به نام خدا

دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده برق و کامپیوتر

درس سیستم‌های هوشمند

تمرین شماره ۵

نام و نام خانوادگی : سیاوش شمس

شماره دانشجویی : ۸۱۰۱۹۷۶۴۴

دی ۱۴۰۰

فهرست سوالات

سوال ۱

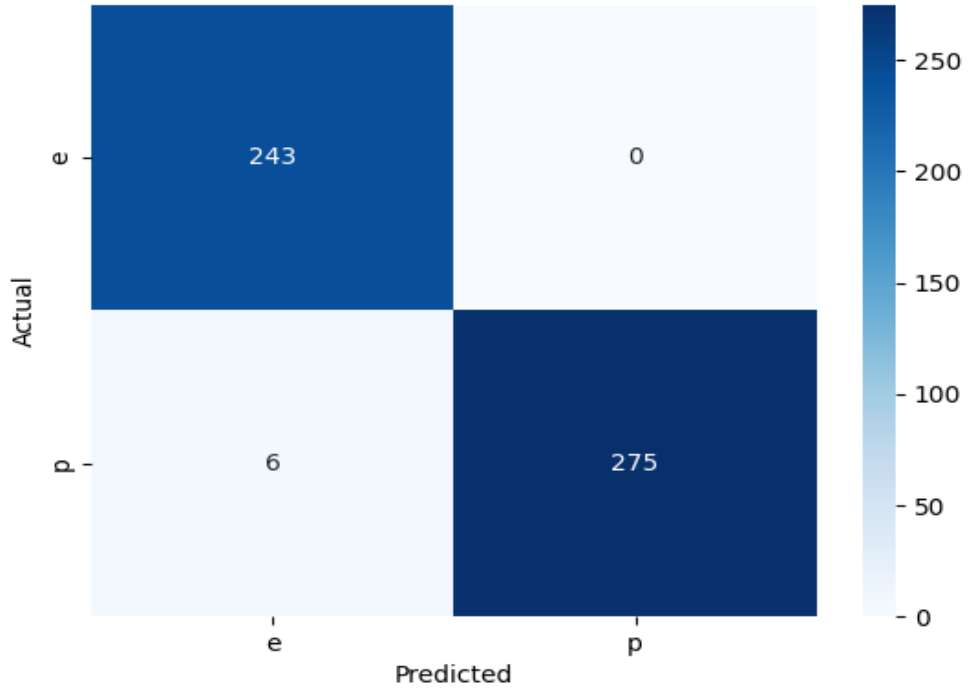
در این سوال با استفاده از روش بیز ساده لوح[1]، ابتدا مقادیر احتمالات قبل[2] و راست نمایی [3]را به کمک داده های آموزش به دست می آوریم سپس احتمال پسین[4] را بر روی داده های تست حساب کرده و با توجه به آن داده های تست را طبقه بندی می کنیم. در نهایت دقت و ماتریس آشفتگی را گزارش می کنیم.

الف:

```
Accuracy: 0.9885496183206107
Predicted    e    p
Actual
e          243    0
p            6  275
```

شکل۱-۱- دقت و ماتریس آشفتگی برای داده های تست

ب:



