A(E)3M33UI — Exercise11:

Markov Decision Processes

Martin Macaš

2015

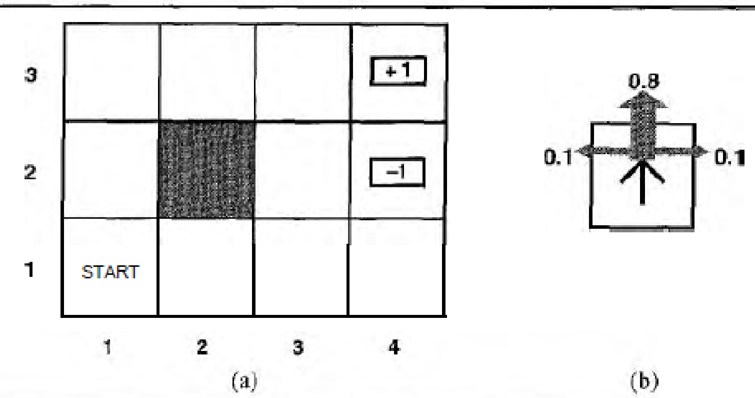

Figure 17.1 (a) A simple 4 x **3** environment that presents the agent with a sequential decision problem. (b) Illustration of the transition model of the environment: the "intended" outcome occurs with probability 0.8, but with probability 0.2 the agent moves at right angles to the intended direction. A collision with a wall results in no movement. The two terminal states have reward +1 and −1, respectively, and all other states have a reward of −0.04.

Exercise 1 For the 4 x **3** world shown in figure, calculate which squares can be reached from (1,l) by the action sequence *[Up, Up, Right]* and with what probabilities.

Exercise 2 Consider an undiscounted MDP having three states 1, 2, 3 with rewards -1, -2, 0, respectively. State 3 is a terminal state. In states 1 and 2 there are two possible actions: *a* and *b*. The transition model is as follows:

- In state 1, action *a* moves the agent to state 2 with probability 0.8 and makes the agent stay put with probability 0.2.
- In state 2, action *a* moves the agent to state 1 with probability 0.8 and makes the agent stay put with probability 0.2.
- In either state 1 or state 2, action *b* moves the agent to state 3 with probability 0.1 and makes the agent stay put with probability 0.9.

Questions:

a. What can be determined *qualitatively* about the optimal policy in states 1 and 2?
b. Apply three steps of value iteration, showing each step in full.
c. Which action will you chose for state 2 after the third iteration and why?