# 10-703 Deep Reinforcement Learning and Control
# Assignment 3
# Fall 2018

ielshar
mharding

October 26, 2018

## Problem 1: Reinforce

1. Describe your implementation:
   - Neural Network Architecture: same as given in JSON file.
   - Learning rate: 0.001
   - Discount factor $\gamma$ : 1

2.

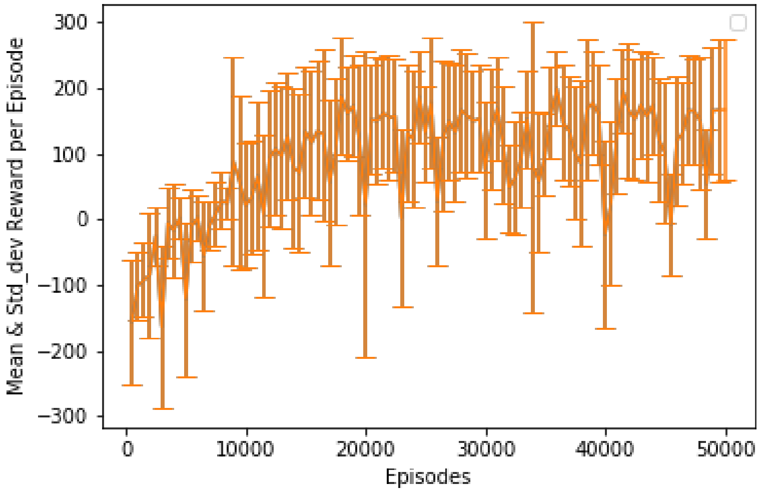

Figure 1: Reinforce Algorithm: learning curve - Every k=500 episodes the current policy is tested on 100 episodes. The plot shows the mean and standard deviation of each of this tests.

Figure 1 above shows that our agent was able to achieve a mean reward of 200 or slightly more or less at several points (after 19000, 36000, 41500 episodes) through out the training. Since no baseline was used in our Reinforce algorithm one can see that the test results have a high variance. This is somehow expected though as the total return at the end of episode varies highly from one episode to another.

# Problem 2: Advantage-Actor Critic

1. Implementation details:

Table 1: A2C Implementation

| A2C | | |
|---|---|---|
| N | Settings | value |
| 1 | Actor NN Architecture | Same as given in JSON file |
| | Critiic NN Architecture | MLP with 3 layers each with 20 hidden units and relu activation |
| | Actor learning rate | 0.001 |
| | Critic learning rate | 0.001 |
| 20 | Actor NN Architecture | Same as given in JSON file |
| | Critiic NN Architecture | MLP with 3 layers each with 20 hidden units and relu activation |
| | Actor learning rate | 0.001 |
| | Critic learning rate | 0.001 |
| 50 | Actor NN Architecture | Same as given in JSON file |
| | Critiic NN Architecture | MLP with 3 layers each with 20 hidden units and relu activation |
| | Actor learning rate | 0.001 |
| | Critic learning rate | 0.001 |
| 100 | Actor NN Architecture | Same as given in JSON file |
| | Critiic NN Architecture | MLP with 3 layers each with 20 hidden units and relu activation |
| | Actor learning rate | 0.001 |
| | Critic learning rate | 0.001 |

