RL Project Proposal

Airi Shimamura Khoi Trinh

For our RL project this semester; we want to build an RL agent that can easily beat the CartPole game.

CartPole v1 is a game environment provided in the Gym package from OpenAI. In this environment, there is a pole, attached to a cart (hence the name CartPole). The cart moves along a frictionless track. Force is applied in the left and right direction of the cart. The goal of the game is to keep the pole upright for as long as possible. For each step taken, a +1 reward is given, including the termination step. The maximum points achievable in the game is 475. There are a few conditions that, if met, will end the game.

First, if the pole angle is greater than 12 degrees, the game ends.

Second, if the cart position is greater than 2.4 (or the center of the cart reaches either end of the display), the game ends.

Finally, if episode length is greater than 500, the game ends.

Airi will implement Q-learning, and Khoi will implement TD learning. We plan to train the cart and get a better Q-table to improve the performance of the agent. Before building the Q-table, we will first set up the environment. There will be four factors, which are cart position, cart velocity, pole angle, and the velocity of the pole at the tip. The cart takes only two actions: moving right or moving left, and it will receive +1 as a reward for every successful step, but it will receive -1 as a penalization if an episode ending criteria is met. As mentioned above, an episode ends when the angle of the pole is greater than 12 degrees, the cart position is more than 2.4 units from the center, or episode length is greater than 500.

The goal for this game is to minimize penalties and get a score as close to 200 as possible. However, if the cart only takes a good action, the range the cart moves is going to be very limited, so we will use the epsilon greedy approach to solve this issue. This approach allows the agent to take random actions at a certain rate, which means the cart can be trained more. Then based on the actions taken, and rewards, get maximum q values and update Q-table using TD learning, and repeat these steps for the number of episodes.

For the purpose of this project, we will end the training if the total score is over 195 for 100 episodes in a row. To visualize the performance of the agent, we will record the values of rewards and total steps until the pole falls over for each episode and then make a bar plot of steps taken in each episode, and a line graph of reward given per episode. Based on these plots, we will find the maximum rewarded scores, and see if the performance increases over time, or if there is room for more training. Addtionally, we will get the average of the total rewarded scores to see the overall performance of the agent.

We plan on doing research and looking at existing CartPole training tutorial (for example, here `https://tinyurl.com/ycknpsen`) to help get started, as we are not too familiar with reinforcement learning. We will make a reference to all of these codes and note any changes made.