

Assignment 2: Policy Gradients

1 Vanilla Policy Gradients

1.1 The learning curves

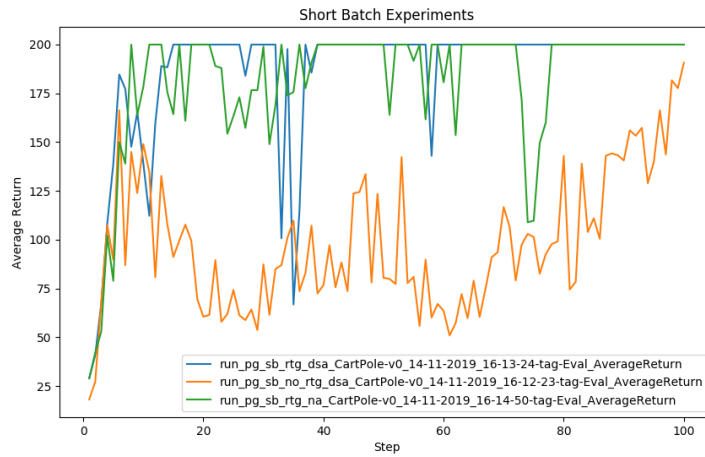


Figure. 1: The learning curves of the small batch experiments

From Figure 1 we can see that in the short batch experiments, using reward-to-go and standardizing the advantages can improve the performance of the model.

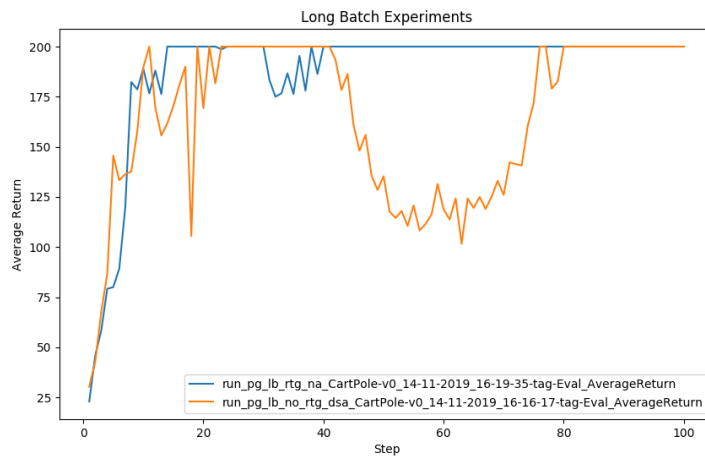


Figure. 2: The learning curves of the long batch experiments

From Figure 2 we can see that in the long batch experiments, using reward-to-go helps reduce the variance of our model.

1.2 Questions

1. The one using reward-to-go has better performance without advantage-standardization.
2. Yes, advantage standardization helps reduce the variance, making the learning curves smoother.
3. Yes, small batch size leads to slow convergence.

1.3 Command line configuration

```
python cs285/scripts/run_hw2_policy_gradient.py --env_name CartPole-v0 -n
    100 -b 1000 -dsa --exp_name sb_no_rtg_dsa
python cs285/scripts/run_hw2_policy_gradient.py --env_name CartPole-v0 -n
    100 -b 1000 -rtg -dsa --exp_name sb_rtg_dsa
python cs285/scripts/run_hw2_policy_gradient.py --env_name CartPole-v0 -n
    100 -b 1000 -rtg --exp_name sb_rtg_na
python cs285/scripts/run_hw2_policy_gradient.py --env_name CartPole-v0 -n
    100 -b 5000 -dsa --exp_name lb_no_rtg_dsa
python cs285/scripts/run_hw2_policy_gradient.py --env_name CartPole-v0 -n
    100 -b 5000 -rtg -dsa --exp_name lb_rtg_dsa
python cs285/scripts/run_hw2_policy_gradient.py --env_name CartPole-v0 -n
    100 -b 5000 -rtg --exp_name lb_rtg_na
```

