National Taiwan University
Department of Electrical Engineering
Name: JUI-CHAO LU
ID: B09901142

112-1 (Fall 2023)
Reinforcement Learning

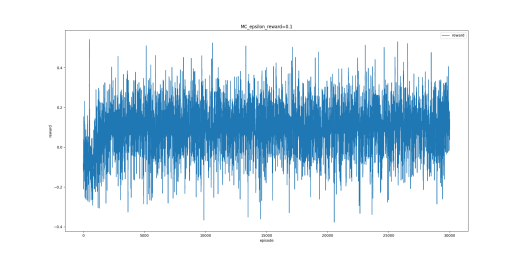
Assignment 2 Report

Q1. Discuss and plot learning curves under ϵ values of $(0.1, 0.2, 0.3, 0.4)$ on MC, SARSA, and Q-Learning

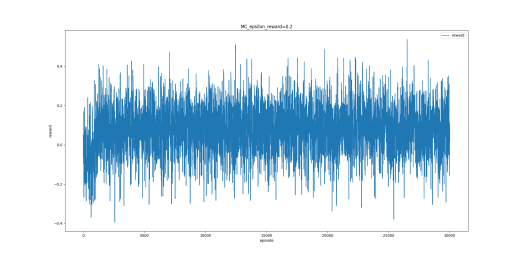
1. Below are the learning curve figures of the three methods, we can observe some properties:
   (a) They all go from low to high.(All three methods.)
   (b) SARSA and Q_learning performs better since it earns more reward rather than gaining oscillating rewards like Monte Carlo methods.
   (c) I didn't discover the difference between each ϵ settings, but I guess when $\epsilon = 0.3$ it performs better(I have this guess from SARSA learning curve)
   (d) In this question, I fixed the max_episodes of MC & SARSA to 30000, Q_learning to 5000 Also in my implementation, I altered the epsilon greedy policy of the first n episodes into random policy
   (e) Implementation of the Average non-discounted Episodic Reward looks like the following:

```
reward_record, loss_record = run_Q_Learning(grid_world, Q_episode
,epsilon=0.2)
avg_reward = [sum(row)/len(row) for row in reward_record]
avg_r = [np.mean(avg_reward[i:i+10]) for i in range(0, len(avg_reward))]
```

