RL Assignment 4

陈麒麟 517030910155

作业要求:

- 利用tensorflow和gym实现 **MountainCar-v0** 分别利用 **DQN** 和 **D(Double)DQN** 和 **Q-Learning** 进行学习的例子，并进行对比分析。

代码实现

详细信息参见 *相应名称文件夹*:

- 模型与训练在 **模型名称.py** 文件中，如 **DDQN.py**

```
python DDQN.py # 训练DDQN模型
```

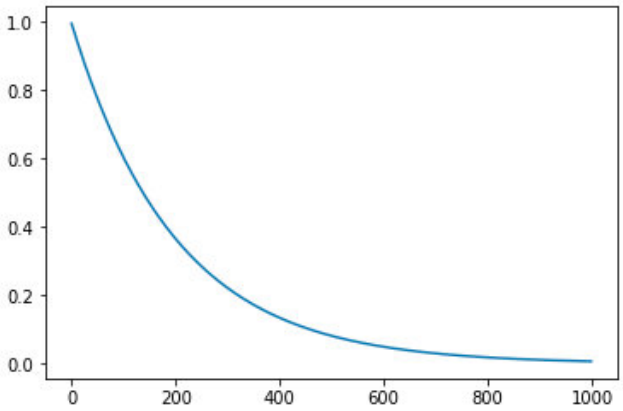
- 训练完成的模型保存为各文件夹 **.h5** 文件
- gym-demo 运行在 **模型名称demo.py** 文件中，如 **DDQNdemo.py**

```
python DDQNdemo.py # 运行DDQN gym-MountainCar-demo
```

测试

Tricks

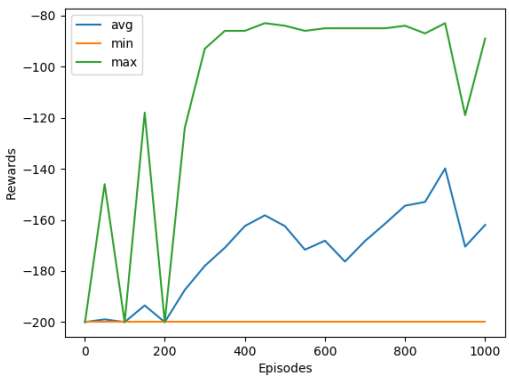
- 在实验中，对于DQN和DDQN我们采用了同一套网络结构 **16X16X16** 的全连接网络。
- epsilon采用了指数衰减，衰减率为 **0.995**，可以在settings更改。



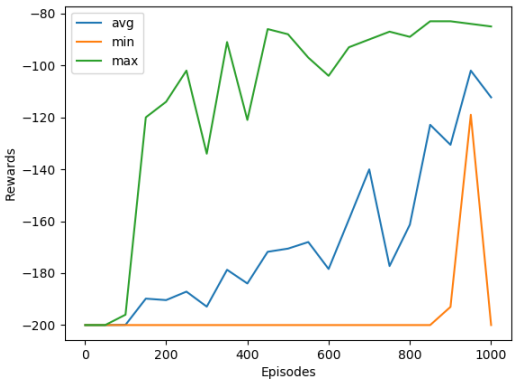
- 网络训练一个batch数据大小为 **64**。
- 记忆回放储存了 **10000** 个episode，回放下限是 **200**，每次以 **mini-batch** 为单位 **随机抽取**。

Reward曲线对比

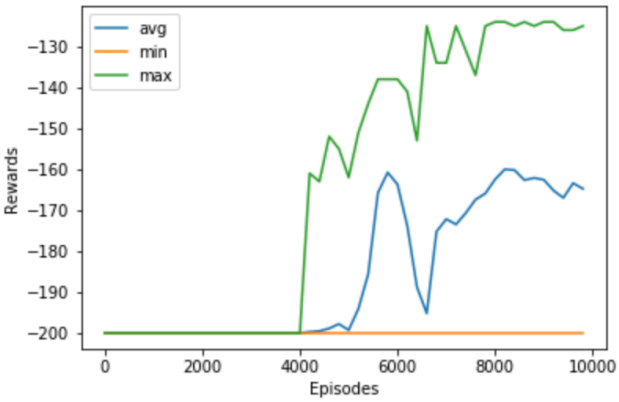
- 每轮更新的max-reward、average-reward、min-reward



DQN



DDQN

> 离散化的Q-learning:
>
> 
>

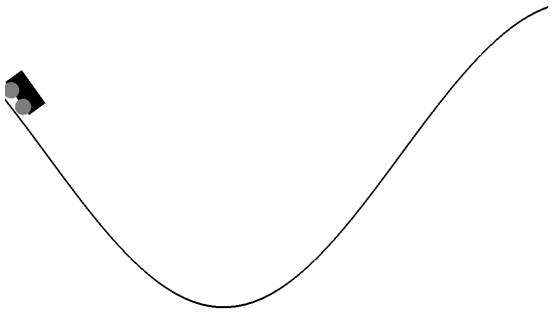
DQN和DDQN的对比：

> - DDQN有着更高更稳定的reward均值。
> - DQN在episodes少时有较大的reward波动，但是·收敛更快。

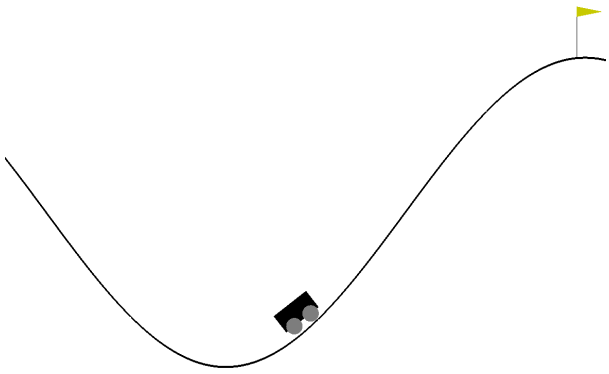
Network和Q-Learning对比：

> - Q-Learning无论是收敛速度还是reward均值都不如DQN。

- 智能体行为对比：



DQN会到达左边最高点



DDQN不会到达左边最高点

> 这里可以自行运行 **DQNdemo.py** 和 **DDQN.demo.py** 对比，训练好的模型附在相应文件夹下

DQN的智能体会在峡谷摆动2次才可以到达目的地，而且中途会到达左边最高点。

DDQN的智能体只在峡谷摆动1次就可以到达目的地，而且中途不会到达左边最高点。

分析

- DQN和DDQN的区别仅在于q-target的更新方式：

DQN在target表上进行最大Q选取更新：

$$Y_t^{\mathrm{Q}} \equiv R_{t+1} + \gamma \max_a Q(S_{t+1}, a; \boldsymbol{\theta}_t) .$$

DDQN在origin表上选取最大Q的action，以此在target表中选取Q更新：

$$Y_t^{\mathrm{Q}} = R_{t+1} + \gamma Q(S_{t+1}, \operatorname*{argmax}_a Q(S_{t+1}, a; \boldsymbol{\theta}_t); \boldsymbol{\theta}_t) .$$

- DQN在收敛性上的较DDQN优是因为DDQN的target更新由于收敛性被平均所以较慢。
- DDQN进行了更接近实际的估计，所以reward波动小。
- 从智能体行为上看，明显DDQN有着更好的智能体学习效果（仅在峡谷往返一次）。