2 Implementing Neural Network Baseline

2.1 Lunar Lander

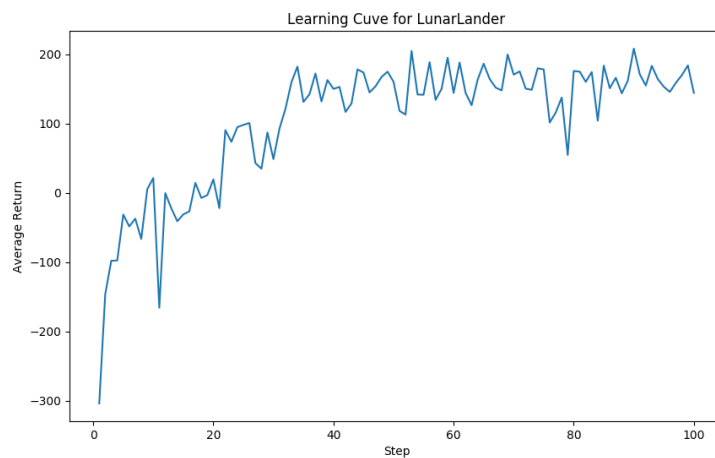


Figure. 3: The learning curve in task "LunarLander"

2.2 Half Cheetah

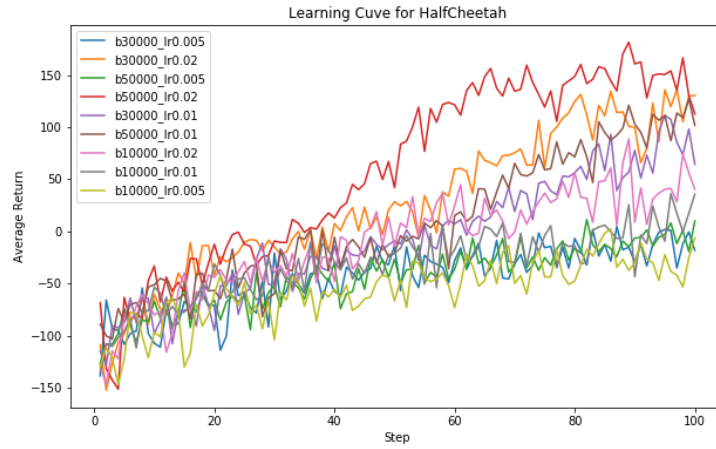


Figure. 4: The learning curve in task "HalfCheetah"

From the above figure we know that the higher the learning rate is, the higher average return will be. Also, the larger the batch size is, the higher average return will be.

2.3 Half Cheetah 2

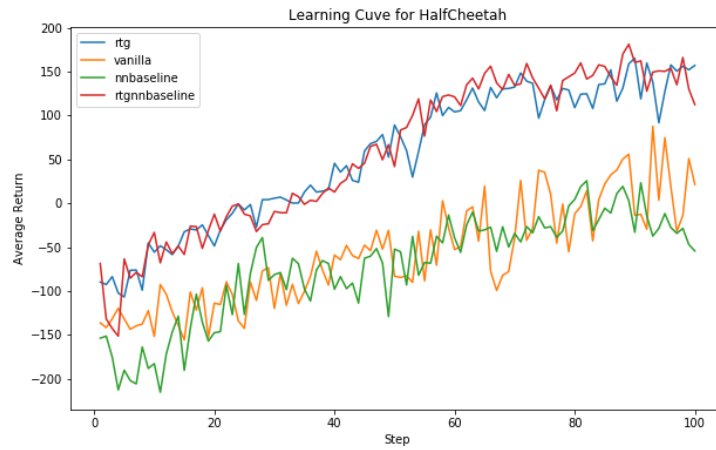


Figure. 5: The learning curve in task "HalfCheetah" using learning rate and batch size selected above