Machine Learning HW12

ML TAs

ntu-ml-2021spring-ta@googlegroups.com

HW Content

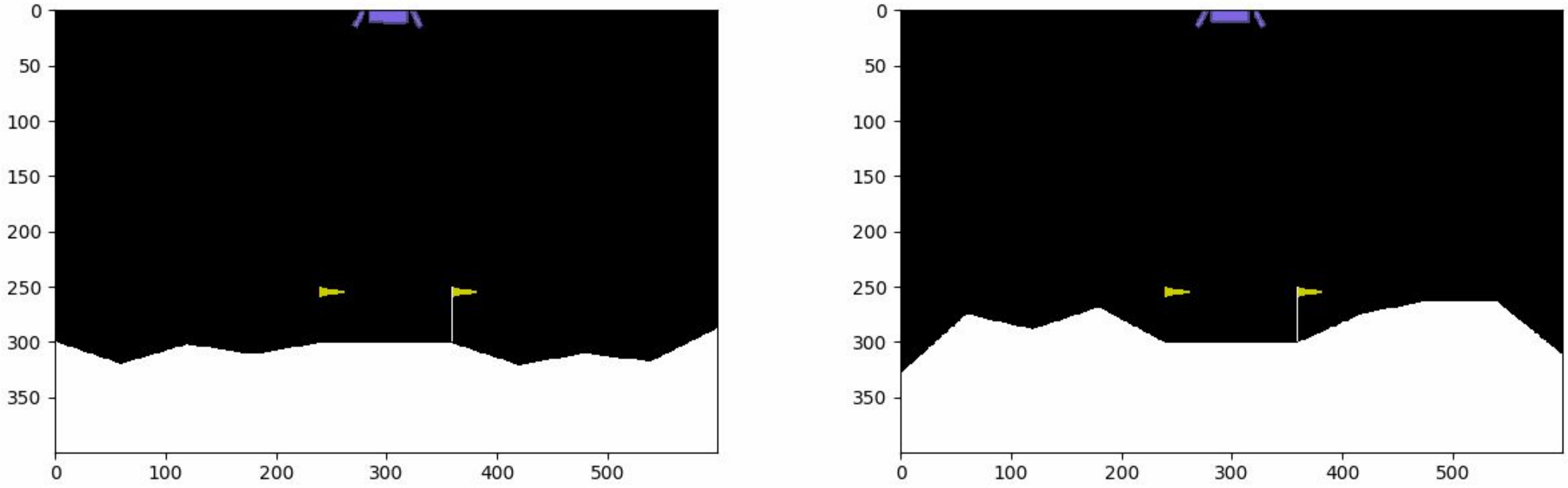
In this HomeWork, you can implement some Deep Reinforcement Learning methods by yourself :

- Policy Gradient
- Actor-Critic (Implement by yourself to get high score !)

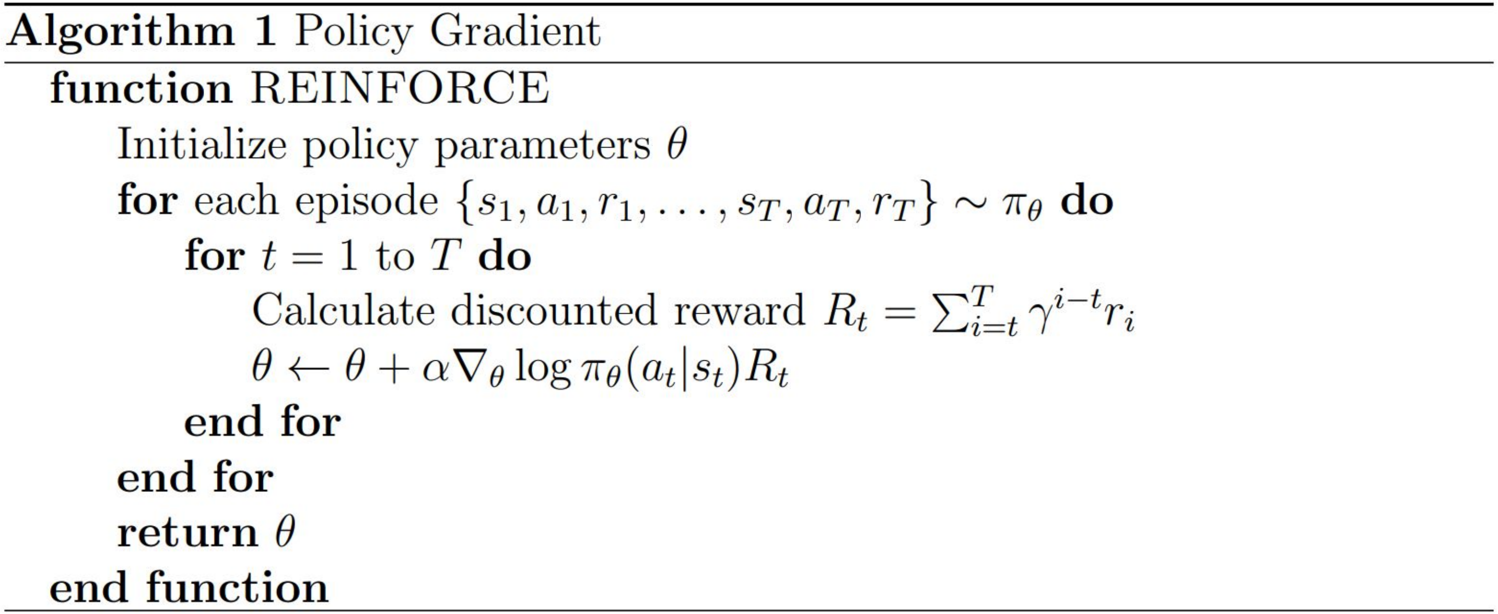
The environment of this HW is **Lunar Lander** in gym of OpenAI.

