Deep Reinforcement Learning Homework 2
Policy Gradients

Abdelkader Benamara | Mohamed Ali Benrekia | Aymen Djelid

February 8, 2022

1 Experiment 1 (CartPole)

Here are our results :

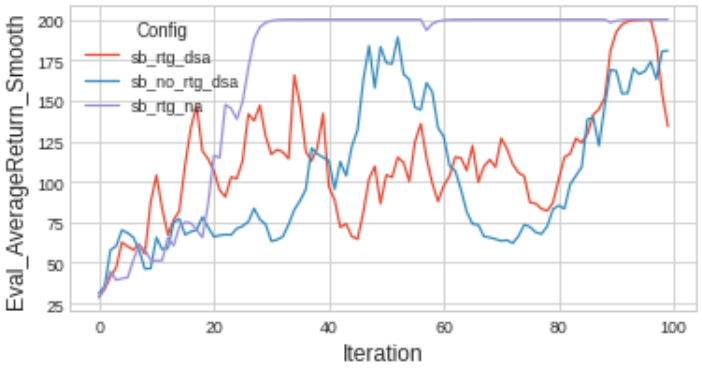

Figure 1
Small Batch Experiments (CartPole Environment)

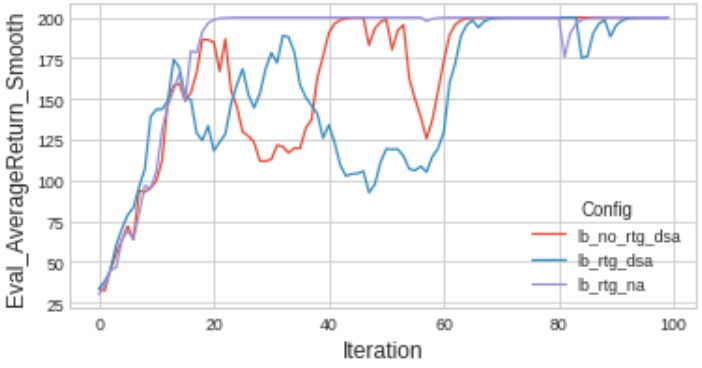

Figure 2
Large Batch Experiments (CartPole Environment)

1. **Reward to go** gives better results without advantage- standardization.
2. Did advantage- standardization helped ? **YES**.
3. Larger batch sizes has better performance.

Using the following command :

```
>>> python path_to_run_hw2.py --env_name CartPole-v0 -n 100 -b 1000 -dsa
        --exp_name q1_sb_no_rtg_dsa
```

2 Experiment 2 (InvertedPendulum)

Here are our results :

