CS 188 Spring 2021 Introduction to Artificial Intelligence

Exam Prep 9

Q1. VPI

You are the latest contestant on Monty Hall's game show, which has undergone a few changes over the years.
In the game, there are n closed doors: behind one door is a car ($U(car) = 1000$), while the other $n - 1$ doors each have a goat behind them ($U(goat) = 10$). You are permitted to open exactly one door and claim the prize behind it.

You begin by choosing a door uniformly at random.

(a) What is your expected utility?

Answer:

(b) After you choose a door but before you open it, Monty offers to open k other doors, each of which are guaranteed to have a goat behind it.
If you accept this offer, should you keep your original choice of a door, or switch to a new door?

$EU(keep)$:

$EU(switch)$:

Action that achieves MEU:

(c) What is the value of the information that Monty is offering you?

Answer:

(d) Monty is changing his offer!
After you choose your initial door, you are given the offer to choose any other door and open this second door. If you do, after you see what is inside the other door, you may switch your initial choice (to the newly opened door) or keep your initial choice.
What is the value of this new offer?

Answer:

(e) Monty is generalizing his offer: you can pay $\$d^3$ to open d doors as in the previous part. (Assume that $U(\$x) = x$.) You may now switch your choice to any of the open doors (or keep your initial choice). What is the largest value of d for which it would be rational to accept the offer?

Answer:

Q2. How do you Value It(eration)?

(a) Fill out the following True/False questions.

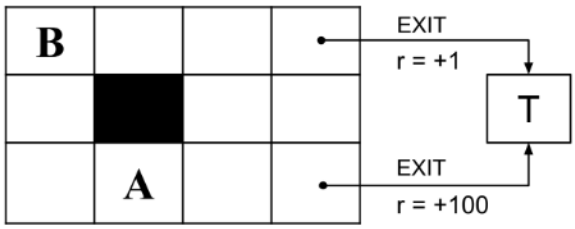
(i) ◯ True ◯ False: Let A be the set of all actions and S the set of states for some MDP. Assuming that $|A| \ll |S|$, one iteration of value iteration is generally faster than one iteration of policy iteration that solves a linear system during policy evaluation.

(ii) ◯ True ◯ False: For any MDP, changing the discount factor does not affect the optimal policy for the MDP.

The following problem will take place in various instances of a grid world MDP. Shaded cells represent walls. In all states, the agent has available actions $\uparrow, \downarrow, \leftarrow, \rightarrow$. Performing an action that would transition to an invalid state (outside the grid or into a wall) results in the agent remaining in its original state. In states with an arrow coming out, the agent has an *additional* action $EXIT$. In the event that the $EXIT$ action is taken, the agent receives the labeled reward and ends the game in the terminal state T. Unless otherwise stated, all other transitions receive no reward, and all transitions are deterministic.

For all parts of the problem, assume that value iteration begins with all states initialized to zero, i.e., $V_0(s) = 0 \ \forall s$. **Let the discount factor be $\gamma = \frac{1}{2}$ for all following parts**.

(b) Suppose that we are performing value iteration on the grid world MDP below.



(i) Fill in the optimal values for A and B in the given boxes.

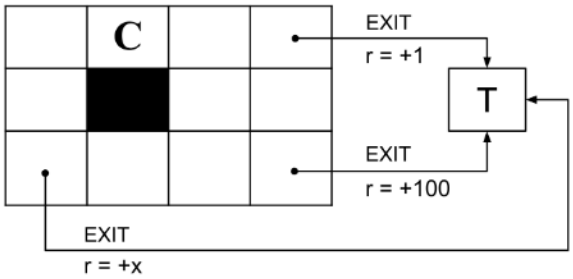
$V^*(A)$: $V^*(B)$:

(ii) After how many iterations k will we have $V_k(s) = V^*(s)$ for all states s? If it never occurs, write "never". Write your answer in the given box.

(iii) Suppose that we wanted to re-design the reward function. For which of the following new reward functions would the optimal policy **remain unchanged**? Let $R(s, a, s')$ be the original reward function.

- ☐ $R_1(s, a, s') = 10R(s, a, s')$
- ☐ $R_2(s, a, s') = 1 + R(s, a, s')$
- ☐ $R_3(s, a, s') = R(s, a, s')^2$
- ☐ $R_4(s, a, s') = -1$
- ☐ None

(c) For the following problem, we add a new state in which we can take the $EXIT$ action with a reward of $+x$.



(i) For what values of x is it *guaranteed* that our optimal policy π^* has $\pi^*(C) = \leftarrow$? Write ∞ and $-\infty$ if there is no upper or lower bound, respectively. Write the upper and lower bounds in each respective box.

$\square < x < \square$

(ii) For what values of x does value iteration take the **minimum** number of iterations k to converge to V^* for all states? Write ∞ and $-\infty$ if there is no upper or lower bound, respectively. Write the upper and lower bounds in each respective box.

$\square \leq x \leq \square$

(iii) Fill the box with value k, the **minimum** number of iterations until V_k has converged to V^* for all states.

$\square$