COMP-767: Reinforcement Learning - Assignment 3

Optional (can be turned in until May 6)

The assignment can be carried out individually or in teams of two.

1. **Policy gradient vs value-based methods [100 points]**

   In this task, you will compare value-based and actor-critic/policy gradient methods for the Mountain-Car task of the classical control suite of the Gym environment:

   https://gym.openai.com/envs/MountainCar-v0/

   You will compare Q-learning and actor-critic, and asynchronous versions of these algorithms with 4 parallel threads. For the feature representation, please discretize the position and velocity of the car into 32 bins each. Descriptions of all the algorithms are available in the following paper: https://arxiv.org/pdf/1602.01783.pdf

   You should use the same set of 3 different learning rates for each algorithm and run at least 5 independent runs in each case. Write a little report describing your experimental setup (choice of exploration, anything else important), include learning curves for each algorithm with all parameter settings, and a plot comparing the learning curves of each algorithm obtained with the best parameter setting. Explain what you conclude form these plots.