RL Assignment 2

陈麒麟 517030910155

作业要求：

- 利用MC-first visit\every visit、TD0算法估计6x6 gridworld(1、35为出口) 的状态值

代码实现

generator方法

> 代码见 *self_gridworld.py*

产生一个6x6 gridworld的随机episode，调用如下：

```
sequence,g=generator()
```

MC1文件

> 代码见 *MC1.py*

文件为MC-first visit方法实现并输出最终状态矩阵，其中迭代次数可以控制，更新公式为：

$$N(S_t) \leftarrow N(S_t) + 1$$

$$V(S_t) \leftarrow V(S_t) + \frac{1}{N(S_t)}(G_t - V(S_t))$$

MC2文件

> 代码见 *MC2.py*

文件为MC-every visit方法实现并输出最终状态矩阵，其中迭代次数可以控制，更新公式同上，区别在于每一个episode中可以更新同一个状态两次。

TD0文件

> 代码见 *TD0.py*

文件为TD(0)方法实现并输出最终状态矩阵，其中迭代次数/alpha=0.1/gamma=0.7可以控制，更新公式为：

$$V(S_t) \leftarrow V(S_t) + \alpha(R_{t+1} + \gamma V(S_{t+1}) - V(S_t))$$

测试

MC-First Visit

10000次迭代后结果：

```
[Running] python -u "/Users/kylinchan/Documents/Spring2020-Git/RL/MC1.py"
-58.032274081430025  -44.887281292059164  -64.86986697513024  -73.82681074766342  -79.14437500000024  -84.59152215799615
-67.57641921397372  -62.94699570815458  -68.19612115929381  -72.71864091226465  -76.02605107041524  -80.23708774113258
-73.68745227793318  -70.36222780569507  -71.87667837223728  -73.51182469082228  -74.57345299952772  -77.38010860245785
-77.61812703991967  -75.03651071272404  -75.54098717112443  -74.62478070175426  -72.47898383371822  -73.02335516946724
-82.23818628789951  -79.88630170918961  -79.69767441860458  -74.87410071942425  -68.63879436071964  -63.58450704225346
-88.29645436638236  -83.19500561167212  -80.97542585869861  -75.84652210644889  -66.04810049019619  -44.361439842209116

[Done] exited with code=0 in 1.522 seconds
```

100000次迭代后结果1：

```
[Running] python -u "/Users/kylinchan/Documents/Spring2020-Git/RL/MC1.py"
-57.741025006189446  -45.374552964287474  -64.08754179885103  -72.58664829803159  -77.66441622793408  -82.58501801630949
-67.22958733148842  -63.11532146844407  -67.59228162239019  -71.64386536373517  -74.7004423204018  -78.96792714925157
-73.8483527752853  -70.26398849188656  -71.61737426790785  -72.7604388512427  -73.5576291850318  -76.65529243937245
-78.42022614661454  -74.24607815635615  -74.72844408952598  -73.65587194828991  -72.17583312015962  -72.62485395947928
-82.13913922114838  -79.23015151887212  -79.33858490323644  -74.01349796211933  -68.66988605799854  -64.01159848165345
-87.30654501925052  -82.61928976095044  -80.34653877075785  -75.23711545576084  -64.85847801100981  -43.48168937464461

[Done] exited with code=0 in 9.839 seconds
```

100000次迭代后结果2：

```
[Running] python -u "/Users/kylinchan/Documents/Spring2020-Git/RL/MC1.py"
-57.35497495334441  -45.07719103392586  -63.96273345136797  -72.50119352333871  -77.04835999253221  -81.68243730318318
-67.22254749846311  -63.109116871704884  -67.4021360069746  -71.38049962714386  -74.29853573622731  -78.33951860227418
-73.72874130004801  -70.18240361651367  -71.34687962634766  -72.55652907580473  -73.46797679649508  -76.11357071357115
-77.87088233796649  -74.2350400314145  -74.39050154772062  -73.41853182077092  -71.63211122218884  -72.03402522707239
-81.75163703115146  -79.11527391822155  -78.9914545360656  -73.72688672455939  -68.24934830803865  -63.374887515747716
-87.25686965602722  -82.42448729380284  -80.19342160278751  -74.7769033361853  -64.58120355265571  -43.12795741324948

[Done] exited with code=0 in 9.251 seconds
```

可以看出，10000次迭代时已经收敛，而且其最优方向是与直观一致。

MC-Every Visit

10000次迭代后结果：

```
[Running] python -u "/Users/kylinchan/Documents/Spring2020-Git/RL/MC2.py"
-65.22970903522202  -45.42461329253806  -77.74155635283927  -92.69937075674757  -100.91393210749678  -103.99773289365243
-80.31372738705618  -79.49342177998884  -88.90966909392586  -97.83943705777159  -102.48499999999966  -102.66002589182965
-96.12553350044018  -96.03480741797483  -98.89831026836957  -101.64170294399777  -102.2583025830253  -99.09435827250626
-107.35333511920687  -106.58505082271024  -104.50998964037277  -100.72869712874373  -95.78858248669573  -90.08457839601544
-114.33656675248237  -113.79464510181887  -105.6996224165347  -97.14446397188031  -85.60889066081936  -72.00344736144268
-118.53246574714765  -115.50194950774856  -107.86357008956479  -95.52438919105568  -75.94288750166051  -43.388146708768566

[Done] exited with code=0 in 1.487 seconds
```

