Reinforcement Learning HW1 Part 2

Name: Jiayi Zhang UNI: jz2856

1

(1) Value Iteration

The optimal value function is:

```
304.988
[ 379.
          273.301
                             258.636
                                       195.604
                                                273.301 174.582
                                                                   206.951
                     304.988
                              206.951
                                        174.582
                                                           174.582
  244.705
           195.604
                                                 195.604
                                                                    258.636
  400.
           288.738
                    322.092
                              273.301
                                       359.05
                                                 258.636
                                                          288.738
                                                                    244.705
           288.738
                                        231.469
  206.951
                     184.823
                              218.896
                                                 184.823
                                                           288.738
                                                                    195.604
  184.823
           206.951
                     184.823
                              273.301
                                        379.
                                                 304.988
                                                           304.988
                                                                    288.738
  304.988
           218.896
                     244.705
                              206.951
                                        244.705
                                                 340.097
                                                           218.896
                                                                    258.636
  218.896
           174.582
                     273.301
                              184.823
                                        195.604
                                                 218.896
                                                           195.604
                                                                    288.738
  322.092
           359.05
                     288.738
                              304.988
                                        288.738
                                                 206.951
                                                           231.469
                                                                    195.604
  258.636
           359.05
                     231.469
                              273.301
                                        206.951
                                                 164.853
                                                           258.636
                                                                    174.582
  206.951
           231.469
                     206.951
                              304.988
                                        304.988
                                                 379.
                                                           273.302
                                                                    322.092
  273.302
           195.604
                     218.896
                              184.823
                                        273.302
                                                 379.
                                                           244.705
                                                                    288.738
  195.604
           155.61
                              164.853
                                        195.604
                     244.705
                                                 218.896
                                                           195.604
                                                                    288.738
                     258.636
  288.738
           400.
                              304.988
                                        359.05
                                                 258.636
                                                           288.738
                                                                    244.705
  206.951
           288.738
                     184.823
                              218.896
                                        258.636
                                                 206.951
                                                           322.092
                                                                    218.896
  184.823
           206.951
                     184.823
                              273.301
                                        379.
                                                 304.988
                                                           340.097
                                                                    288.738
                              231.469
                                                           195.604
  340.097
           244.705
                     273.301
                                        218.896
                                                 304.988
                                                                    231.469
  244.705
           195.604
                     304.988
                              206.951
                                        195.604
                                                 218.896
                                                           195.604
                                                                    288.738
  359.05
           322.092
                     322.092
                              304.988
                                        322.092
                                                 231.469
                                                           258.636
                                                                    218.896
  231.469
           322.092
                     206.951
                              244.705
                                        231.469
                                                 184.823
                                                           288.738
                                                                    195.604
  206.951
           231.469
                     206.951
                              304.988
                                        340.097
                                                 340.097
                                                           304.988
                                                                    322.092
  304.988
           218.896
                     244.705
                              206.951
                                        244.705
                                                 340.097
                                                           218.896
                                                                    258.636
  218.896
           174.582
                     273.302
                              184.823
                                        218.896
                                                 244.705
                                                           218.896
                                                                    322.092
  322.092
           359.05
                     288.738
                              340.097
                                        288.738
                                                 206.951
                                                           231.469
                                                                    195.604
  258.636
           359.05
                     231.469
                              273.301
                                        206.951
                                                 164.853
                                                           258.636
                                                                    174.582
  206.951
           231.469
                     206.951
                              304.988
                                        304.988
                                                 379.
                                                           273.302
                                                                    322.092
  340.097
           244.705
                     273.301
                              231.469
                                        195.604
                                                 273.301
                                                           174.582
                                                                    206.951
           218.896
  273.302
                     340.097
                              231.469
                                        195.604
                                                 218.896
                                                           195.604
                                                                    288.738
  359.05
                     359.05
                                        322.092
           288.738
                              304.988
                                                 231.469
                                                           258.636
                                                                    218.896
  206.951
           288.738
                     184.823
                              218.896
                                        258.636
                                                 206.951
                                                           322.092
                                                                    218.896
```