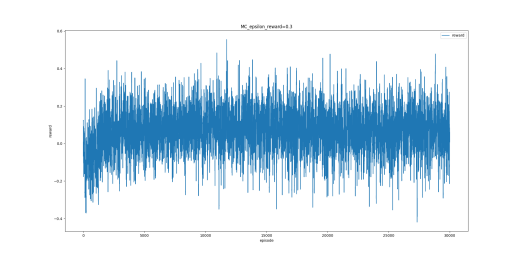
2. Monte Carlo



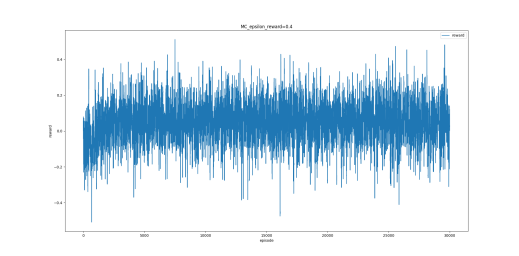
(a). epsilon=0.1



(b). epsilon=0.2



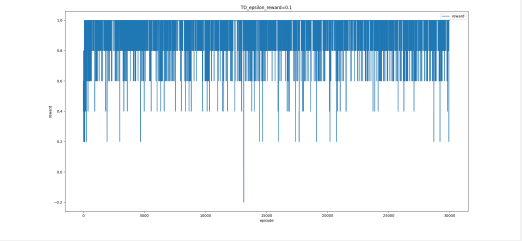
(c). epsilon=0.3



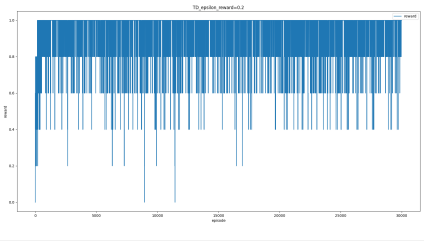
(d). epsilon=0.4

Figure 1: Monte Carlo Learning Curve

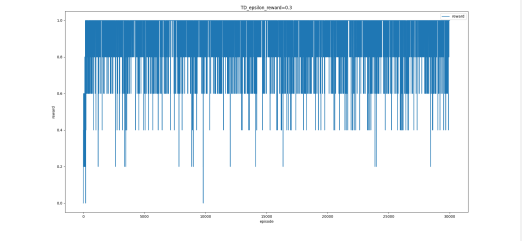
3. SARSA



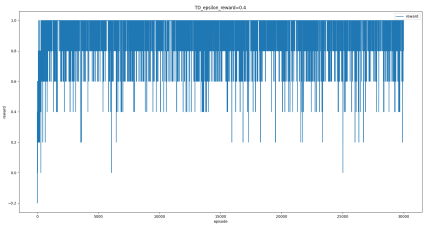
(a). epsilon=0.1



(b). epsilon=0.2



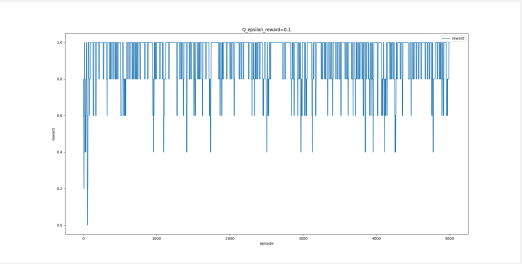
(c). epsilon=0.3



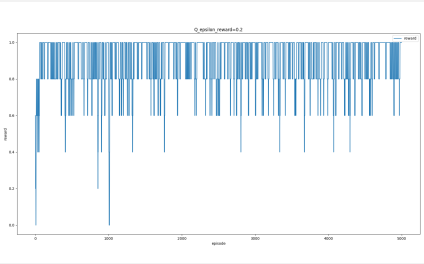
(d). epsilon=0.4

Figure 2: SARSA Learning Curve

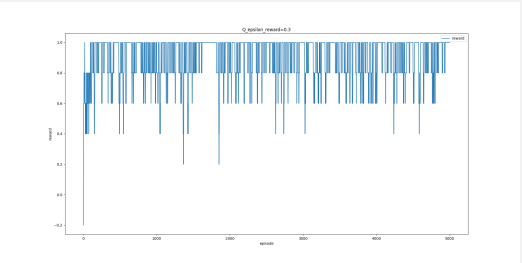
4. Q_learning



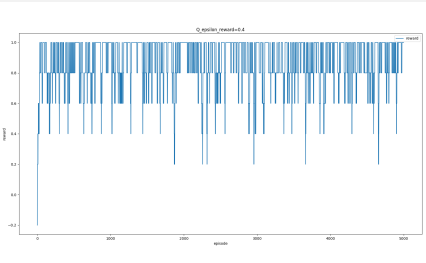
(a). epsilon=0.1



(b). epsilon=0.2



(c). epsilon=0.3



(d). epsilon=0.4

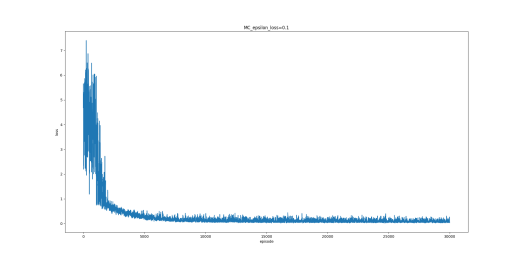
Figure 3: Q Learning Curve

Q2. Discuss and plot loss curves under ϵ values of $(0.1, 0.2, 0.3, 0.4)$ on MC, SARSA, and Q-Learning

