UNIVERSITÄT BONN
RHEINISCHE FRIEDRICH-WILHELMS-UNIVERSITÄT BONN
COMPUTER SCIENCE VI AUTONOMOUS INTELLIGENT SYSTEMS
Prof. Dr. Sven Behnke
Endenicher Allee 19a

Robot Learning

Assignment 4

Due Tuesday, May 7th, before class.

4.1) Consider the following grid world:

			W	W	G	
				W		W
	W	W		W		
S		W			W	
		W				

The agent starts in state S.
When it reaches state G, it will receive a reward of 1000 and the episode ends.
When it reaches a state W, it will receive a reward of -100 and the episode ends.
Every other step will be rewarded with -1.
The agent has eight actions: It can move to an adjacent cell (according to eight-neighborhood).
The actions are not deterministic. Only with probability 0.6, the desired action is carried out. With probability 0.2, the agent deviates from the desired direction by one cell to the left or to the right (axis-parallel moves become diagonal moves and vice versa).
Actions that would move the agent off the grid are handled by truncating the resulting cell coordinates to valid grid coordinates. Hence, almost all diagonal moves outside the grid (except for moves into cells extending corners) will result in a horizontal or vertical move.
Randomly generate 1000 episodes according to the policy:
Move to the right with probability 0.5. Move up with probability 0.25, move down with probability 0.25.
Compute the state-value V(s) for each visited state using TD(0) Policy Evaluation.

10 points

4.2) Starting from the above policy, choose actions using ε-greedy action selection (ε=0.1).
Improve the policy by Q-learning for 10,000 episodes.
Compute the resulting state value V(s) for each grid cell.
Visualize the resulting policy using arrows.

10 points