206.951	231.469	206.951	304.988	340.097	304.988	340.097	322.092
304.988	218.896	244.705	206.951	218.896	304.988	195.604	231.469
244.705	195.604	304.988	206.951	218.896	244.705	218.896	322.092
322.092	322.092	322.092	340.097	288.738	206.951	231.469	195.604
231.469	322.092	206.951	244.705	231.469	184.823	288.738	195.604
231.469	258.636	231.469	340.097	304.988	340.097	304.988	359.05
273.302	195.604	218.896	184.823	244.705	340.097	218.896	258.636
218.896	174.582	273.302	184.823	218.896	244.705	218.896	322.092
288.738	359.05	288.738	340.097	322.092	231.469	258.636	218.896
184.823	258.636	164.853	195.604	288.738	231.469	359.05	244.705
184.823	206.951	184.823	273.301	340.097	273.302	379.	288.738
304.988	218.896	244.705	206.951	195.604	273.302	174.582	206.951
244.705	195.604	304.988	206.951	195.604	218.896	195.604	288.738
322.092	288.738	322.092	304.988	288.738	206.951	231.469	195.604
206.951	288.738	184.823	218.896	231.469	184.823	288.738	195.604
206.951	231.469	206.951	304.988	304.988	304.988	304.988	322.092
273.302	195.604	218.896	184.823	218.896	304.988	195.604	231.469
218.896	174.582	273.302	184.823	244.705	273.302	244.705	359.05
288.738	322.092	288.738	379.	258.636	184.823	206.951	174.582
231.469	322.092	206.951	244.705	206.951	164.853	258.636	174.582
231.469	258.636	231.469	340.097	273.302	340.097	273.302	359.05
304.988	218.896	244.705	206.951	174.582	244.705	155.61	184.823
304.988	244.705	379.	258.636	174.582	195.604	174.582	258.636
322.092	258.636	400.	273.301	288.738	206.951	231.469	195.604
184.823	258.636	164.853	195.604	231.469	184.823	288.738	195.604
184.823	206.951	184.823	273.301	304.988	273.302	304.988	288.738
273.302			184.823			174.582	206.951
							288.738
	288.738	288.738	304.988	258.636	184.823	206.951	174.582
206.951	288.738	184.823	218.896	206.951	164.853	258.636	174.582
258.636	288.738	258.636	379.	273.302	304.988	273.302	400.
244.705	174.582	195.604	164.853	218.896	304.988	195.604	231.469
195.604	155.61	244.705	164.853	244.705	273.302	244.705	359.05
258.636	322.092	258.636	379.]				
	304.988 244.705 322.092 231.469 231.469 273.302 218.896 288.738 184.823 304.988 244.705 322.092 206.951 273.302 218.896 288.738 231.469 231.469 304.988 304.988 304.988 322.092 184.823 184.823 273.302 218.896 288.738 231.469 231.469 304.988	304.988 218.896 244.705 195.604 322.092 322.092 231.469 258.636 273.302 195.604 218.896 174.582 288.738 359.05 184.823 258.636 184.823 206.951 304.988 218.896 244.705 195.604 322.092 288.738 206.951 231.469 273.302 195.604 218.896 174.582 288.738 322.092 231.469 258.636 304.988 218.896 304.988 244.705 322.092 258.636 184.823 258.636 184.823 258.636 184.823 206.951 273.302 195.604 218.896 174.582 288.738 288.738 206.951 288.738 258.636 288.738 26.951 288.738 258.636 288.738 258.636 288.738 25	