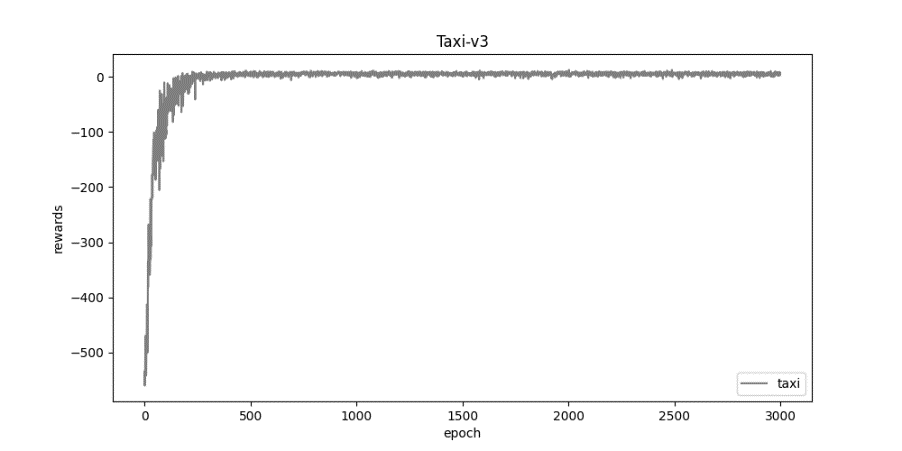
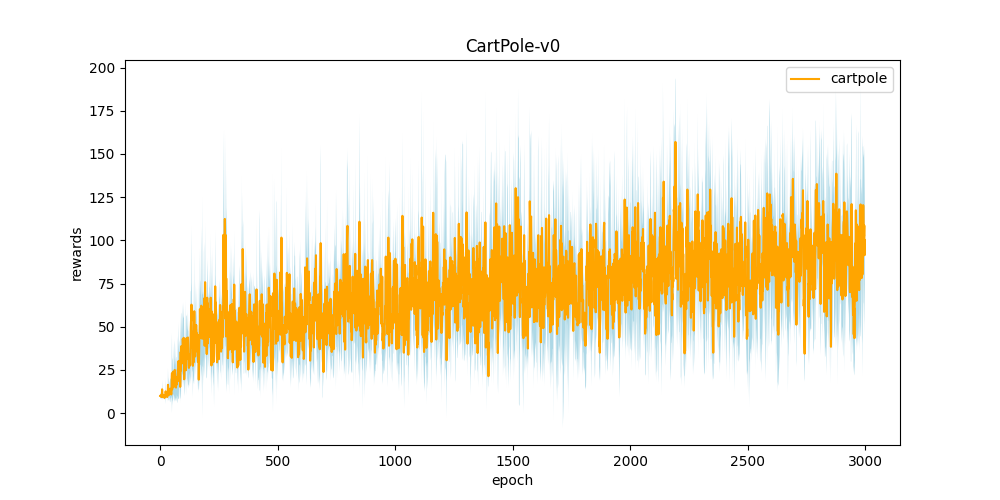
Part I. Experiment Results (the score here is included in your implementation):

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

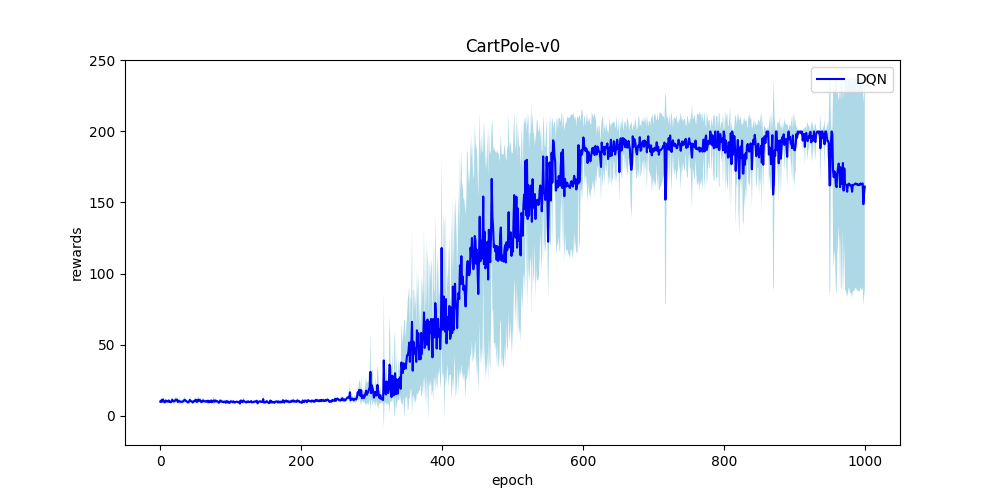
1. taxi.png:



