

# 强化学习 Lab 2

## DQN

王若麟 2023/11/28





# Review: DQN

## A simple DQN Algorithm

1. take action  $a_i$  in ENV, then insert  $(s_i, a_i, s_{i+1}, r_i)$  in Experience Replay
2. if Experience Replay have enough  $\tau$ , then randomly sample  $(s_i, a_i, s_{i+1}, r_i)$  from Experience Replay
3.  $target(s_i) = r_i + \gamma \max_{a'} Q_{\phi^-}(s_{i+1}, a')$
4.  $\phi \leftarrow \phi + \alpha(target(s_i) - Q_{\phi}(s_i, a_i)) \frac{dQ_{\phi}}{d\phi}$
5. every N step, let  $\phi^- = \phi$
6. go to step 1



# Review: Double DQN

- ▶ 为何存在过估计问题？问题存在于下面划蓝线部分

$$target(s_t) = r_t + \gamma \max_{\underline{a'}} Q_{\phi^-}(s_{t+1}, a')$$

## 定理

如果我们有俩个随机变量  $X_1, X_2$

$$E[\max(X_1, X_2)] \geq \max(E[X_1], E[X_2])$$



# Review: Double DQN (cont.)

- ▶ DQN 有两个网络  $\phi$  和  $\phi^-$ ，所以使用如下方法降低估计问题：
  - ▶  $\phi$  用于选择最好的动作
  - ▶  $\phi^-$  用于估计最好的动作

## 1. DQN target

$$target(s_t) = r_t + \gamma Q_{\phi^-}(s_{t+1}, \arg \max_{a'} Q_{\phi^-}(s_{t+1}, a'))$$

## 2. Double DQN target

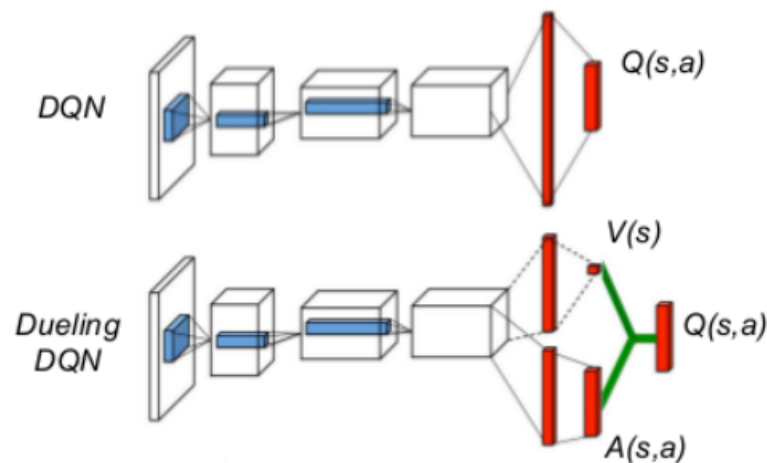
$$target(s_t) = r_t + \gamma Q_{\phi^-}(s_{t+1}, \arg \max_{a'} Q_{\phi}(s_{t+1}, a'))$$

# Review: Dueling DQN

- ▶ 如下图，DQN 的网络直接输出 Q 值；而 Dueling DQN 网络由如下公式确定

$$Q(s, a) = V(s) + A(s, a) - \frac{1}{|\mathbb{A}|} \sum_{a=1}^{\mathbb{A}} A(s, a)$$

$$Q(s, a; \theta, \alpha, \beta) = V(s; \theta, \beta) + \left( A(s, a; \theta, \alpha) - \frac{1}{|\mathbb{A}|} \sum_{a'} A(s, a'; \theta, \alpha) \right)$$





## 环境准备

- 安装anaconda/miniconda，配置镜像源
  - <https://mirror.tuna.tsinghua.edu.cn/help/anaconda/>
- pytorch ~= 1.12，根据自身情况选择GPU或CPU版本
  - <https://pytorch.org/get-started/previous-versions/>
  - 30系及以上的显卡，不要选择CUDA10.x的版本
- tensorboard
  - pip install tensorboard



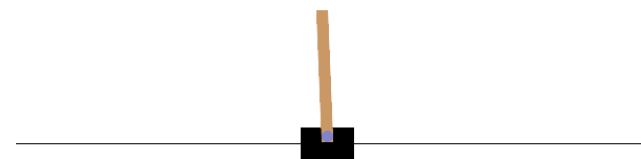
## 环境准备 (cont.)

