CSYE 7374
Autonomous Learning in Games
Assignment 4 – Deep Reinforcement Learning
Open Gym

Professor: Nik Bear Brown
Due: **Sunday, April 5, 2020**

TAs

Dikshant Rathi <rathi.d@husky.neu.edu>

Deep Reinforcement Learning

In this assignment you will use deep reinforcement learning and Open AI to play games see http://gym.openai.com/envs/#classic_control and https://github.com/openai/gym

We will start with the http://gym.openai.com/envs/#toy_text or http://gym.openai.com/envs/#classic_control examples

You will be using more advanced reinforcement learning like deep reinforcement learning algorithms, (if you wish you can substitute Proximal Policy Optimization (PPO) and Soft Actor-Critic (SAC))

Part 1 30 Points

Use one of the toy text http://gym.openai.com/envs/#toy_text or http://gym.openai.com/envs/#classic_control examples

Implement some form of deep reinforcement learning to play the game or

Answer the following questions for all of the:

* Establish a baseline performance. How well did your RL Q-learning do on your problem?
* What are the states, the actions and the size of the Q-table?
* What are the rewards? What did you choose them?
* How did you choose alpha and gamma in the following equation?

$$newQ(s,a) = Q(s,a) + \alpha[R(s,a) + \gamma[maxQ'(s',a') - Q(s,a)]]$$

Try at least one additional value for alpha and gamma. How did it change the baseline performance?
* Try a policy other than *maxQ(s', a').* How did it change the baseline performance?

* How did you choose your decay rate and starting epsilon? Try at least one additional value for epsilon and the decay rate. How did it change the baseline performance? What is the value of epsilon when if you reach the max steps per episode?
* What is the average number of steps taken per episode?
* Does Q-learning use value-based or policy-based iteration?
* What is meant by expected lifetime value in the Bellman equation?

Part 2 40 Points

You must explain the CNN hyperparameters you used and show the effect on the performance of at least one important hyperparameter.

Part 3 Professionalism 30 Points

Did I explain my idea clearly? (5 Points)
How effective are you at explaining what you are doing? You MUST write an abstract and a conclusion.

Did I explain my evaluation clearly? (5 Points)
Just saying "accuracy" is not a clear explanation of an evaluation scheme. Clearly explain the evaluation scheme. Do the metrics make sense?

It MUST run. (5 Points)
The code must run on a laptop other than yours. There MUST be a clear README on how to run it.

What code is yours and what have you adapted and licensing*? (5 Points)*
You must explain what code you wrote and what you have done that is different. Failure to cite ANY code will result in a zero for this section. Did I explain my licensing clearly? Failure to cite a clear license will result in a zero for this section.

Did I explain my code clearly? (10 Points) Your code review score will be scaled to a range of 0 to 10 and be used for this score.