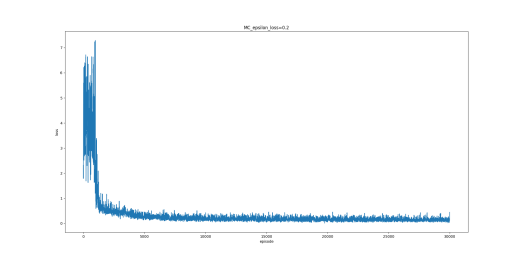
1. Below are the loss curve figures of the three methods, we can observe some properties:
   (a) They all go from high to low.(All three methods.)
   (b) SARSA and Q_learning performs worse since it has bigger bias because of the estimations
   (c) I didn't discover the difference between each ϵ settings, but I guess when $\epsilon = 0.1, 0.4$ it performs better(I have this guess from all three curves)

(d) In this question, I fixed the max_episodes of Monte Carlo & SARSA to 30000, Q_learning to 5000.

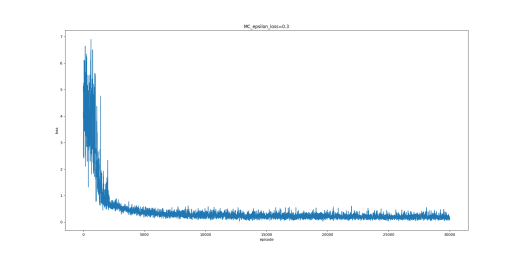
2. Monte Carlo



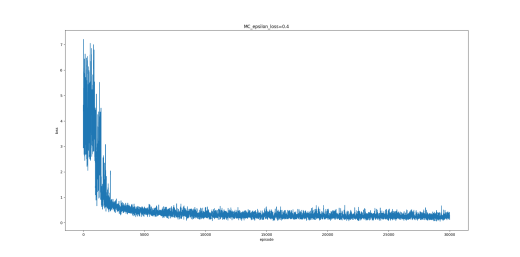
(a). epsilon=0.1



(b). epsilon=0.2



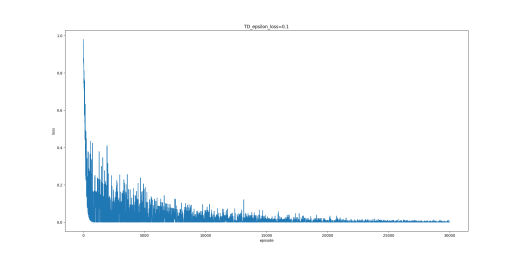
(c). epsilon=0.3



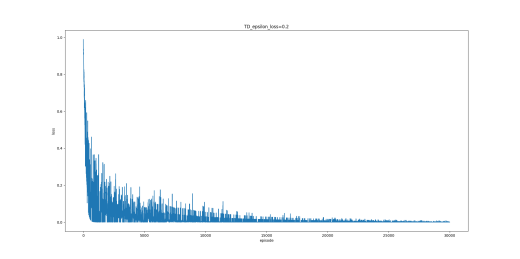
(d). epsilon=0.4

Figure 4: Monte Carlo Loss Curve

