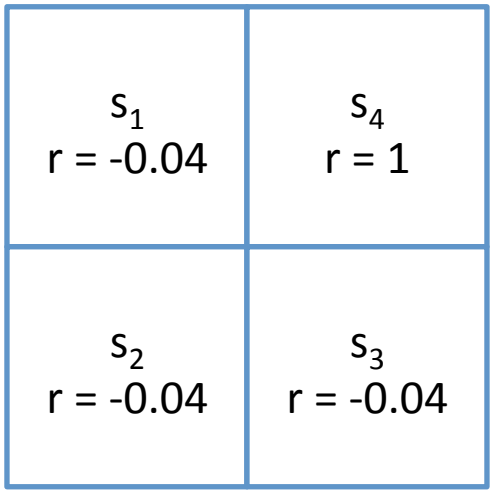
CS 3600: Markov Decision Process problem

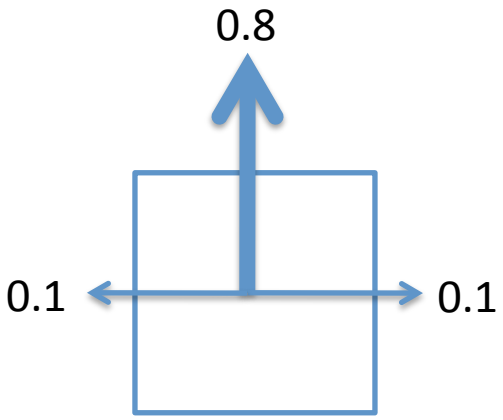
Given the following Markov Decision Problem, use Value Iteration to find the policy after two iterations.

There are four states, s1, s2, s3, and s4, arranged in a grid. State s4 is a sink (absorbing) state. The immediate rewards are given above. The agent can move UP, DOWN, LEFT, RIGHT and the transition model is such that there is an 80% chance of a correct move, and a 10% chance of moving to either side in error (e.g., if performing UP, there is a 10% chance of performing LEFT instead and a 10% chance of performing RIGHT instead).

Let the initial utility values for states are shown below.

- $U_0(s_1) = 0.1$
- $U_0(s_2) = 0.1$
- $U_0(s_3) = 0.1$
- $U_0(s_4) = 1.0$
- $\gamma = 0.5$





$U_1(s_1) = R(s_1) + \gamma \max_a \{$

up: (0.8)(0.1) + (0.1)(0.1) + (0.1)(1) ←0.19

down: (0.8)(0.1) + (0.1)(0.1) + (0.1)(1) ←0.19

left: (0.8)(0.1) + (0.1)(0.1) + (0.1)(0.1) ←0.1

right: (0.8)(1) + (0.1)(0.1) + (0.1)(0.1) ←0.82

$U_1(s_1) = -0.04 + (0.5)(0.82)$

$U_1(s_1) = 0.37$

$\pi_1(s_1) = \text{Right}$

$U_1(s_2) = R(s_2) + \gamma \max_a \{$

up: (0.8)(0.1) + (0.1)(0.1) + (0.1)(0.1) ←0.1

down: (0.8)(0.1) + (0.1)(0.1) + (0.1)(0.1) ←0.1

left: (0.8)(0.1) + (0.1)(0.1) + (0.1)(0.1) ←0.1

right: (0.8)(0.1) + (0.1)(0.1) + (0.1)(0.1) ←0.1

$U_1(s_2) = -0.04 + (0.5)(0.1)$

$U_1(s_2) = 0.01$

$\pi_1(s_2) = \text{any}$

$U_1(s_3) = R(s_3) + \gamma \max_a \{$

up: (0.8)(1) + (0.1)(0.1) + (0.1)(0.1) ←0.82

down: (0.8)(0.1) + (0.1)(0.1) + (0.1)(0.1) ←0.1

left: (0.8)(0.1) + (0.1)(1) + (0.1)(0.1) ←0.19

right: (0.8)(0.1) + (0.1)(1) + (0.1)(0.1) ←0.19

$U_1(s_3) = -0.04 + (0.5)(0.82)$

$U_1(s_3) = 0.37$

$\pi_1(s_3) = \text{Up}$

$U_2(s_2) = R(s_2) + \gamma \max_a \{$

up: (0.8)(0.37) + (0.1)(0.01) + (0.1)(0.37) ←0.334

down: (0.8)(0.01) + (0.1)(0.37) + (0.1)(0.01) ←0.046

left: (0.8)(0.01) + (0.1)(0.37) + (0.1)(0.01) ←0.046

right: (0.8)(0.37) + (0.1)(0.01) + (0.1)(0.37) ←0.334

$U_2(s_2) = -0.04 + (0.5)(0.334)$

$U_2(s_2) = 0.127$

$\pi_2(s_2) =$ up or right

MDP

- $U_1(s_1) = 0.37$
- $\pi_1(s_1) = \text{right}$

- $U_1(s_2) = 0.01$
- $\pi_1(s_2) = \text{any}$

- $U_1(s_3) = 0.37$
- $\pi_1(s_3) = \text{up}$

- $U_2(s_1) = 0.379$
- $\pi_2(s_1) = \text{right}$

- $U_2(s_2) = 0.127$
- $\pi_2(s_2) = \text{up or right}$

- $U_2(s_3) = 0.379$
- $\pi_2(s_3) = \text{up}$