Introduction to Artificial Intelligence
Problem Set 3 — Reinforcement Learning

Instructions

In this problem set we're going to run through value iteration and Q-Learning by hand.

Questions

For the next few questions, use the environment described in figure

1. Run Value Iteration *by hand* for 10 iterations on the given environment. Make sure you note down the utility at each iteration separately to keep things organized.
2. Write some code to automate the utility updates for this environment. Make sure your code agrees with the work you did by hand in the previous question.
3. How long does it take for the utility to converge? Pick a threshold for small differences and see how long it takes until none of the utility values for any of the states is changing more than this threshold.
4. Run Q-Learning by hand on the same environment using a learning rate of $\alpha = 0.5$, an initial Q-table of all zeros, and the following experience traces (given as (s, a, s', r) tuples):
   (a) $(s_2, Up, s_1, -0.04)$
   (b) $(s_1, Right, s_4, 1.0)$
   (c) $(s_2, Right, s_3, -0.04)$
   (d) $(s_3, Up, s_2, -0.04)$
   (e) $(s_2, Up, s_1, -0.04)$
   (f) $(s_1, Right, s_4, 1.0)$
5. Assuming that the world resets after the agent visits state s_4, and the agent starts in state s_2, do these experience traces suggest that this is a greedy agent? Why or why not?
6. If a greedy agent were being used to generate experience traces for Q-Learning in this environment, would we be guaranteed to visit every state (in the limit)? What single aspect of the environment could be changed to flip your answer (yes to no, no to yes)?

Submission

Write up your answers to the given questions as a single PDF file, and put any code into a single python file.

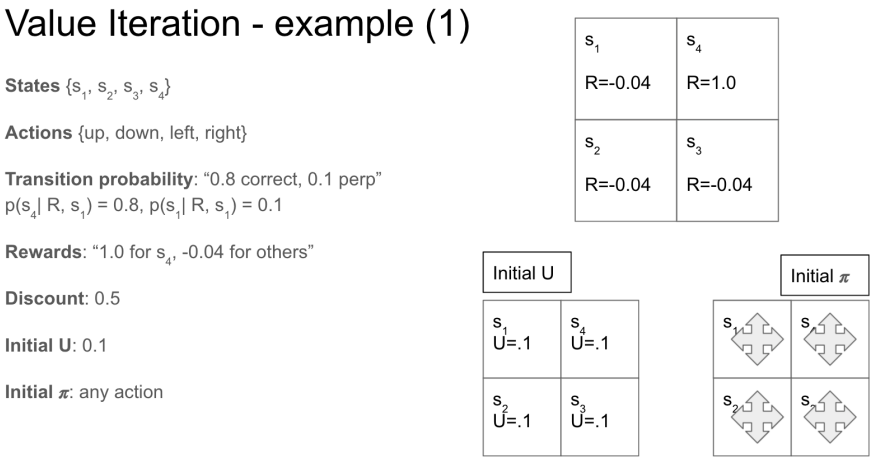

Figure 1: A tiny gridworld.