Reinforcement Learning Lecture: Homework 02

Ngo Anh Vien

MLR, University of Stuttgart

April 14, 2016

1 Exercise 1

You are asked to get familiar with the course's code base, download and test (*http://ipvs.informatik.uni-stuttgart.de/mlr/16-RL/16-RL.tgz*).

- Have a look at examples/Core/array/main.cpp with many examples on how to use the array class. Report on problems with installation.

2 Exercise 2

[Programming] Given the problem setting as in Figure 1. There are two terminal states with the rewards of +100 and -100 (the reward function is $R(s')$). There is a non-accessible state labeled with black color. Assuming that there are 4 possible actions: move-left, move-right, move-down, move-up. The movement rewards are c. The dynamics is **stochastic** with probability of 0.8 that brings the agent to its intended direction, and 0.2 randomly uniform probability to its perpendicular directions.

The code of this exercise can be found in *examples/Ex02/*. This code consists of: 1) parsing code to return the transition and reward functions T, R of the above problem given a map file; 2) code structure for different questions in this exercise; 3) displaying code to test your results.

- write a program using the iteration iteration algorithm to find an optimal policy. Set $c = 0.0, \gamma = 1.0$.
- write a program using the policy iteration algorithm (this requires two parts: policy evaluation and policy improvement) to find an optimal policy. Set $c = 0.0, \gamma = 1.0$.

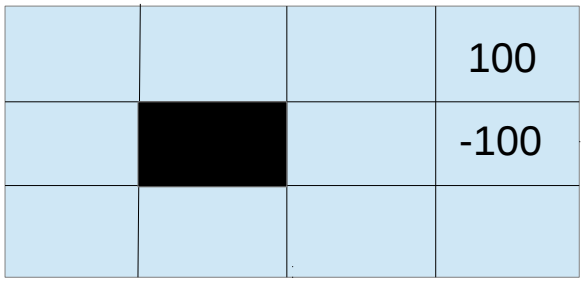

Figure 1: MDP for exercise 1

- rerun the above code with a different setting $c = -200.0, \gamma = 1.0$. Do you receive a different optimal policy? Explain your results!

3 Exercise 2

(Bonus) Prove the contraction mapping property in slide 26.