304.988218.896244.705244.705195.604304.988322.092322.092322.092231.469258.636231.469273.302195.604218.896218.896174.582273.302288.738359.05288.738184.823206.951184.823304.988218.896244.705244.705195.604304.988322.092288.738322.092206.951288.738184.823206.951231.469206.951273.302195.604218.896218.896174.582273.302288.738322.092288.738231.469322.092288.738231.469258.636231.469304.988218.896244.705304.988244.705379.322.092258.636400.184.823258.636164.853184.823266.951184.823273.302195.604218.896218.896174.582273.302288.738288.738288.738266.951288.738288.738268.951288.738288.738258.636288.738258.636244.705174.582195.604195.604155.61244.705	304.988 218.896 244.705 206.951 244.705 195.604 304.988 206.951 322.092 322.092 340.097 231.469 322.092 206.951 244.705 231.469 258.636 231.469 340.097 273.302 195.604 218.896 184.823 218.896 174.582 273.302 184.823 288.738 359.05 288.738 340.097 184.823 258.636 164.853 195.604 184.823 258.636 164.853 195.604 184.823 206.951 184.823 273.301 304.988 218.896 244.705 206.951 244.705 195.604 304.988 206.951 2306.951 231.469 206.951 304.988 273.302 195.604 218.896 184.823 218.896 174.582 273.302 184.823 288.738 322.092 288.738 379. 231.469 258.636 231.469 340.097 304.988 244.705 <	304.988 218.896 244.705 206.951 218.896 244.705 195.604 304.988 206.951 218.896 322.092 322.092 340.097 288.738 231.469 322.092 206.951 244.705 231.469 231.469 258.636 231.469 340.097 304.988 273.302 195.604 218.896 184.823 244.705 218.896 174.582 273.302 184.823 218.896 288.738 359.05 288.738 340.097 322.092 184.823 258.636 164.853 195.604 288.738 184.823 206.951 184.823 273.301 340.097 304.988 218.896 244.705 206.951 195.604 244.705 195.604 304.988 206.951 195.604 231.469 288.738 184.823 218.896 244.705 288.738 184.823 218.896 218.896 174.582 273.302 184.823<	304.988 218.896 244.705 206.951 218.896 244.705 224.705 195.604 304.988 206.951 218.896 244.705 322.092 322.092 340.097 288.738 206.951 231.469 322.092 206.951 244.705 231.469 184.823 231.469 258.636 231.469 340.097 304.988 340.097 273.302 195.604 218.896 184.823 244.705 340.097 288.738 359.05 288.738 340.097 322.092 231.469 184.823 258.636 164.853 195.604 288.738 231.469 184.823 266.951 184.823 273.301 340.097 273.302 304.988 218.896 244.705 206.951 195.604 218.896 244.705 195.604 304.988 206.951 195.604 218.896 22092 288.738 322.092 304.988 304.988 304.988 231.469 231.469 268.51 184.823 218.896 184.823 2	304.988218.896244.705206.951218.896204.970218.896244.705195.604304.988206.951218.896244.705218.896322.092322.092340.097288.738206.951231.469231.469226.965244.705231.469184.823288.738231.469258.636231.469340.097304.988340.097218.896218.896174.582273.302184.823244.705244.705218.896288.738359.05288.738340.097220.92231.469258.636184.823258.636164.853195.604288.738231.469259.636184.823269.951184.823273.301340.097273.302379.304.988218.896244.705206.951195.604218.896195.604224.07195.604304.988206.951195.604218.896195.604226.951231.469266.951304.988304.988304.988304.988233.469195.604218.896344.705273.302244.705288.738322.092288.738379.258.636148.823266.951231.469258.636244.705266.951174.582244.705155.61304.988244.705379.258.636144.705273.302244.705288.738258.636244.705258.636144.823266.951244.705256.951304.98824