Other details can be found in the sample code.

Illustraion



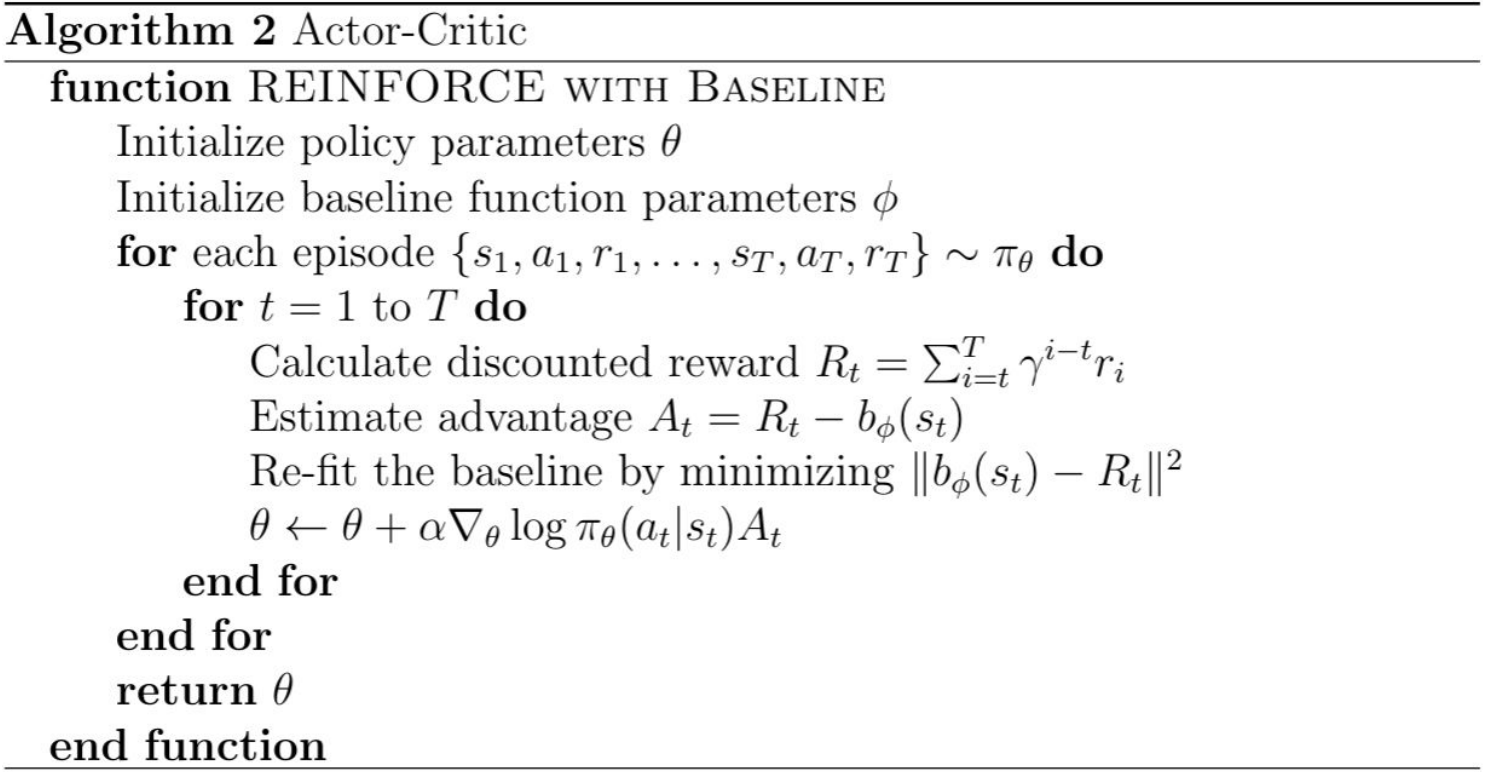
Policy Gradient(to get 8 points)