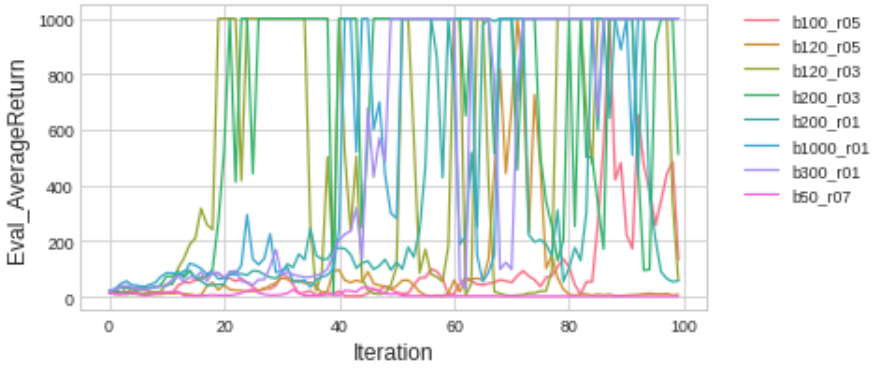

Figure 3
InvertedPendulum Environment Experiments

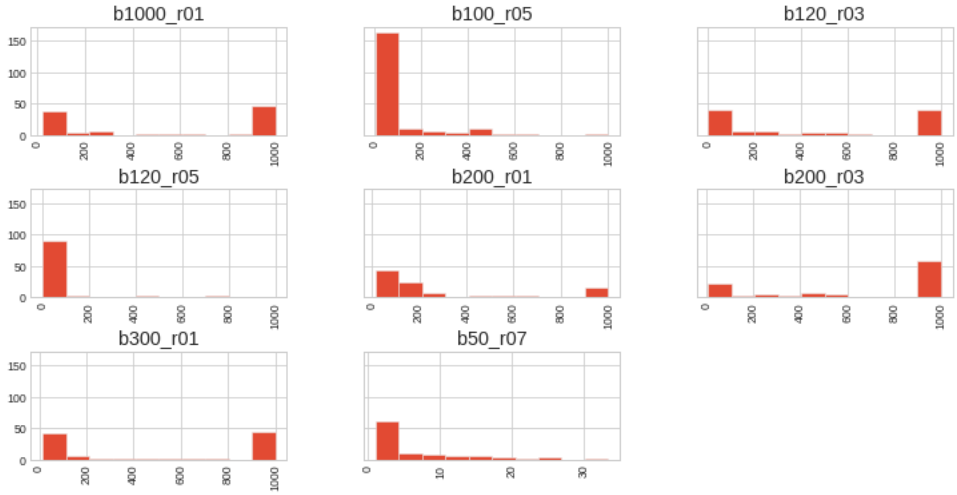

Figure 4
Best parameters for InvertedPendulum Environment

Note. With the help of this graphs we can take the best hyperparameters : **batch-size : 200** and **learning-rate : 0.03**

Using the following command :

```
>>> python path_to_run_hw2.py --env_name InvertedPendulum-v2 --ep_len 1000
          --discount 0.9 -n 100 -l 2 -s 64 -b <b*> -lr <r*> -rtg --exp_name q2_b<b*>_r<r*>
```

3 Experiment 3 (LunarLander)

Here are our results :

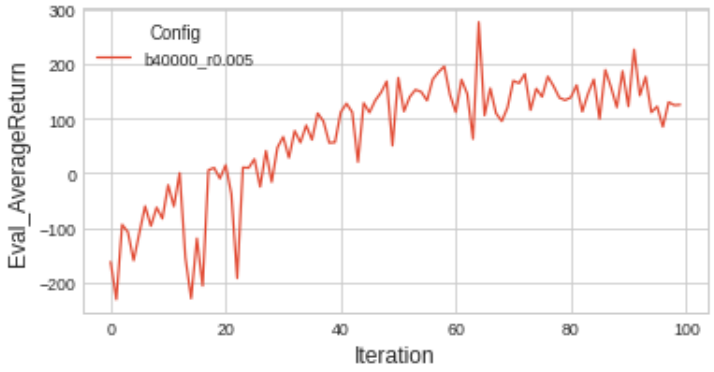

Figure 5
LunarLander Environment Experiments

Using the following command :

```
>>> python path_to_run_hw2.py --env_name LunarLanderContinuous-v2
          --ep_len 1000 --discount 0.99 -n 100 -l 2 -s 64 -b 40000
          -lr 0.005 --reward_to_go --nn_baseline --exp_name q3_b40000_r0.005
```

4 Experiment 4 (HalfCheetah)

Here are our results :