The optimal policy function is:

The average rewards for running 100 episodes using optimal policy is:

8.4141414141413

(2) Policy Iteration

The optimal value function is:

```
273.301 304.988 258.636 195.604
[ 379.
                                              273.301 174.582 206.951
  244.705
                    304.988
           195.604
                             206.951
                                      174.582
                                               195.604
                                                        174.582 258.636
  400.
           288.738
                    322.092
                             273.301
                                      359.05
                                               258.636
                                                        288.738
                                                                 244.705
  206.951
          288.738
                    184.823 218.896
                                      231.469
                                               184.823
                                                        288.738 195.604
  184.823
           206.951
                    184.823
                             273.301
                                      379.
                                               304.988
                                                        304.988
                                                                 288.738
  304.988
          218.896
                    244.705
                             206.951
                                      244.705
                                               340.097
                                                        218.896 258.636
  218.896
          174.582
                    273.301
                             184.823
                                      195.604
                                               218.896
                                                        195.604
                                                                 288.738
           359.05
                                      288.738
                                               206.951
  322.092
                    288.738
                             304.988
                                                         231.469
                                                                 195.604
  258.636 359.05
                    231.469
                             273.301
                                      206.951
                                                        258.636 174.582
                                               164.853
  206.951
          231.469
                    206.951
                             304.988
                                      304.988
                                               379.
                                                         273.302
                                                                 322.092
  273.302
          195.604
                    218.896
                             184.823
                                      273.302
                                               379.
                                                         244.705
                                                                 288.738
  195.604
          155.61
                    244.705
                             164.853
                                      195.604
                                               218.896
                                                        195.604
                                                                 288.738
  288.738
          400.
                    258.636
                             304.988
                                      359.05
                                               258.636
                                                        288.738
                                                                 244.705
  206.951
          288.738
                    184.823
                                      258.636
                             218.896
                                               206.951
                                                        322.092 218.896
  184.823
           206.951
                    184.823
                             273.301
                                      379.
                                               304.988
                                                        340.097
                                                                 288.738
  340.097
           244.705
                    273.301
                             231.469
                                      218.896
                                               304.988
                                                        195.604
                                                                 231.469
  244.705
          195.604
                    304.988
                             206.951
                                      195.604
                                               218.896
                                                        195.604
                                                                 288.738
  359.05
           322.092
                    322.092
                             304.988
                                      322.092
                                               231.469
                                                        258.636
                                                                 218.896
  231.469 322.092
                    206.951
                             244.705
                                      231.469
                                               184.823
                                                        288.738 195.604
                    206.951
  206.951
           231.469
                             304.988
                                      340.097
                                                                 322.092
                                               340.097
                                                         304.988
  304.988
          218.896
                    244.705
                             206.951
                                      244.705
                                               340.097
                                                         218.896
                                                                  258.636
  218.896
          174.582
                    273.302
                             184.823
                                      218.896
                                               244.705
                                                        218.896
                                                                 322.092
  322.092
           359.05
                    288.738
                             340.097
                                      288.738
                                               206.951
                                                        231.469
                                                                 195.604
  258.636
           359.05
                    231.469
                             273.301
                                      206.951
                                               164.853
                                                        258.636
                                                                 174.582
  206.951
           231.469
                    206.951
                             304.988
                                      304.988
                                               379.
                                                         273.302
                                                                 322.092
  340.097
           244.705
                    273.301
                             231.469
                                      195.604
                                               273.301
                                                        174.582
                                                                 206.951
  273.302
          218.896
                    340.097
                             231.469
                                      195.604
                                               218.896
                                                        195.604
                                                                 288.738
  359.05
           288.738
                    359.05
                             304.988
                                      322.092
                                               231.469
                                                        258.636
                                                                 218.896
  206.951
          288.738
                    184.823
                             218.896
                                      258.636
                                               206.951
                                                         322.092
                                                                 218.896
  206.951
          231.469
                    206.951
                             304.988
                                      340.097
                                               304.988
                                                         340.097
                                                                  322.092
```

304.988	218.896	244.705	206.951	218.896	304.988	195.604	231.469
244.705	195.604	304.988	206.951	218.896	244.705	218.896	322.092
322.092	322.092	322.092	340.097	288.738	206.951	231.469	195.604
231.469	322.092	206.951	244.705	231.469	184.823	288.738	195.604
231.469	258.636	231.469	340.097	304.988	340.097	304.988	359.05
273.302	195.604	218.896	184.823	244.705	340.097	218.896	258.636
218.896	174.582	273.302	184.823	218.896	244.705	218.896	322.092
288.738	359.05	288.738	340.097	322.092	231.469	258.636	218.896
184.823	258.636	164.853	195.604	288.738	231.469	359.05	244.705
184.823	206.951	184.823	273.301	340.097	273.302	379.	288.738
304.988	218.896	244.705	206.951	195.604	273.302	174.582	206.951
244.705	195.604	304.988	206.951	195.604	218.896	195.604	288.738
322.092	288.738	322.092	304.988	288.738	206.951	231.469	195.604
206.951	288.738	184.823	218.896	231.469	184.823	288.738	195.604
206.951	231.469	206.951	304.988	304.988	304.988	304.988	322.092
273.302	195.604	218.896	184.823	218.896	304.988	195.604	231.469
218.896	174.582	273.302	184.823	244.705	273.302	244.705	359.05
288.738	322.092	288.738	379.	258.636	184.823	206.951	174.582
231.469	322.092	206.951	244.705	206.951	164.853	258.636	174.582
231.469	258.636	231.469	340.097	273.302	340.097	273.302	359.05
304.988	218.896	244.705	206.951	174.582	244.705	155.61	184.823
304.988	244.705	379.	258.636	174.582	195.604	174.582	258.636
322.092	258.636	400.	273.301	288.738	206.951	231.469	195.604
184.823	258.636	164.853	195.604	231.469	184.823	288.738	195.604
184.823	206.951	184.823	273.301	304.988	273.302	304.988	288.738
273.302	195.604	218.896	184.823	195.604	273.302	174.582	206.951
218.896	174.582	273.302	184.823	195.604	218.896	195.604	288.738
288.738	288.738	288.738	304.988	258.636	184.823	206.951	174.582
206.951	288.738	184.823	218.896	206.951	164.853	258.636	174.582
258.636	288.738	258.636	379.	273.302	304.988	273.302	400.
244.705	174.582	195.604	164.853	218.896	304.988	195.604	231.469
195.604	155.61	244.705	164.853	244.705	273.302	244.705	359.05
258.636	322.092	258.636	379.]				

