

افروز شیخ الاسلامی 9729393

به نام خدا

سوال اول

(الف)

انجام بازی دوستانه : F

انجام بازی در لیگ: L

تیم قوی باشد: S

نتیجه بازی در لیگ: R

U : Utility

f2 (F L, S) :

F	L	S	U
T	T	T	80
T	T	F	-100
T	F	T	30
T	F	F	-30
F	T	T	100
F	T	F	-80
F	F	T	0
F	F	F	0

ب) باید از آخر به اول حذف شوند. در این سوال ابتدا بازی دوستانه انجام می شود و سپس بازی در لیگ انجام می شود. بنابراین باید از حذف بازی در لیگ (L) شروع کنیم.

سوال دوم

باید سه مورد را مشخص کنیم:

1. استیت ها برای هر دو مسئله راس های گراف هستند.
2. اکشن ها نیز یال های خارج شده از هر راس است.
3. تنها نقطه تمایز دو مسئله مقادیر reward است.

برای مسئله ی کوتاه ترین مسیر نود پاداش نود شروع 0 و بقیه نود ها مقدار منفی و نود نهایی یک مقدار مثبت زیاد باشد و مقدار منفی نود های دیگر باید متناسب با این مقدار مثبت باشد (مثلا مقادیر -1 و 20 مقادیر خوبی هستند).

برای مسئله یافتن مجموعه های همبند، مقادیر پاداش یکسان و مثبتی برای هر نود در نظر می گیریم و هر بار وارد نودی شدیم آن را نشان می کنیم. هر بار دیگر در حرکت ها نود جدیدی مشاهده نشد نود های مشاهده شده را حذف و دوباره الگوریتم را اجرا می کنیم تا زمانی که همه ی نود ها حذف شوند.

مقدار +1 برای پاداش مناسب است.

برای discount factor هر دو مسئله 0.9 کافی است.

سوال سوم)

$V^*(s) \rightarrow$ start from s , acting optimally for ∞ steps.

$V_H^*(s) \rightarrow$ " " " " for H steps.

Additional reward collected over steps $H+1, H+2$

$$\gamma^{H+1} \cdot R(s_{H+1}) + \gamma^{H+2} R(s_{H+2}) + \dots \leq \gamma^{H+1} R_{max} + \gamma^{H+2} R_{max} + \dots$$

$$= \frac{\gamma^{H+1}}{1-\gamma} R_{max}$$

$$\lim_{H \rightarrow \infty} \frac{\gamma^{H+1}}{1-\gamma} R_{max} = 0 \Rightarrow V_H^* \xrightarrow{H \rightarrow \infty} V^*$$

سوال چهارم)

(الف)

```

number of states = 5
-----
slip chance = 0.2
-----
small reward = 2
-----
large reward = 10
-----
gamma = 0.9
-----
V* = [32.473856 34.206336 39.390336 46.590336 56.590336]

```

(ب)

N=5

```

number of states = 5
-----
slip chance = 0.2
-----
small reward = 2
-----
large reward = 10
-----
gamma = 0.99
-----
V* = [435.0601257 438.02805658 444.30069658 452.22069658 462.22069658]

```

```

number of states = 5
-----
slip chance = 0.2
-----
small reward = 2
-----
large reward = 10
-----
gamma = 0.999
-----
V* = [4481.24055962 4484.34521497 4490.73242137 4498.72442137 4508.72442137]

```

N=10

```
number of states = 10
-----
slip chance = 0.2
-----
small reward = 2
-----
large reward = 10
-----
gamma = 0.9
-----
V* = [16.45288779 14.4822699 14.64550381 15.55235892 17.48727655 20.17466215
      23.90714215 29.09114215 36.29114215 46.29114215]
```

```
number of states = 10
-----
slip chance = 0.2
-----
small reward = 2
-----
large reward = 10
-----
gamma = 0.99
-----
V* = [177.96195163 176.84891888 178.80360159 181.27163531 184.38783951
      188.32244077 193.29037165 199.56301165 207.48301165 217.48301165]
```

```
number of states = 10
-----
slip chance = 0.2
-----
small reward = 2
-----
large reward = 10
-----
gamma = 0.999
-----
V* = [1832.59229299 1831.75441658 1833.83693248 1836.44268311 1839.70313185
      1843.78277241 1848.88742776 1855.27463416 1863.26663416 1873.26663416]
```