شکل۱-۲- ماتریس آشفتگی برای داده های تست

[1] Naïve Bayes
[2] Prior
[3] Likelihood
[4] Posterior

سوال ۲

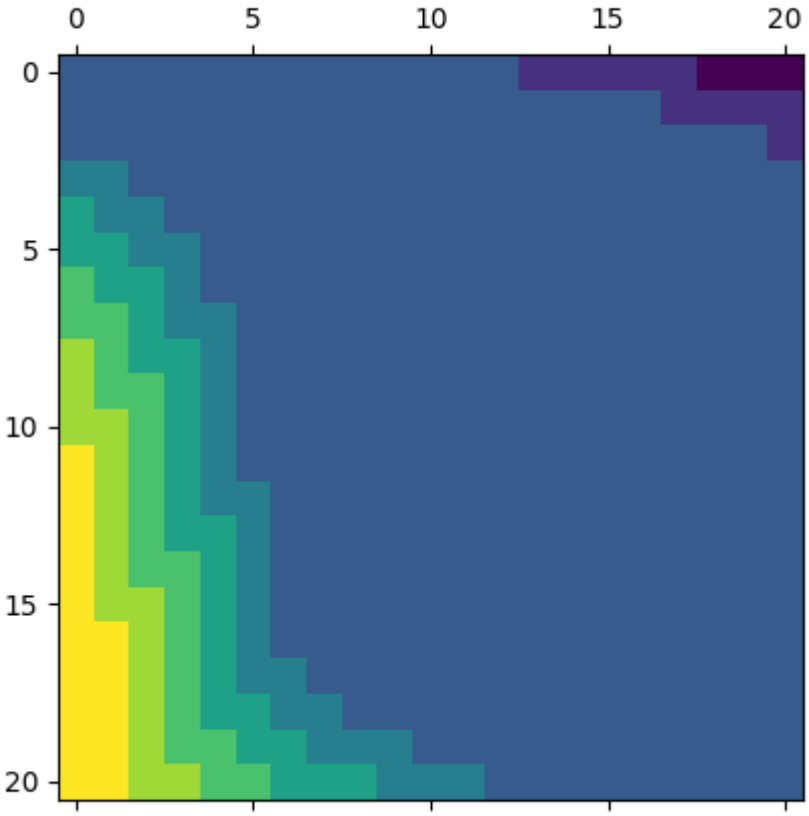
الف:

توضیح رویکرد حل سوال: حالت [1]ها را تعداد ظرفیت خالی هر شرکت در نظر می گیریم که می تواند عددی بین ۰ تا ۲۰ برای هر شرکت باشد. اعمال [2]را برابر خرید ظرفیت اضافه از شرکت رقیب در نظر می گیریم که برای هر شرکت می تواند عددی بین ۰ و ۵ باشد. بنابراین مسئله بهینه سازی ما به دنبال پیدا کردن تعداد ظرفیتی که هر شرکت باید از شرکت رقیب در هر حالت بخرد تا سود آن شرکت بیشینه شود می باشد

```
[[ 0  0  0  0  0  0  0  0  0  0  0  0  0 -1 -1 -1 -1 -1 -2 -2 -2]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 -1 -1 -1 -1]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 -1]
 [ 1  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 2  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 2  2  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 3  2  2  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 3  3  2  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 4  3  2  2  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 4  3  3  2  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 4  4  3  2  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 5  4  3  2  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 5  4  3  2  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 5  4  3  2  2  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 5  4  3  3  2  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 5  4  4  3  2  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 5  5  4  3  2  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 5  5  4  3  2  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 5  5  4  3  2  2  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 5  5  4  3  3  2  2  1  1  1  0  0  0  0  0  0  0  0  0  0  0]
 [ 5  5  4  4  3  3  2  2  2  1  1  1  0  0  0  0  0  0  0  0  0]]
```

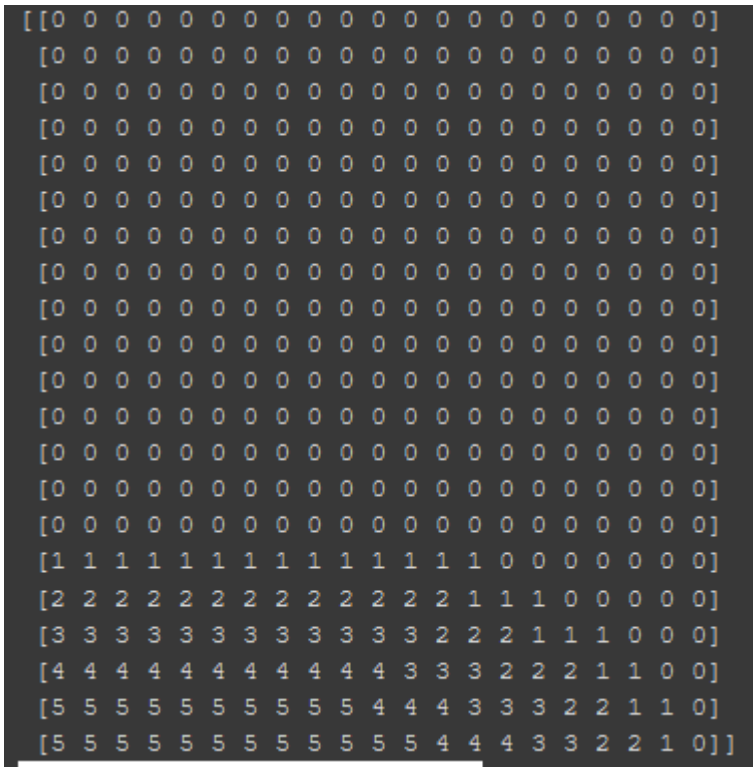
شکل۲-۱- سیاست [3]های خرید و فروش ظرفیت در هر حالت به ازای ضریب تخفیف 0.9[4]