The optimal policy function is:

The average rewards for running 100 episodes using optimal policy is:

8.6767676767676

(3) Discussion

Through our program, we can find that in policy iteration, the policy evaluation and policy improvement are executed 17 times. And in policy evaluation, except for the first time of execution, every time the function will iterate 284 times. In value iteration, it also cost 284 times to converge.

It is obvious to find that the final value function and optimal policy function of those two algorithms are the same.

2

(4) Q-learning

The Q value array of each (s, a) pair is:

```
[[ 1.57695811e+01
                    1.38977082e+01 -1.97588113e-01
                                                       1.22927318e+01
   2.24964869e+02
                     1.96495207e+01]
[-2.01874904e+00 -2.07173833e+00 -2.03768696e+00
                                                       3.59134895e-01
    5.55325646e+01 -2.96810310e+00]
[-1.27313426e+00 7.66540215e+00 -1.28375812e+00
                                                       4.52851756e+00
   1.22922777e+02 -1.97468322e+00]
 [-9.77797391e-01 -8.75268062e-01 -9.77797391e-01 -1.06100615e+00
   -2.85106829e+00 -1.95851084e+00]
\begin{bmatrix} -2.26251646e+00 & -2.29003948e+00 & -2.34404189e+00 & -2.33294893e+00 \end{bmatrix}
   -2.91536208e+00 -2.97352331e+00]
[-1.99500000e-01 -1.99500000e-01 -1.99500000e-01
                                                      9.52935511e+01
   -1.00000000e+00 -1.0000000e+00]]
(too long to show all of those)
```

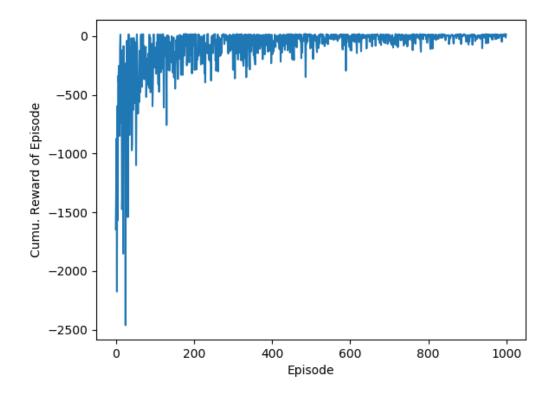
(5) SARSA

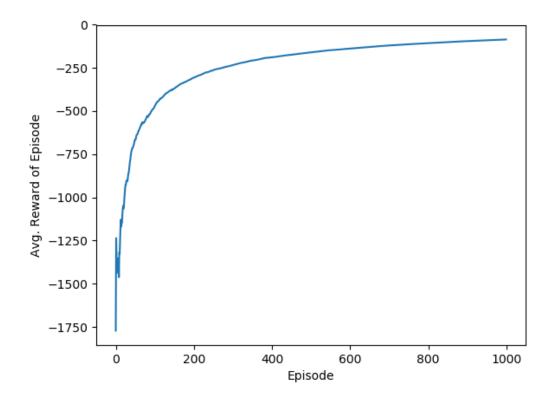
The Q value array of each (s, a) pair is:

```
[[-1.36657484 -1.29722863 -1.30307844 -1.15618847 -1.09954938 -2.81857346]
[-5.29191273 -5.25848374 -5.37791735 -5.28929674 -5.26014069 -9.37054361]
[-4.76142261 -4.79852202 -4.78706306 -4.78450661 -4.69004058 -7.96759454]
...,
[-3.59128127 -3.59226594 -3.66037699 -3.61907009 -6.74396321 -6.26997637]
[-6.02201626 -6.02959142 -5.98894215 -5.97649343 -8.23659351 -6.61345698]
[-1.04011687 -1.131997 -1.11934517 -0.28194184 -1.94485187 -2.7366 ]
[(too long to show all of those)
```