Large = 10

```
number of states = 5
-----
slip chance = 0.2
-----
small reward = 2
-----
large reward = 10
-----
gamma = 0.9
-----
V* = [32.473856 34.206336 39.390336 46.590336 56.590336]
```

```
number of states = 5
-----
slip chance = 0.2
-----
small reward = 2
-----
large reward = 10
-----
gamma = 0.99
-----
V* = [435.0601257 438.02805658 444.30069658 452.22069658 462.22069658]
```

```
number of states = 5
-----
slip chance = 0.2
-----
small reward = 2
-----
large reward = 10
-----
gamma = 0.999
-----
V* = [4481.24055962 4484.34521497 4490.73242137 4498.72442137 4508.72442137]
```

Large = 100

```
number of states = 5
-----
slip chance = 0.2
-----
small reward = 2
-----
large reward = 100
-----
gamma = 0.9
-----
V* = [274.33856 309.66336 361.50336 433.50336 533.50336]
```

```
number of states = 5
-----
slip chance = 0.2
-----
small reward = 2
-----
large reward = 100
-----
gamma = 0.99
-----
V* = [3976.20125696 4023.88056576 4086.60696576 4165.80696576 4265.80696576]
```

```
number of states = 5
-----
slip chance = 0.2
-----
small reward = 2
-----
large reward = 100
-----
gamma = 0.999
-----
V* = [41198.0055962 41247.05214975 41310.92421375 41390.84421375
41490.84421375]
```

Slip_chance = 0.2:

```
number of states = 5
-----
slip chance = 0.2
-----
small reward = 2
-----
large reward = 10
-----
gamma = 0.99
-----
V* = [435.0601257  438.02805658 444.30069658 452.22069658 462.22069658]
```

```
number of states = 5
-----
slip chance = 0.2
-----
small reward = 2
-----
large reward = 10
-----
gamma = 0.999
-----
V* = [4481.24055962 4484.34521497 4490.73242137 4498.72442137 4508.72442137]
```

Slip_chance = 0.05:

```
number of states = 5
-----
slip chance = 0.05
-----
small reward = 2
-----
large reward = 10
-----
gamma = 0.999
-----
V* = [8214.4310878  8220.97914226 8229.98610129 8239.47660129 8249.47660129]
```

```

number of states = 5
-----
slip chance = 0.05
-----
small reward = 2
-----
large reward = 10
-----
gamma = 0.99
-----
V* = [794.31145387 800.63055492 809.47595742 818.88095742 828.88095742]

```

```

number of states = 5
-----
slip chance = 0.05
-----
small reward = 2
-----
large reward = 10
-----
gamma = 0.9
-----
V* = [56.33975506 60.59001881 67.90026881 76.45026881 86.45026881]

```

و اما نتایج Average Reward عبارتند از:

	Gamma=0.9	Gamma=0.99	Gamma=0.9999
N=5	3.63	3.69	3.74
N=10	1.66	1.64	1.63
Slip=0.05	7.83	7.92	7.68
Slip=0.2	3.65	3.67	3.71
Large=10	3.51	3.71	3.75
Large=100	32.17	34.06	33.1

هنگامیکه N را زیاد می شود، بهتر است از گاما کمتر استفاده شود.

هنگامیکه large زیاد می شود بهتر است گاما را کم کنیم.

هنگامیکه slip کم میشود بهتر است کمی گاما را کم کنیم.

(ج)

	N=5	N=10	N=1000
Large=10	3.4	1.62	1.54
Large=100	32.69	12.62	1.63
Gamma=0.9	3.5	1.61	1.55
Gamma=0.9999	3.44	1.78	1.35

نتایج Average Reward عبارتند از:

در حالت کلی هر چه تعداد استیت ها کمتر باشد، بهتر است.

(د)

	Small=2 Large=10	Small=2 Large=100	Small=4 Large=10	Small=4 Large=100
N=5	3.68	34.2	3.96	34.81
N=20	1.55	1.61	3.23	3.2

در حالت کلی زمانی که تعداد استیت ها بیشتر می شود، اگر پاداش کوچک بیشتر باشد، نتیجه بهتری خواهیم گرفت.

اما اگر تعداد استیت ها کم باشد، اگر پاداش بزرگ بیشتر باشد بهتر است، البته اگر پاداش کوچک هم بزرگ باشد بهترین نتیجه را خواهیم گرفت.

(ه)

خیر در بعضی شرایط، ممکن است به عقب رفتن و ماندن در استیت ابتدایی حرکت بهتری باشد.