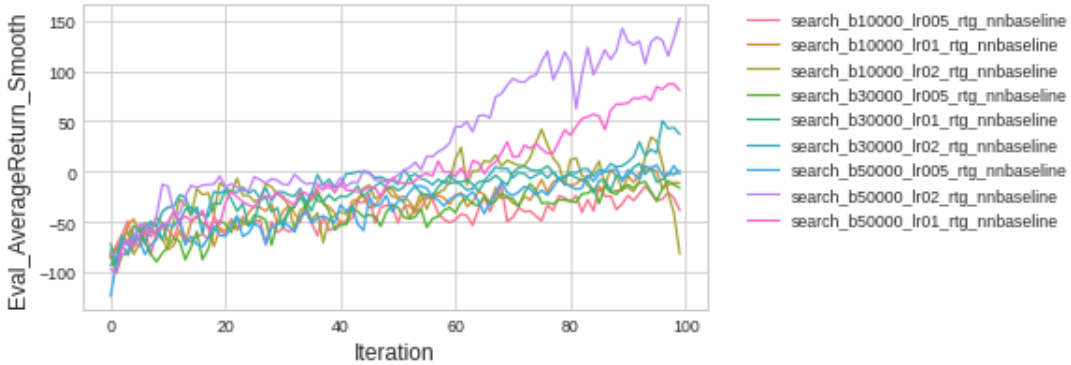

Figure 6
HalfCheetah Environment Experiments Hyperparameter Search

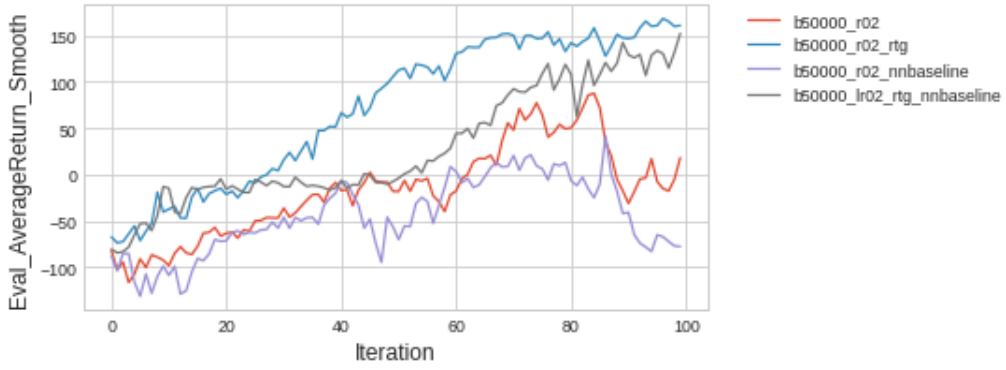

Figure 7
HalfCheetah Environment Experiments Optimal params

Note. We notice that larger batch size and higher learning rate give better results. And for the best combination we got **size : 50k** and **lr : 0.02**.

Using the following command :

```
>>> python path_to_run_hw2.py --env_name HalfCheetah-v2 --ep_len 150
          --discount 0.95 -n 100 -l 2 -s 32 -b 10000 -lr 0.005 -rtg
          --nn_baseline --exp_name q4_search_b10000_lr0.005_rtg_nnbaseline
```

Note. Hyperparameters need to be changed regarding other configurations.

5 Experiment 5 (Hopper)

Here are our results :

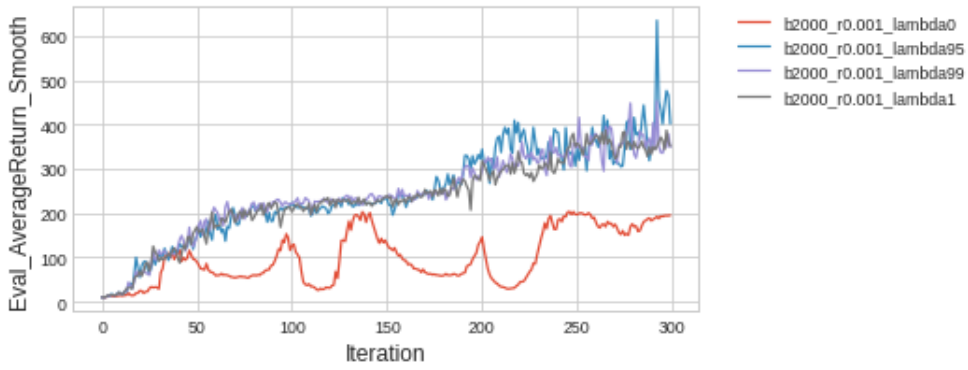

Figure 8
Hopper Environment Experiments Hyperparameter Search

Note. The best performance is achieved for $\lambda = 0.95$ and it under-performs for $\lambda = 0$

```
>>>     python path_to_run_hw2.py --env_name Hopper-v2 --ep_len 1000
            --discount 0.99 -n 300 -l 2 -s 32 -b 2000 -lr 0.001
            --reward_to_go --nn_baseline --action_noise_std 0.5 --gae_lambda
            --exp_name q5_b2000_r0.001_lambda<$lambda$>
```

Note. Hyperparameters need to be changed regarding other configurations.