[1] State
[2] Action
[3] Policy
[4] Discount factor



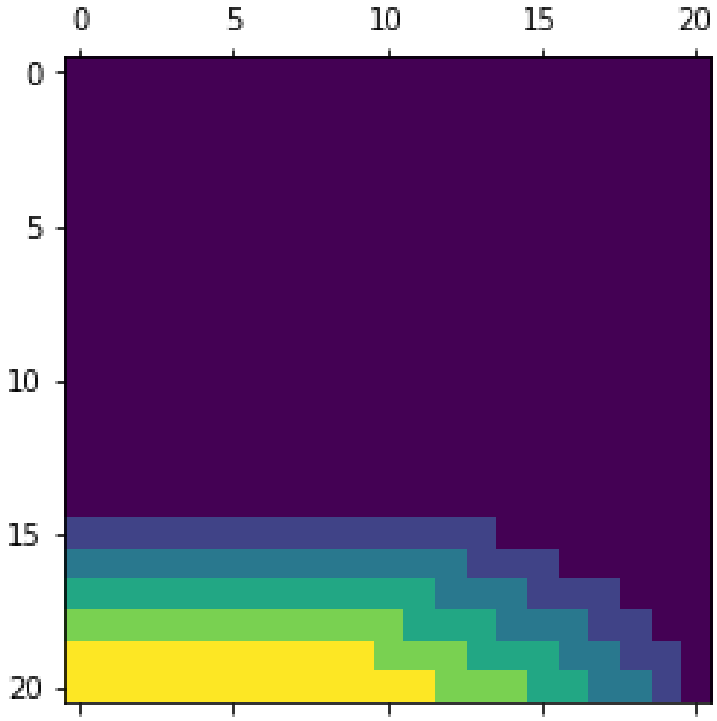
شکل۲-۲- سیاست های خرید و فروش ظرفیت در هر حالت به ازای ضریب تخفیف 0.9

تفسیر نتیجه بالا: جدول های بالا مشخص می کند در هر حالت هر شرکت برای بیشینه کردن سود خود باید چه تعداد ظرفیت از شرکت رقیب خود بخرد، مثلا در حالتی که شرکت B ظرفیت خالی ندارد و شرکت A دارای 20 ظرفیت خالی است بهترین کار خرید 5 ظرفیت از شرکت A می باشد.