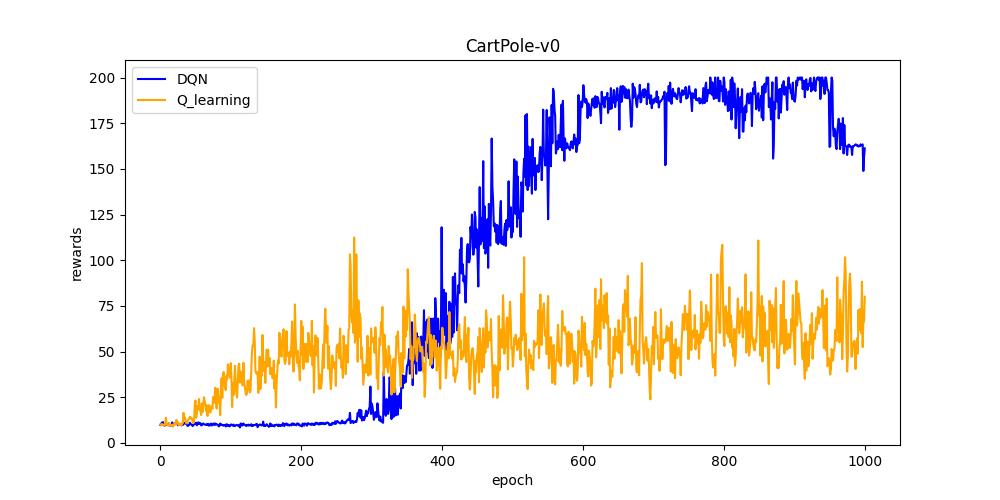
2. cartpole.png:



3. DQN.png:



4. compare.png:



Part II. Question Answering (50%):

1. Calculate the optimal Q-value of a given state in Taxi-v3 (the state is assigned

in google sheet), 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). (4%)

一張含有 文字 的圖片

自動產生的描述

The optimal Q value = -1(1+r+r^2+r^3+r^4+r^5+...+r^12+r^13)+r^13\*20 =

-3.13696226351

The result is different since the Q value we learned is an estimated value, it’s reasonable to be different from the optimal Q value.

2. 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) (4%)

一張含有 文字 的圖片

自動產生的描述

The optimal Q value = 1(1+r+r^2+….+r^199) = 33.25795863300011

The result is different from we learned since the learned value is an estimate value, it’s reasonable to be different from the optimal Q value.

3.

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

Ans: Since we can’t categorize infinite states with finite array index.

b. How do you expect the performance will be if we increase “num\_bins”?

(2%)

Ans: It would go better since more states means it express the continuous states better.

c. Is there any concern if we increase “num\_bins”? (2%)

Ans: Since the states increases, the calculation time would grow.

4. Which model (DQN, discretized Q learning) performs better in Cartpole-v0,

and what are the reasons? (3%)

Ans: DQN will perform better since it can include continuous states, which can express every possible variations.

5.

a. What is the purpose of using the epsilon greedy algorithm while

choosing an action? (2%)

Ans: At first, since the machine hasn’t learn anything, we had no choice but to randomly pick actions, as the machine learns more, we would prefer the machine’s answer more as the machine becomes ’smarter’.

b. What will happen, if we don’t use the epsilon greedy algorithm in the

CartPole-v0 environment? (3%)

Ans: The machine can’t use its current optimal action to improve itself, which means it won’t learn at all.

c. Is it possible to achieve the same performance without the epsilon

greedy algorithm in the CartPole-v0 environment? Why or Why not?

(3%)

Ans: it’s possible for achieving the same performance if we can find a function adequately choose action between the machine’s output or random one.

d. Why don’t we need the epsilon greedy algorithm during the testing

section? (2%)

Ans: since we believe the machine is fully learned, we don’t need random actions but to believe the machine’s optimal choice.

6. Why is there “with torch.no\_grad():“ in the “choose\_action” function in

DQN? (3%)

Ans: we can’t have partial derivative value for choose\_action since it only chooses action.

7.

a. Is it necessary to have two networks when implementing DQN? (1%)

Ans: yes, otherwise we would have trouble of changing network values.

b. What are the advantages of having two networks? (3%)

Ans: it’s easier for management and changes value for network in specific loop process.

c. What are the disadvantages? (2%)

Ans: we need extra memory to save the other network.

8.

a. What is a replay buffer(memory)? Is it necessary to implement a replay

buffer? What are the advantages of implementing a replay buffer? (5%)

Ans: reply buffer is a device for giving adequate input data for RAM to calculate.

It is necessary since we may not have everything calculate at one time for limited RAM.

The advantage is that buffer would distribute adequately data for RAM, which increases speed while not causing RAM overload.

b. Why do we need batch size? (3%)

Ans: We need it because RAM can’t afford to calculate all data in memory at once.

c. Is there any effect if we adjust the size of the replay buffer(memory) or

batch size? Please list some advantages and disadvantages. (2%)

Ans: If we increase the batch size, we can calculate the data faster but increase the load of RAM, which could result in overload.

If we decrease the batch size, we can reduce the load of RAM, but have longer calculation time for data.

9.

a. What is the condition that you save your neural network? (1%)

Ans: We save the network when the reward value is bigger than a specific value.

b. What are the reasons? (2%)

Ans: We believe the network is fully trained when the reward value is big enough since higher reward means better performance.

10. What have you learned in the homework? (2%)

Ans: I learned how to use lots of different libraries to help me deal with the homework, and how to train neural networks with different models.