3/4/17
580-03, Spring 2017, Programming Project 3

In this homework, you will implement a temporal difference learning. Showing that this method works will give you 7 points.

I recommend you use Open AI Gym, as discussed in class because Gabe is very familiar with this platform. However, if you would like to use a different framework (e.g., BURLAP, RL-Glue, RL-PI, etc.) you can. We just won't be able to help as much.

Recall that homework is graded on a 10-point scale. Here is a guide to how we will grade your homework. To get a 7/10, you only need to successfully implement a TD method (i.e., Q-Learning, Sarsa, TD(0), or TD(1) and show that it works.

Ideas for making your report more interesting, with rough point values:

+1.5 point: Implement eligibility traces and show how different values of lambda change the speed of learning.
+1 point: Compare on-policy vs. off-policy learning (e.g., Sarsa and Q-Learning). Which works better in this domain?
+0.5 point: Compare different values of alpha and epsilon and see how this affects learning.
+0.5 point: Use a different task or modify the task provided to show that your method works in stochastic domains.
+0.5 points: How does changing the rewards and/or the amount of randomness in the domain change the speed of learning?