**Homework 4:**

**Reinforcement Learning**

**Report**

**110652019林楷傑**

**Part I. Implementation (-5 if not explain in detail):**

* **Part1  
  一張含有 文字 的圖片

  自動產生的描述一張含有 文字 的圖片

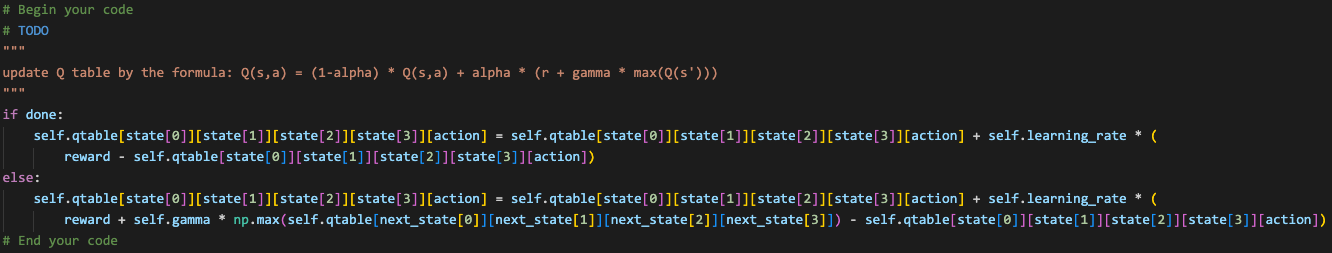
  自動產生的描述一張含有 文字 的圖片

  自動產生的描述**
* **Part2  
  一張含有 文字 的圖片

  自動產生的描述一張含有 文字 的圖片

  自動產生的描述一張含有 文字 的圖片

  自動產生的描述一張含有 文字 的圖片

  自動產生的描述一張含有 文字 的圖片

  自動產生的描述**
* **Part3一張含有 文字 的圖片

  自動產生的描述一張含有 文字 的圖片

  自動產生的描述一張含有 文字 的圖片

  自動產生的描述**

**Part II. Experiment Results:**

**Please paste taxi.png, cartpole.png, DQN.png and compare.png here.**

**1. taxi.png:一張含有 圖表 的圖片

自動產生的描述**

**2. cartpole.png一張含有 圖表 的圖片

自動產生的描述**

**3. DQN.png一張含有 圖表 的圖片

自動產生的描述**

**4. compare.png一張含有 圖表 的圖片

自動產生的描述**

**Part III. Question Answering (50%):**

1. Calculate the optimal Q-value of a given state in Taxi-v3, and compare with the Q-value you learned (Please screenshot the result of the “check\_max\_Q” function to show the Q-value you learned). **(10%)**一張含有 文字, 裝置, 儀表 的圖片

   自動產生的描述

一張含有 文字 的圖片

自動產生的描述

1. Calculate the max Q-value of the initial state in CartPole-v0, and compare with the Q-value you learned. (Please screenshot the result of the “check\_max\_Q” function to show the Q-value you learned) **(10%)**

一張含有 文字 的圖片

自動產生的描述  


1. Why do we need to discretize the observation in Part 2? **(3%)**

Since the observation of cartpole environment is continuous. We need to discretize the observation to present on the Q table.

1. How do you expect the performance will be if we increase “num\_bins”? **(3%)**

The run time of the program will be longer, but the performance will be better. Since the agent can gain more knowledge from the environment.

1. Is there any concern if we increase “num\_bins”? **(3%)**The runtime of the program would be longer.
2. Which model (DQN, discretized Q learning) performs better in Cartpole-v0, and what are the reasons? **(5%)**

DQN. The neural network can deal the continuous space. It solves the problem that discretized Q learning encountered, that is continuous space is hard to presented on the tabular method. Thus, DQN would perform better.

1. What is the purpose of using the epsilon greedy algorithm while choosing an action? **(3%)**

For exploration, use epsilon greedy algorithm to make the agent fully explore the evironment.

1. What will happen, if we don’t use the epsilon greedy algorithm in the CartPole-v0 environment? **(3%)**

The policy will stuck in a local optimal action since the agent will always choose the max Q which first update previously and is not alctually optimal. The agent is lack of exploration.

1. Is it possible to achieve the same performance without the epsilon greedy algorithm in the CartPole-v0 environment? Why or Why not? **(3%)**

No, without exploration, the agent can not fully explore actions that is likely to be better.

1. Why don’t we need the epsilon greedy algorithm during the testing section? **(3%)**

Since we have successfully trained out agent, the agent can perform well policy. We do not need randomness to disturb its decision.

1. Why does “with torch.no\_grad():“ do inside the “choose\_action” function in DQN? **(4%)**

with torch.no\_grad(): means the back propagation will not propagate through the computation. Since the decision making is depend on choosing the maximum Q value, we don’t need to tack the gradient of this process.