Actions

Homework 4

Problem

Policy Iteration

Note: Submission details to come soon.

Policy iteration (PI) is perhaps the most under appreciated algorithm for solving MDPs. Although each iteration is expensive, it generally requires very few iterations to find an optimal policy. In this problem, you'll gain an appreciation for how hard it is to get policy iteration to break a sweat.

For this assignment, construct an MDP with 30 states and at most 2 actions per state. You may assume the discount factor is 3/4. The MDP may have stochasticity in its transitions.

You should specify the transition function and reward function using four 30 x 30 arrays in the following way: $cell(i, j)$ in...

1. the first array should specify the probability of transitioning from state i to state j under the first action.
2. the second array should specify the probability of transitioning from state i to state j under the second action.
3. the third array should specify the reward for transitioning from state i to state j under the first action.
4. the fourth array should specify the reward for transitioning from state i to state j under the second action.

Currently it is not known whether there is an MDP which requires more than a linear number of PI iterations in the number of states of the MDP. Your goal is to create an MDP that attains at least 15 iterations of PI before the algorithm terminates. (If you're able to obtain more than 31 iterations,

please send a self-addressed, stamped envelope to Box 1910, Computer Science Department, Brown University, 115 Waterman St., Providence, RI 02912 for a smiley-face sticker.)