- Gymnasium
  - <https://github.com/Farama-Foundation/Gymnasium>
  - 替换原来的Gym (Gym被OpenAI转手出去以后处于欠维护状态)
- 以下环境<sub>3</sub>选<sub>2</sub>



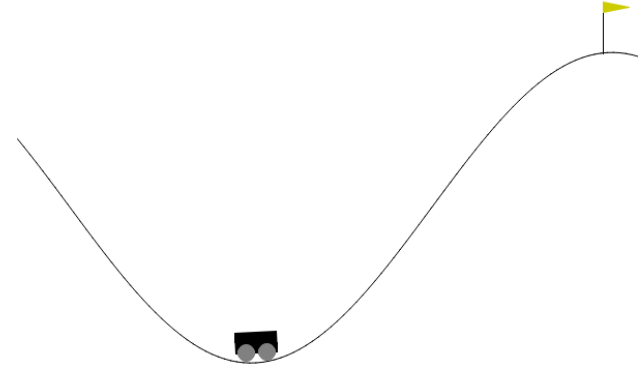
# 实验环境

- 环境1: CartPole-v1
- [https://gymnasium.farama.org/environments/classic\\_control/cart\\_pole/](https://gymnasium.farama.org/environments/classic_control/cart_pole/)
- 状态: Box(4)
- 动作: Discrete(2)
- `pip install gymnasium[classic-control]`



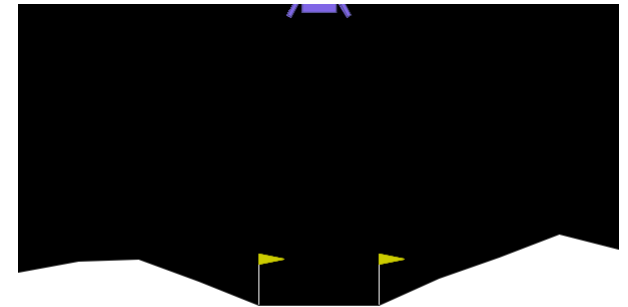


## 实验环境 (cont.)

- 环境2: MountainCar-v0
  - [https://gymnasium.farama.org/environments/classic\\_control/mountain\\_car/](https://gymnasium.farama.org/environments/classic_control/mountain_car/)
  - 状态: Box(2)
  - 动作: Discrete(3)
- 
- The diagram illustrates the MountainCar environment. It shows a black car with two wheels at the bottom of a U-shaped valley. A yellow flag is positioned at the top of the right-hand slope, representing the goal. The track is a smooth, continuous curve.
- `pip install gymnasium[classic-control]`

## 实验环境 (cont.)

- 环境3: LunarLander-v2
- <https://gymnasium.farama.org/environments/box2d/lunarlander/>
- 状态: Box(8)
- 动作: Discrete(4)
- `pip install gymnasium[box2d]`
- # 如果出现找不到swig的错误, 尝试用apt安装swig





# 实验要求

- 基于助教给出的代码，完善DQN算法的实现
  - 一共有3处TODO需要你补全
- 在此基础上，实现Double DQN (DDQN), Dueling DQN, Dueling DDQN，并对DQN和它们的表现进行比较
  - 绘制Reward曲线（4条：DQN, DDQN, Dueling DQN, Dueling DDQN）。为了更好的视觉效果，可以从Tensorboard中导出CSV，用seaborn等重绘
  - 进行简要的分析，包括收敛速度、最优性、稳定性等角度
  - 录制各方法最好策略的视频，10秒以内
    - ubuntu下可使用kazam
  - 不需要太过关注训练的分数



## 实验要求 (cont.)

- 加分项:
  - 实现Rainbow中其他改进手段, 并进行对比
  - Prioritized Replay
  - Multi-Step
  - Noisy-Net
  - ...
  - 可参考arXiv: 1710.02298 (<https://arxiv.org/pdf/1710.02298.pdf>)



# 实验提交

- 将代码压缩包、一份PDF格式的报告和视频压缩包提交到BB系统。
  - 如果文件太大，可上传至睿客网并在BB系统中留下链接
- DDL: 2023/12/31 23:59 UTC+8