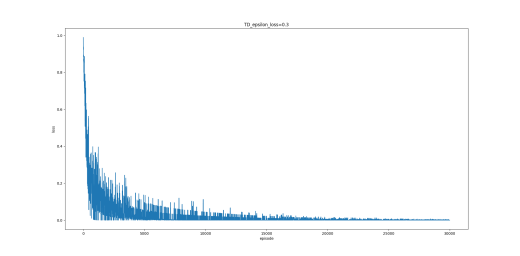
3. SARSA



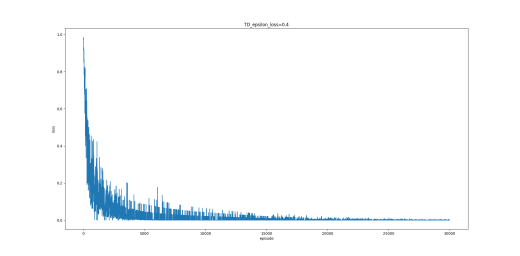
(a). epsilon=0.1



(b). epsilon=0.2



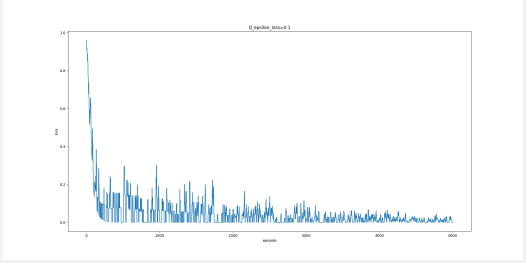
(c). epsilon=0.3



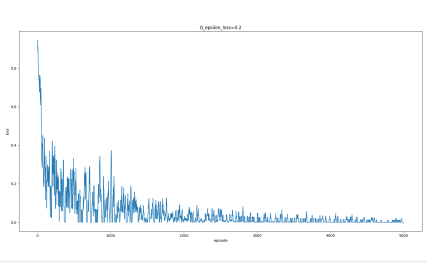
(d). epsilon=0.4

Figure 5: SARSA Loss Curve

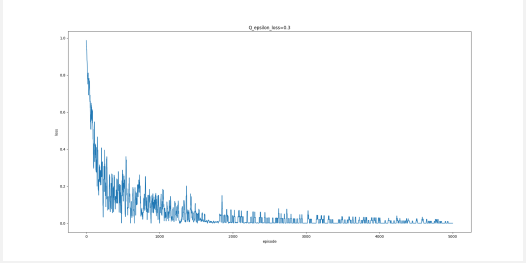
4. Q_learning



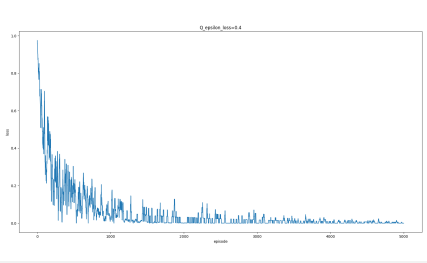
(a). epsilon=0.1



(b). epsilon=0.2



(c). epsilon=0.3



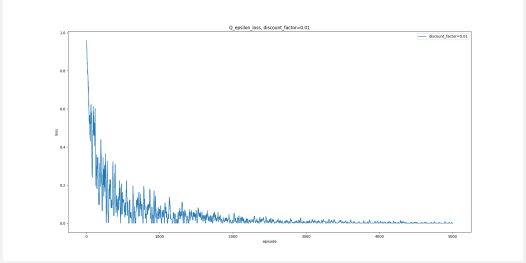
(d). epsilon=0.4

Figure 6: Q Loss Curve

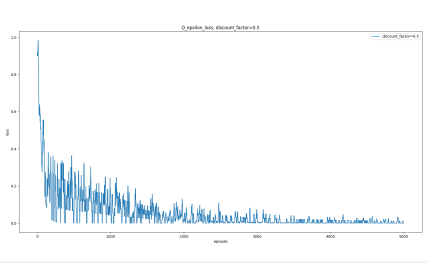
Q3. Discuss and plot ...

1. Below are the figures of different settings of the three methods, we can observe some properties:(In this question, I fixed the max_episodes settings identical with the previous two questions, other hyperparameters are set to default value if not discussed in each section below.)

