1 State-Action Value Function and Policy Iteration

(a) Do exercise 4.10 from Sutton and Barto’s textbook, page 67. A link to the pdf version of the book was provided on the last exercise sheet.

(b) (From the last exercise sheet) Show that modified policy iteration is equivalent to value iteration. Modified policy iteration is the same as policy iteration but with doing only one step of policy evaluation.

2 Policy Iteration

Use the codebase from last week’s exercise (gridworld3.zip) on value iteration. Short descriptions of all the files contained in the archive and an instruction for how to run the code can be found in exercise sheet 4.

(a) Create a new class PolicyIterationAgent in agent.py and implement policy iteration. You can use the ValueIterationAgent class as a template for the new class. To select the policy iteration agent in the command line interface using the agent switch (-a or --agent), you have to edit gridworld.py and add an appropriate option after line 393.

(b) How many rounds of policy iteration are needed before the start state of MazeGrid becomes non-zero?

(c) How many iterations of policy iteration do we need for the algorithm to converge to a optimal policy?

(d) Compared with value iteration, what are the advantages and disadvantages of policy iteration? Give a detailed list of the pros and cons.