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

Robot Learning

Assignment 3

Due Tuesday, April 30th, before class.

3.1) Consider the following grid world:

*	*	*		*	*		*	*
					X			*
	X	X	X	X	X	X		
			*	*	*		X	
		X	X	X	X	X		
							X	
	X	X	X	X	X	X	G	
	*	*	*	X	*	*		X
	*	*	*		*	*		X

The agent may start in any cell.
It can choose between four actions: moving one cell up, moving one cell down, moving one cell left, and moving one cell right.
When it reaches cell G, it will receive 100 points and the episode ends.
When it reaches a cell marked *, it will receive 5 points and the episode continues.
When it attempts to enter a cell marked X, it will receive -20 points and stay in the cell it came from.
When it attempts to leave the grid, it will receive -5 points and stay where it is.
All actions entering an unmarked cell will receive -1 point.

Compute the expected value of all cells for a policy that chooses with probability 0.5 a random action and otherwise moves down.
The discount parameter shall be γ=0.9.

5 points

3.2) Use the Policy Iteration algorithm to compute the optimal value $V^*(s)$ for each cell.
Indicate the resulting optimal policy $\pi^*(s)$ with arrows in each cell.

5 points

3.3) Extend the action set by also allowing diagonal moves, such that the agent can move to its eight neighboring cells.
Use the Value Iteration algorithm to compute the optimal value $V^*(s)$ for each cell.
Indicate the resulting optimal policy $\pi^*(s)$ with arrows in each cell.
Discuss how value and policy changed, compared to 3.2)!

5 points

3.4) Consider non-deterministic actions, where the agent moves with probability 0.7 into the desired direction, but with probability 0.15 deviates 45° to the left and with probability 0.15 deviates 45° to the right of the desired direction. Compute again $V^*(s)$ and indicate $\pi^*(s)$ with arrows.
Discuss how value and policy changed, compared to 3.3) !

5 points