Foo Lin Geng 1002504

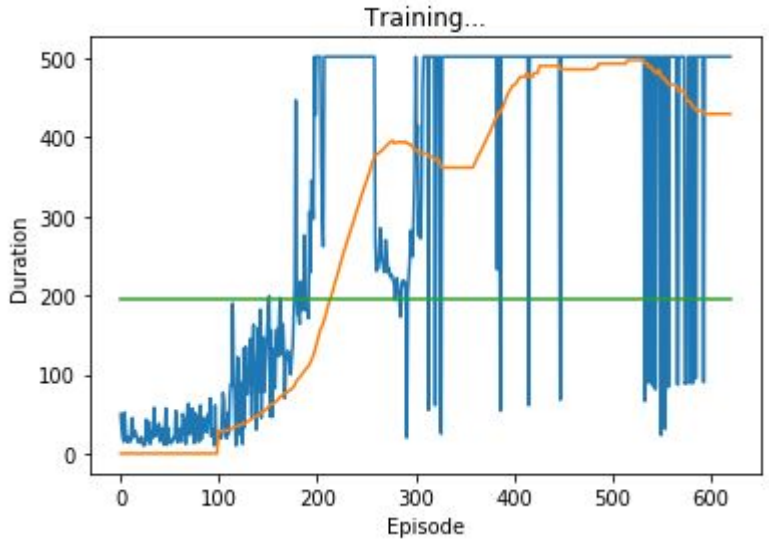
Coding Homework Week 9

Part 1: reinforce with baselines (Cartpole)

The full code can be found in "**week9codinghmk_part1.py**". I chose to use the neural network with two hidden layers (of output size 24).

The settings are:

- GAMMA = 0.99
- EPS = 0.05
- LEARNING_RATE = 0.005
- WEIGHT_DECAY = 0.000001



We observe that the network successfully mastered Cartpole after about 200 episodes.

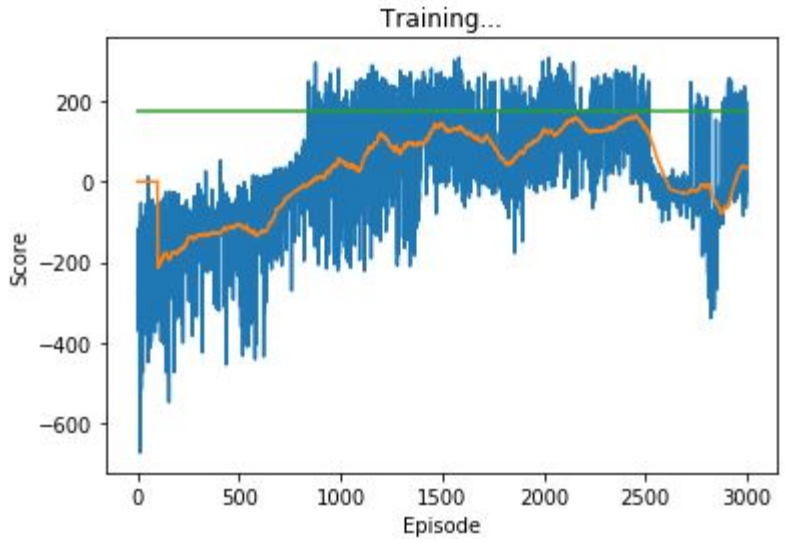
Implementing reinforce: Apart from recording the rewards and states, we also record the log probabilities of the actions for each episode. The state values are recorded to act as the baseline.

Part 2: reinforce with baselines (Lunar Lander)

The full code can be found in "**week9codinghmk_part2.py**". I chose to use the neural network with two hidden layers (of output size 24).

I chose to use similar settings as the Cartpole task above:

- GAMMA = 0.99
- EPS = 0.05
- LEARNING_RATE = 0.005
- WEIGHT_DECAY = 0.000001



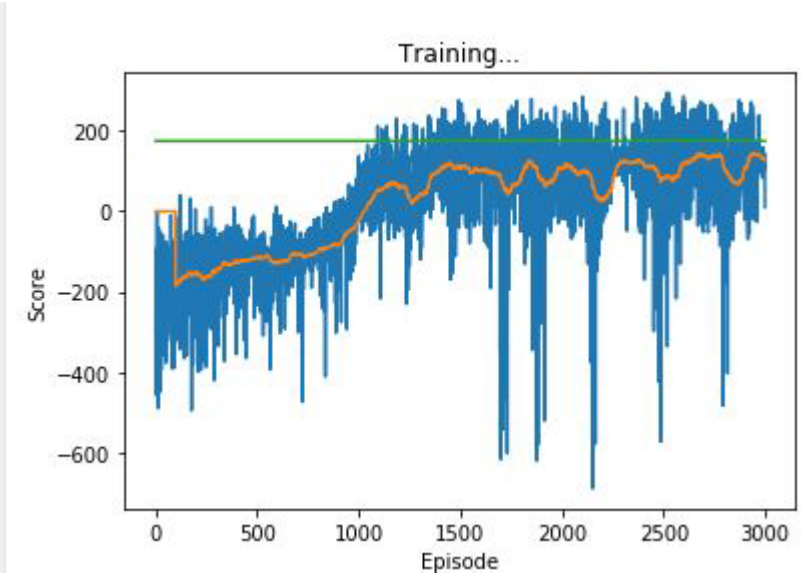
We observe that the agent under "reinforce with baselines" generally learns well, and almost "masters" the game at around episode 2100, with a average score of 164.0047.

Part 3: reinforce without baselines (Lunar Lander)

The full code can be found in "**week9codinghmk_part3.py**". I chose to use the neural network with two hidden layers (of output size 24).

I chose to use similar settings as the Part 2 task above:

- GAMMA = 0.99
- EPS = 0.05
- LEARNING_RATE = 0.005
- WEIGHT_DECAY = 0.000001



The highest "average value reached" is 144.93, quite a bit lower than in Part 2. We also observe that there are dips (periods which perform very badly), which did not happen when baselines were used.