

Reinforcement Learning: Coursework Assignment 1 (Semester 2, 2013)

Instructions

- This homework assignment is to be done *individually*, without help from your classmates or others. Plagiarism will be dealt with strictly as per University policy.
- Solve all problems and provide your complete solutions (with adequate reasoning behind each step) in a report in computer-printed or *legibly* handwritten form.
- For computational questions, include your code (e.g., Matlab commands) and all major numerical parameters involved.
- Before you start to write a program, read all questions below carefully.
- Only the report will be marked. The code itself will not be marked, but will be used to clarify any questions arising from the report.
- Use graphical representations wherever suitable. If you use numerical output for demonstrating your results, make sure that the numbers are appropriately rounded and presented in an accessible way. If your problem involves randomness, try to obtain a result that is representative for most of the possible realisations of the underlying random effects.
- This assignment will count for 10% of your final course mark.
- Please submit your assignment by 4 pm on 6th March as a paper copy to ITO as well as an electronic version via the submit system (including code).

Questions

Design and build a learning agent that operates in a grid world, specified in figure 1. The goal of the robot is on the top right hand corner, where the agent should expect to receive a reward. The grid world includes barriers which prevent certain actions from being used in corresponding cells. The objective of learning is to find policies that take the agent to the goal from all possible initial states. In this problem, you may assume the following:

State: you know what square the agent is occupying at any given time.

Actions: N, S, E, W.

State transitions: Initially, deterministic and can't pass through walls. Agent remains in the same square if it attempts a disallowed action. In a later section, we'll modify this further.

Rewards: -1 for each time step, 0 for attaining the goal.

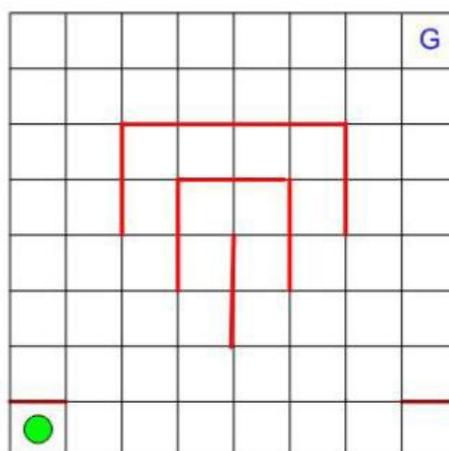


Figure 1: Grid world for Question 1. The green dot is one instance of an initial state and G denotes the goal cell.