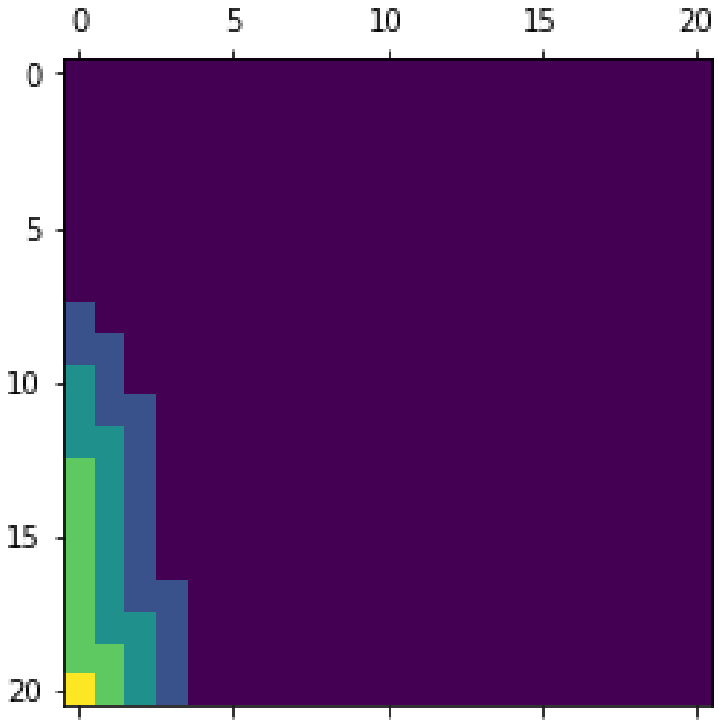


```
[[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0]
 [2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 0 0 0 0 0]
 [3 3 3 3 3 3 3 3 3 3 3 3 2 2 2 1 1 1 0 0 0]
 [4 4 4 4 4 4 4 4 4 4 4 3 3 3 2 2 2 1 1 0 0]
 [5 5 5 5 5 5 5 5 5 5 4 4 4 3 3 3 2 2 1 1 0]
 [5 5 5 5 5 5 5 5 5 5 5 5 4 4 4 3 3 2 2 1 0]]
```

شکل۲-۳- سیاست های خرید و فروش ظرفیت در هر حالت به ازای ضریب تخفیف 1



شکل۲-۴- سیاست های خرید و فروش ظرفیت در هر حالت به ازای ضریب تخفیف 1



```
[[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [2 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [2 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [3 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [3 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [3 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [3 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [3 2 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [3 2 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [3 3 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [4 3 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]]
```

شکل۲-۵- سیاست های خرید و فروش ظرفیت در هر حالت به ازای مبلغ خرید ظرفیت ۶

شکل۲-۶- سیاست های خرید و فروش ظرفیت در هر حالت به ازای مبلغ خرید ظرفیت ۶

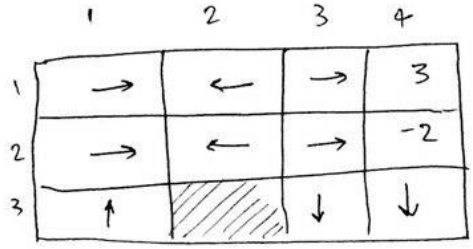
می بینیم که با تغییر هزینه ظرفیت اضافی، خرید و فروش بین دو شرکت کاهش می یابد.

ب:

در این سوال از روش تکراری (Iterative) و تحلیلی استفاده می‌کنیم

روش ۱: تکراری: از یک سیاست اولیه دلخواه شروع می‌کنیم.

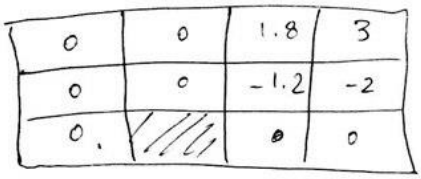
تکرار اول:



$V(1,3) = 0.6 \times 3 + 0.2 \times 0 + 0.2 \times 0 = 1.8$

$V(2,3) = 0.6 \times (-2) + 0.2 \times (0) + 0.2(0) = -1.2$

V بقیه $= 0$



Improve policy:

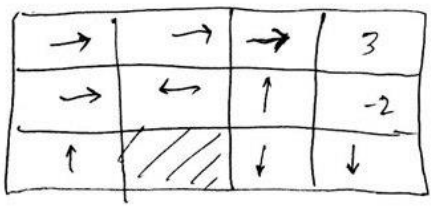
$\pi_1(1,2) = \begin{cases} \uparrow : 0.6(0) + 0.2(0) + 0.2(1.8) = 0.3 \\ \rightarrow : 0.6(1.8) + 0.2(0) + 0.2(0) = 1.08 \\ \leftarrow : 0 \\ \downarrow : = \uparrow = 0.36 \end{cases}$

