ENSTA-ParisTech, ROB311 – 24/09/2018

Problem:

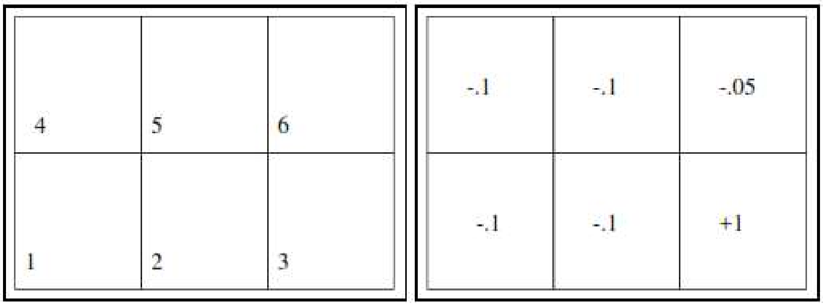
Pretend an agent is trying to plan how to act in a 3x2 world. This world is similar to the Russel and Norvig example of chapter 17 (“AI - A Modern Approach”). Figures 1(a) show the world, the rewards associated with each state, and the dynamics.

There are 5 possible actions: north, east, south, west and stay still. The first 4 actions succeed with probability 0.9 and go to a right or left angle of the desired direction with probability .05(see Figure 2) for an illustration of this. The fifth action, “do nothing” succeeds with probability 1.

The rewards associated with each state are

R(1 : 6) = [-0.1, -0.1, + 1, -0.1, -0.1, - 0.05] and are also shown in Figure 1(b).

State 3 is the only terminal state.



(a) States of 3x2 World

(b) Rewards of 3x2 World

Figure 1: 3x2 World Specification

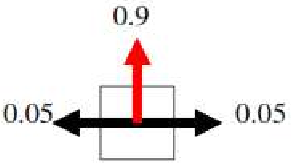


Figure 2: Transition model given a desired movement in red

1. Implement the value iteration for this problem. Initialize your utility vector to be 0 for all the states. Use γ = 0.999 as your discount factor.

2. How many iterations does it take the utilities converge?

3. The value of γ greatly influences the best policy. Change γ to 0.1 and again compute the converged state utilities and the best policy given these utilities. What is this policy? What's one interesting thing about the difference between this policy and the policy for using γ = 0.999?