Algorithm 1 Policy Gradient

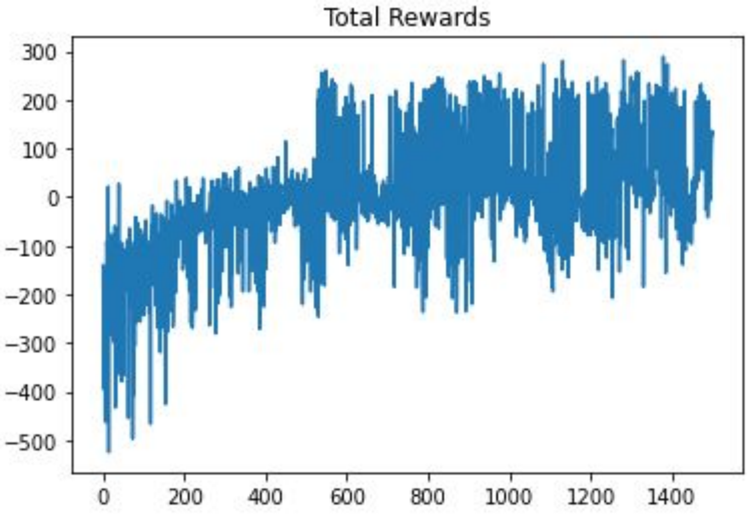
```
function REINFORCE
    Initialize policy parameters $\theta$
    for each episode $\{s_1, a_1, r_1, \ldots, s_T, a_T, r_T\} \sim \pi_\theta$ do
        for $t = 1$ to $T$ do
            Calculate discounted reward $R_t = \sum_{i=t}^{T} \gamma^{i-t} r_i$
            $\theta \leftarrow \theta + \alpha \nabla_\theta \log \pi_\theta(a_t|s_t) R_t$
        end for
    end for
    return $\theta$
end function
```

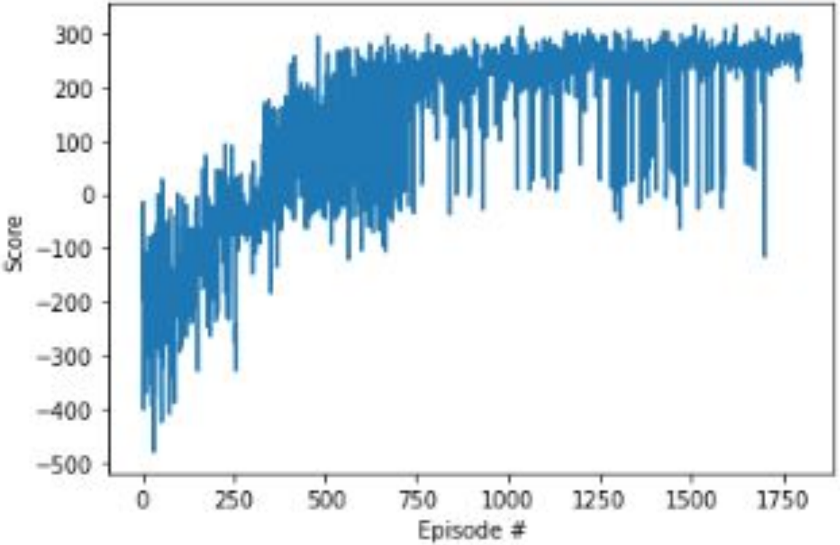
Actor-Critic(to get 10 points)



```
Algorithm 2 Actor-Critic
function REINFORCE WITH BASELINE
    Initialize policy parameters $\theta$
    Initialize baseline function parameters $\phi$
    for each episode $\{s_1, a_1, r_1, \ldots, s_T, a_T, r_T\} \sim \pi_\theta$ do
        for $t = 1$ to $T$ do
            Calculate discounted reward $R_t = \sum_{i=t}^{T} \gamma^{i-t} r_i$
            Estimate advantage $A_t = R_t - b_\phi(s_t)$
            Re-fit the baseline by minimizing $\|b_\phi(s_t) - R_t\|^2$
            $\theta \leftarrow \theta + \alpha \nabla_\theta \log \pi_\theta(a_t|s_t) A_t$
        end for
    end for
    return $\theta$
end function
```