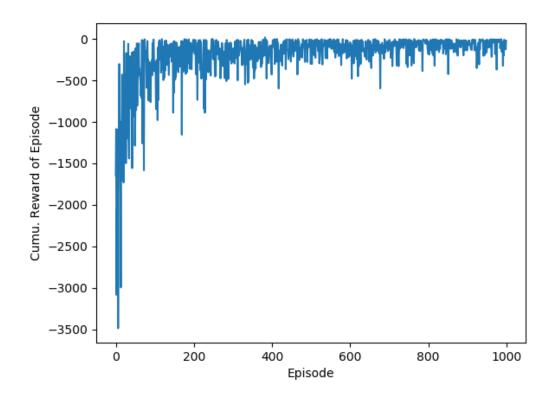
(6) Plot of the learning progress for training 1000 episodes

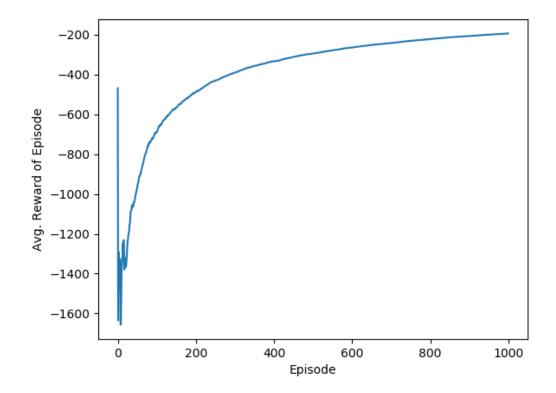
Plot of the learning progress of Q-learning:





Plot of the learning progress of SARSA:

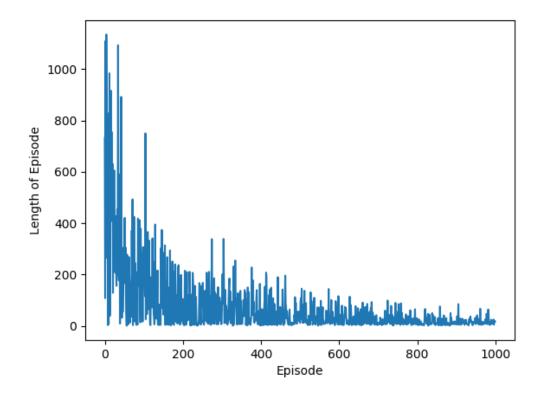




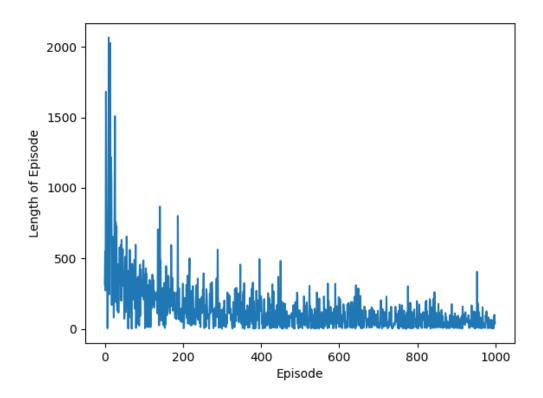
Discussion: From those four plots above, we can find that Q-learning algorithm performs better. In comparison, the Q-learning converges better and more stable in this condition.

(7) Plot of the episode length for training 1000 episodes

Plot of the episode length of Q-learning:



Plot of the episode length of SARSA:



Discussion: From those two plots above, we can find that the episode length of both algorithm also converges. And similar to accumulated reward of episode, the episode length convergence progress of Q-learning is also more stable and have better performance than the progress of SARSA.