared to case of $\gamma = 1$) ?

(1 Point)

(g) Would a different choice of γ result in a different optimal policy for state S ? If so, for what choices of γ would that occur ?

(3 Points)

Problem 3

Consider a two state MRP with states S and T with the state T being the terminal state. The transition probability from state S to T is $\frac{1}{3}$. The reward for being in S and T are 1 and 0 respectively. Let the discount factor $\gamma = 1$. What is the true value of state S and what will be the value of state S estimated via FVMC and EVMC ?

(5 Points)

ALL THE BEST