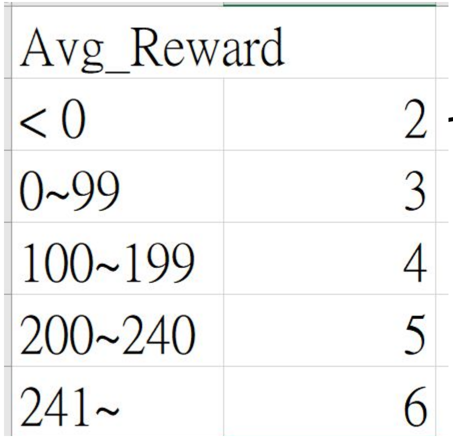
Sample Result





What you need to submit & Grading

1. Python file (Submit on NTU COOL) (**4 points**)
2. Action List (On JudgeBoi, **the highest one is automatically selected**)
3. Submission must be valid



Avg_Reward	
< 0	2
0~99	3
100~199	4
200~240	5
241~	6

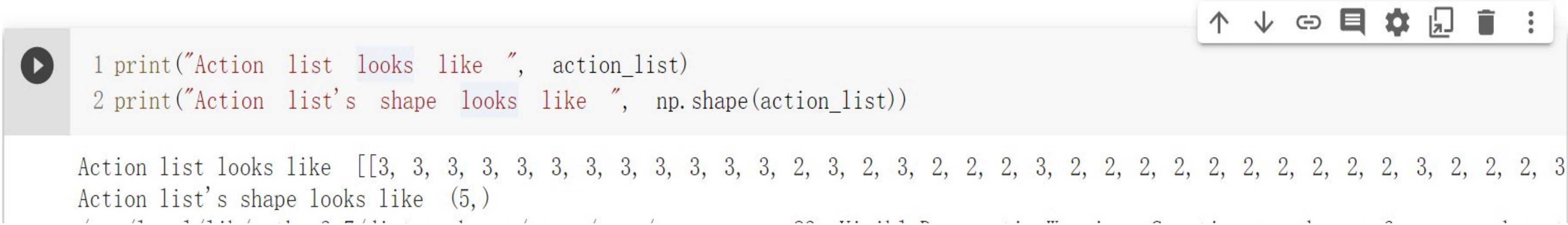
Task failed successfully.

What you need to submit & Grading

More on a "valid submission “:

Your agent should output done after the last input of your action list, action list with mismatched length will be rejected。

Action list 的長相



```
print("Action list looks like ", action_list)
print("Action list's shape looks like ", np.shape(action_list))
```

```
Action list looks like  [[3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 2, 3, 2, 3, 2, 2, 2, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 2, 2, 2, 3
Action list's shape looks like  (5,)
```

Bonus

- If you successfully get 10 pts:
  - Your code will be made public to students.
  - You can submit a report in **PDF** format briefly describing what you have done (in English, less than 100 words) for **extra 0.5 pts**.
  - Reports will also be made public to students.

- Report template

Announcement

- You should finish your homework on your own.
- You should NOT modify your prediction files manually.
- Do NOT share codes or prediction files with any living creatures.
- Do NOT use any approaches to submit your results more than 5 times a day.
- **Do NOT search or use additional data or pre-trained models.**
- Your **final grade x 0.9** if you violate any of the above rules.
- Prof. Lee & TAs preserve the rights to change the rules & grades.

Announcement

- This HW will be graded by the score on JudgeBoi
- Any questions or concerns about HW can be post on NTU COOL(Recommend) or send email to ntu-ml-2021spring-ta@googlegroups.com . Please denote the subject of email by **[HW12]**

Submit Deadline: 6/04 - 6/25 (23:59)