Assignment 6: Reinforcement learning

Submit as group to prof4155@cs.dal.ca with subject line A6 Thursday, November 1, 4pm.

a. Modify the program the chain example from the policy iteration to value iteration and on-policy TD learning and compare these methods. Evaluate how these methods converge, that is, how many iterations are necessary on average. Attach the programs to your email submission.

b. Train the Lego tribot to move forward close a wall without running into it. You can hardcode that the tribot should not move more than a certain distance from the wall, but the robot should learn to move close to the wall but not bumping into it. Thus, the robot should receive a large reward close to the wall but when the touch sensor is pressed it should receive a large negative reward (punishment).

Please submit a paragraph describing your experience.