Introduction to Multi-Agent-Programming

B. Nebel, A. Kleiner
C. Dornhege, D. Zhang
Winter Semester 2010/2011

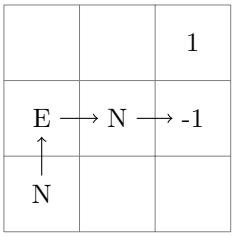
University of Freiburg
Department of Computer Science

Exercise Sheet 11

Due: January 25th, 2011

Exercise 11.1 (**Q-Learning (3pt)**)

Consider the following grid world, where the numbers are rewards associated with the cells. The numbered cells are terminal states, unnumbered cells have a reward of -0.1. An agent starts at the left-bottom corner. It can perform four possible actions: North, South, East and West.



The agent executed the actions marked in each cell and the executed trajectory is marked by the arrows. Let $\gamma = 0.9$ and $\alpha = 0.5$ and the Q-function be initialized to 0 for all $s \in S, a \in A$. Show the updates that Q-learning performed in each step along the trajectory.

Exercise 11.2 (**Optimal Policy (2pt)**)

Consider the gridworld above. Assume the rewards of unnumbered cells are 0. How will the optimal policy in each cell look?

How will this change if the rewards are larger negative numbers, e.g. -5?