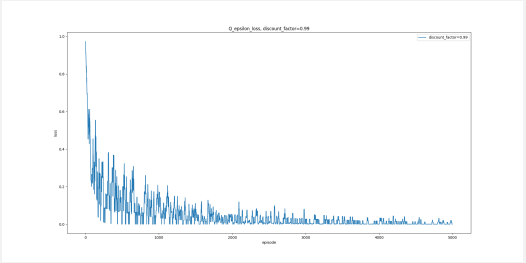
2. discount_factor: In this section, we can find out that as the discount_factor increases, the loss converges slower.



(a). discount_factor=0.01



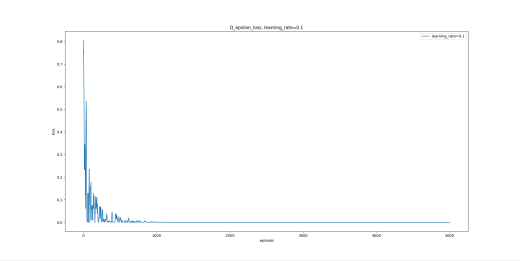
(b). discount_factor=0.5



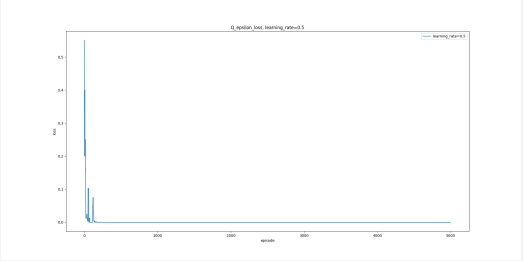
(c). discount_factor=0.99

Figure 7: Q Learning Loss Curve Different Discount Factor

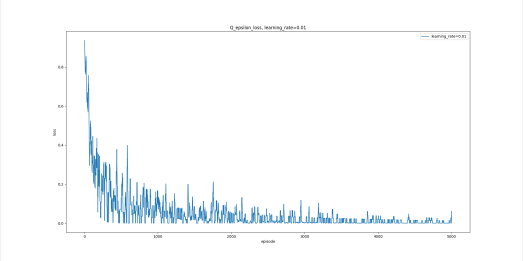
3. learning_rate: In this section, we can find out that as the discount_factor decreases, the loss converges slower.



(a). learning_rate=0.1



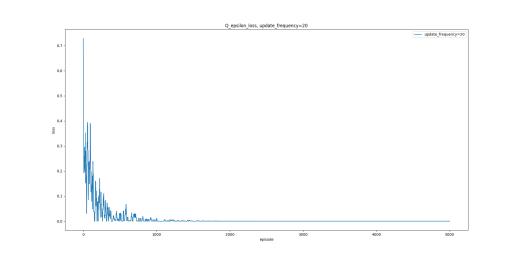
(b). learning_rate=0.5



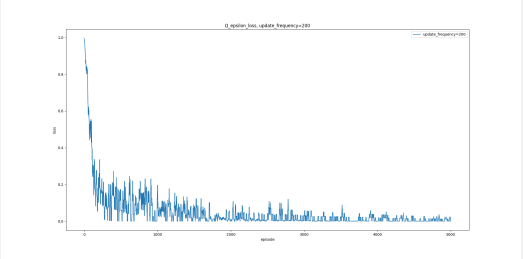
(c). learning_rate=0.01

Figure 8: Q Learning Loss Curve Different Learning Rate

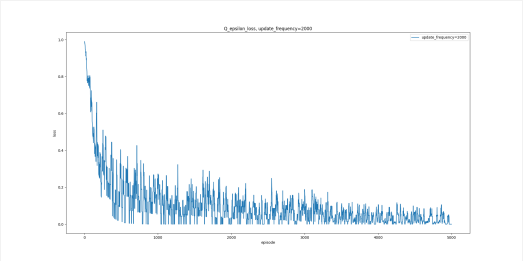
4. Update Frequency: In this section, we can find out that as the discount_factor increases, the loss converges slower.



(a). update_frequency=20



(b). update_frequency=200



(c). update_frequency=2000

Figure 9: Q Learning Loss Curve Different Update Frequency