$\pi_1(2,2) = \begin{cases} \rightarrow : 0.6(-1.2) + 0.2(0) + 0.2(0) = -0.72 \\ \uparrow = \downarrow : 0.6(0) + 0.2(0) + 0.2(-2) = -0.4 \\ \leftarrow : 0 \end{cases}$

$\pi_1(2,3) = \begin{cases} \rightarrow : 0.6(-2) + 0.2(1.8) + 0.2(0) = -0.84 \\ \uparrow : 0.6(1.8) + 0.2(-2) + 0.2(0) = 0.68 \\ \downarrow : 0.6(0) + 0.2(-2) + 0.2(0) = -0.4 \\ \leftarrow : 0.2(1.8) + 0 = 0.36 \end{cases}$

بقیه ↑ ها تغییر می‌کند

updated policy

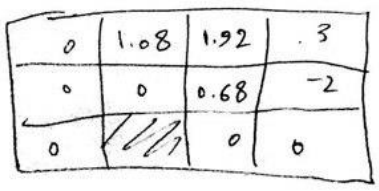


$V(1,3) = 0.6 \times 3 + 0.2(1.8) + 0.2(-1.2) = 1.92$

$V(2,3) = 0.6 \times 1.8 + 0.2(-2) + 0.2(0) = 0.68$

$V(1,2) = 0.6 \times 1.8 + 0.4(0) = 1.08$

V بقیه $= 0$



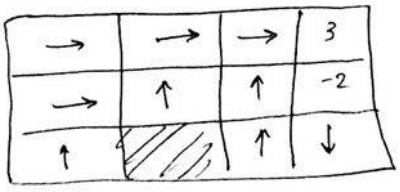
$\pi(1,1) = \begin{cases} \uparrow \\ \rightarrow \checkmark \\ \downarrow \\ \leftarrow \end{cases}$

$\pi(2,2) = \begin{cases} \uparrow \checkmark \\ \rightarrow \\ \downarrow \\ \leftarrow \end{cases}$

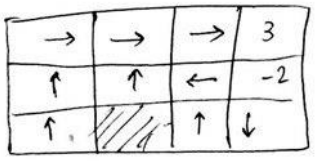
$\pi(3,3) = \begin{cases} \uparrow \checkmark \\ \leftarrow \\ \downarrow \\ \rightarrow \end{cases}$

بقیه π ها بی‌تغییر

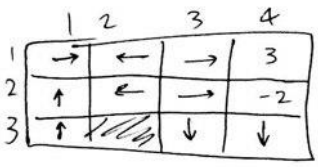
سیاست نمایی بعد از 2 تکرار :



اگر الگوریتم را بیشتر تکرار کنیم سیاست نمایی به صورت زیر می شود :



روش 2 : تحلیلی



حالت اولیه

$$\begin{cases} V(1,3) = 0.6\times 3 + 0.2\,V(1,3) + 0.2\,V(2,3) \\ V(2,3) = 0.6(-2) + 0.2\,V(1,3) + 0.2\,V(3,3) \\ V(3,3) = 0.6\,V(3,3) + 0.2\,V(3,3) + 0.2\,V(3,4) \\ V(3,4) = 0.6\,V(3,4) + 0.2\,V(3,4) + 0.2\,V(3,3) \end{cases}$$

حل دستگاه معادلات ←

$$\begin{cases} V(1,3) = 2.05 \\ V(2,3) = -0.79 \\ V(3,3) = 0 \\ V(3,4) = 0 \end{cases}$$

Improve Policy:

$\pi_1(2,3)$: ↑

$\pi_1(1,2)$: →

$\pi_1(2,2)$: ←

تکرار دوم ←

$$\begin{cases} V(1,3) = 0.6\times 3 + 0.2\,V(1,3) + 0.2\,V(2,3) \\ V(2,3) = 0.6\,V(1,3) + 0.2(-2) + 0.2\,V(2,2) \\ V(2,2) = 0.6\,V(2,1) + 0.2\,V(1,2) + 0.2\,V(2,2) \end{cases}$$