2. Plots:

   - N=20

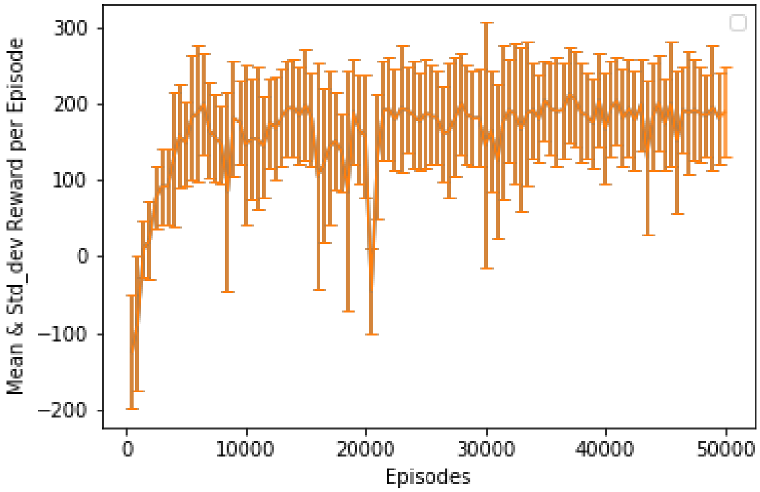

Figure 2: A2C Algorithm N=20: learning curve - Every k=500 episodes the current policy is tested on 100 episodes. The plot shows the mean and standard deviation of each of this tests.

- N=50

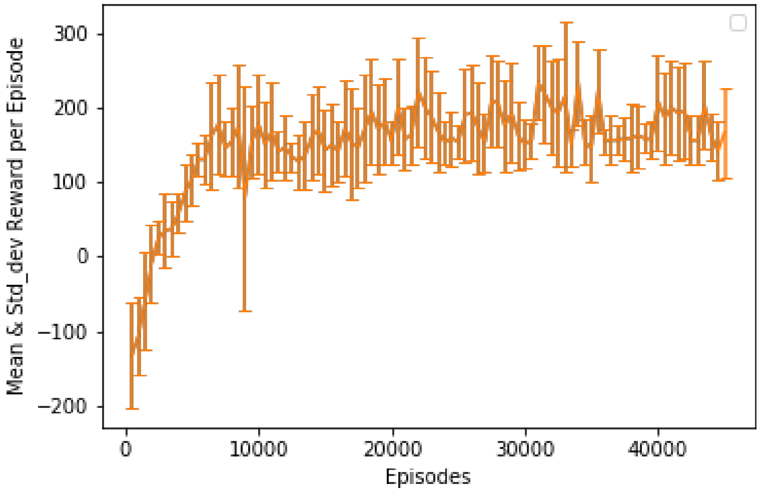

Figure 3: A2C Algorithm N=50: learning curve - Every k=500 episodes the current policy is tested on 100 episodes. The plot shows the mean and standard deviation of each of this tests.

- N=100

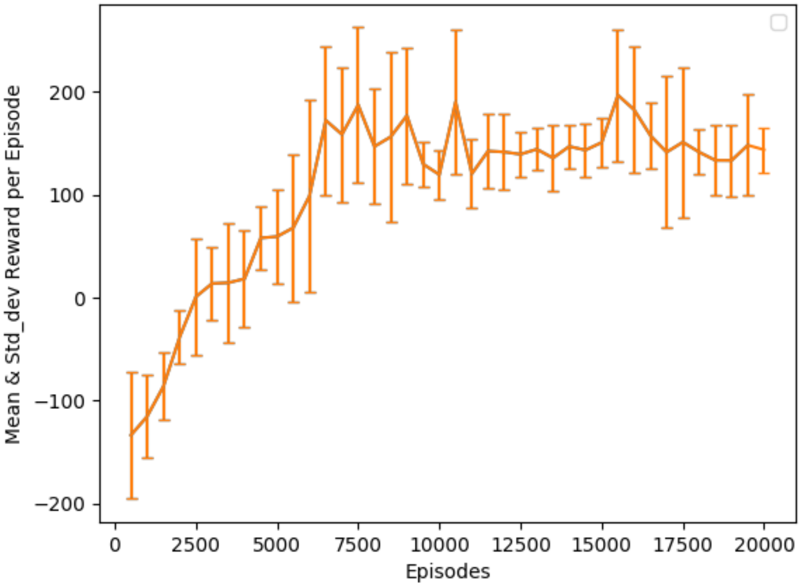

Figure 4: A2C Algorithm N=100: learning curve - Every k=500 episodes the current policy is tested on 100 episodes. The plot shows the mean and standard deviation of each of this tests.

3. Reinforce and A2C comparisons:
   Compared to Reinforce, n-steps A2Cs have much lower variance and accelerated learning. This is due to the critic which replaces G in reinforce.