Lehrstuhl für STEUERUNGS- UND REGELUNGSTECHNIK Technische Universität München Prof. Dongheui Lee	**MACHINE LEARNING IN ROBOTICS** Exercises 11: RL

Exercise 1

Consider a robot that lives in a grid world. The robot can execute only four actions i.e. move to North, South, East or West. The G corresponds to a terminal goal state and reaching goal state yields a reward of $+10$. Similarly D corresponds to a ditch (another terminal state) that should be avoided and reaching ditch yields a negative reward of -20. X represent blocked states that are not accessible. If the robot executes an action that tries to get to the blocked state or out of the grid then it stays at the same position. All other states yield a negative reward of -1 (for encouraging the robot to reach the goal state as quickly as possible). The model of robot is stochastic and with probability 0.8, it moves in the commanded direction but with probability of 0.2 it either moves left or right (0.1 for each).

Using a discount factor of $\gamma = 0.99$, write few steps of value iteration using asynchronous update and use it to get the optimal policy $\pi^*(s)$.

G			D
X	X		
X	X		

Exercise 2

Solve the same problem but now with Policy Iteration.