100000次迭代后结果1：

```
[Running] python -u "/Users/kylinchan/Documents/Spring2020-Git/RL/MC2.py"
-63.59681297778365  -45.272740989103156  -76.16303167144609  -92.07304227986437  -99.80005932957343  -103.22240066380171
-80.92731198272811  -79.34121055377311  -88.66603503754054  -96.19580601280833  -100.33019567936674  -102.4552803542047
-96.50815777301298  -96.392999653018  -97.93904143896272  -98.70711186291103  -99.14641628774453  -99.24412145695656
-107.82614020666446  -106.4727112073871  -103.58213520954358  -99.70665973263284  -95.18919203112459  -91.04041701878236
-113.88964503176265  -112.45092735537011  -104.90565515331825  -97.30702007610527  -86.94752999719158  -75.02487502975481
-116.38381636256474  -113.529826802458  -106.25870991687532  -95.19070855380774  -77.24386628023474  -43.22760842627008

[Done] exited with code=0 in 12.825 seconds
```

100000次迭代后结果2：

```
[Running] python -u "/Users/kylinchan/Documents/Spring2020-Git/RL/MC2.py"
-63.722046079052824  -45.10571050914368  -75.41841076215577  -91.19630442736832  -99.43830279983672  -102.79904625201569
-80.53131924659526  -77.83196653822141  -86.85669095177953  -94.9190151474109  -99.97002608286387  -102.15755179481512
-95.57344646419875  -94.57614069327747  -96.45287999340296  -98.39124233983377  -98.86236213743136  -98.6067711617348
-106.13547239046918  -105.17212223973338  -102.19896957268406  -98.80371221822817  -94.36445869120028  -90.52839305530003
-112.78110974245767  -111.72963428066963  -104.23932316633231  -96.23742849599708  -85.96971780431457  -74.63197100596433
-114.84370540895465  -112.36827537617593  -105.4816051622739  -93.41698875614891  -75.40518426657333  -42.97924589267489

[Done] exited with code=0 in 13.022 seconds
```

相比MC-first visit，其状态值偏大，因为每个episode可以计入多次同一状态值。且在10000次迭代时已经收敛，其最优方向是与直观一致。

TD0

> alpha=0.1 gamma=0.7

10000次迭代后结果：

```
[Running] python -u "/Users/kylinchan/Documents/Spring2020-Git/RL/TD0.py"
-1.4011405027932253  0.0  -2.0960658666826393  -3.0339153036401347  -3.2656670446663543  -3.3032838881016926
-2.603948252248251  -1.9119700843477254  -2.9001430674255486  -3.153110973924513  -3.287212988735282  -3.2950964804672207
-3.1377189909999355  -3.100304454412846  -3.220850560576114  -3.2720027485213268  -3.2752944976196887  -3.2110389719526764
-3.259469857404593  -3.26932318603583  -3.2892391421648104  -3.2700375163071214  -3.219215377063616  -2.884682456565302
-3.308398845003221  -3.3134097145904144  -3.2844237995850296  -3.191860897015623  -2.8252574521110714  -1.5872722559936339
-3.32191982744949  -3.306443807555457  -3.2458705356647206  -2.947885434797782  -1.8948366955794849  0.0

[Done] exited with code=0 in 1.808 seconds
```

100000次迭代后结果1：

```
[Running] python -u "/Users/kylinchan/Documents/Spring2020-Git/RL/TD0.py"
-2.0694477252596064  0.0  -2.0767548892784804  -3.0653229491727414  -3.234910282200316  -3.3070264906495828
-2.8120364850059003  -2.200253739973374  -2.8863601902833467  -3.2230697262840398  -3.2840274920386063  -3.3051189228261206
-3.122465817184202  -2.994570727963999  -3.2012608260835487  -3.2573437990231255  -3.2730757329652196  -3.2651091393141027
-3.257172776551447  -3.24783347215981  -3.2632647261919505  -3.2341574096749452  -3.2051063327252276  -3.000314876712975
-3.3100557541290807  -3.307769976231598  -3.2798954901957895  -3.1280377226053746  -2.981100996005023  -2.1666711396957363
-3.320971515881775  -3.3128451825491627  -3.235520643618275  -3.1019073410268554  -2.4032691559084287  0.0

[Done] exited with code=0 in 12.991 seconds
```

100000次迭代后结果2：

```
[Running] python -u "/Users/kylinchan/Documents/Spring2020-Git/RL/TD0.py"
-2.211239467522885  0.0  -2.3790842457717214  -3.142306230545162  -3.260877499308514  -3.303595667718595
-2.6645723286730454  -2.3428743631137827  -3.010430855115857  -3.197646708736897  -3.2760499693908898  -3.298436097628074
-3.15110699689335  -3.0129193941145913  -3.2011024456328854  -3.2481152736633083  -3.2764890807361113  -3.2523144440706617
-3.2677052125660127  -3.2598977319663227  -3.275236839801405  -3.2444583174618478  -3.1588856479030527  -3.040324593541634
-3.3176121482445176  -3.300780493416651  -3.286522456544127  -3.1698248190939697  -2.8189795113937475  -2.424971125204279
-3.322684851819316  -3.303498299876818  -3.227323903990622  -2.950640906311793  -1.6362996696664596  0.0

[Done] exited with code=0 in 123.82 seconds
```

相比MC方法，其收敛较慢，因为每次更新都是由非episode-compete进行更新，存在误差大于MC。