۷ معادله ۷ مجهول

سوال ۳

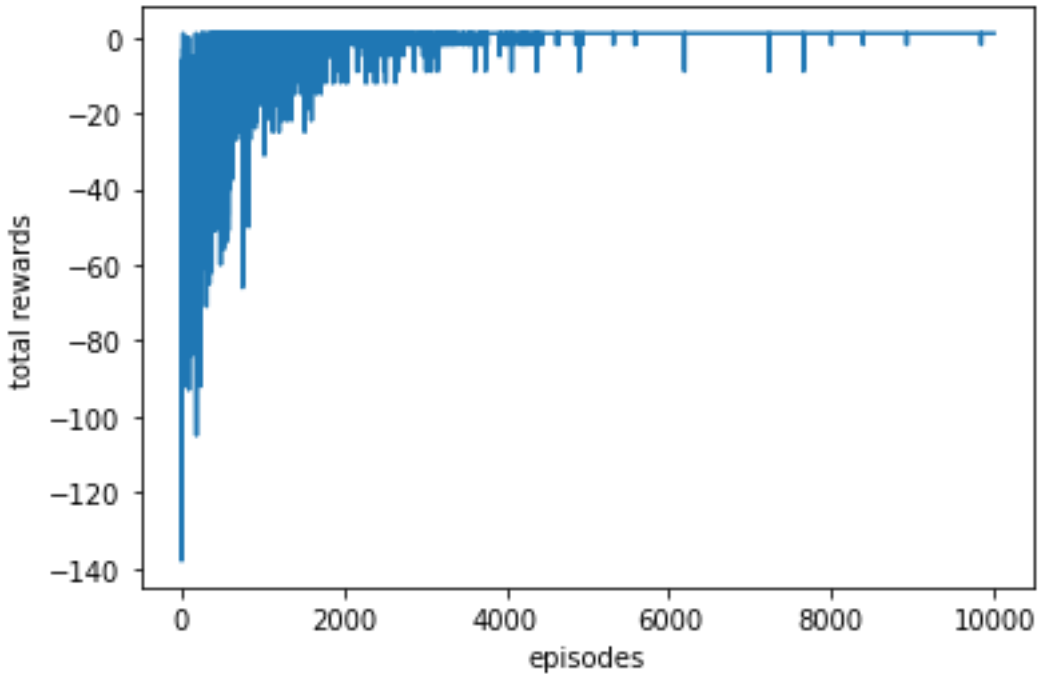
در این سوال الگوریتم Q-learning را پیاده سازی می کنیم تا عامل [1]یاد بگیرد تا به هدف برسد. برای الگوریتم، نرخ یادگیری [2]0.5 و ضریب تخفیف[3]0.99 در نظر می گیریم تا اثر تصمیمات اشتباه کم رنگ نشود و الگوریتم سعی در حداکثر کردن پاداش طولانی مدت داشته باشد. و برای اینکه عامل فرصت جست و جوی کل فضا را داشته باشد هر عمل به صورت احتمالی انجام می شود و هر چه تعداد تکرار بیشتر شود، احتمال انجام شدن عمل بهینه بیشتر می شود و الگوریتم به صورت حریصانه [4]اعمال را انتخاب می کند. احتمال انتخاب عمل تصادفی در هر مرحله به صورت ε^n در نظر گرفته می شود که n همان شماره تکرار و ε برابر 0.999 در نظر گرفتیم تا در ابتدا فضا به خوبی جست و جو شود. همچنین برای اینکه احتمال جست و جو صفر نشود، حداقل احتمال را برابر 0.001 در نظر می گیریم. الگوریتم را ۱۰۰۰۰ بار تکرار می کنیم.

$$Q^{new}(s_t, a_t) \leftarrow \underbrace{Q(s_t, a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \overbrace{\Big(\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1}, a)}_{\text{estimate of optimal future value}} - \underbrace{Q(s_t, a_t)}_{\text{old value}} \Big)}^{\text{temporal difference}}$$
$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}_{\text{new value (temporal difference target)}}$$

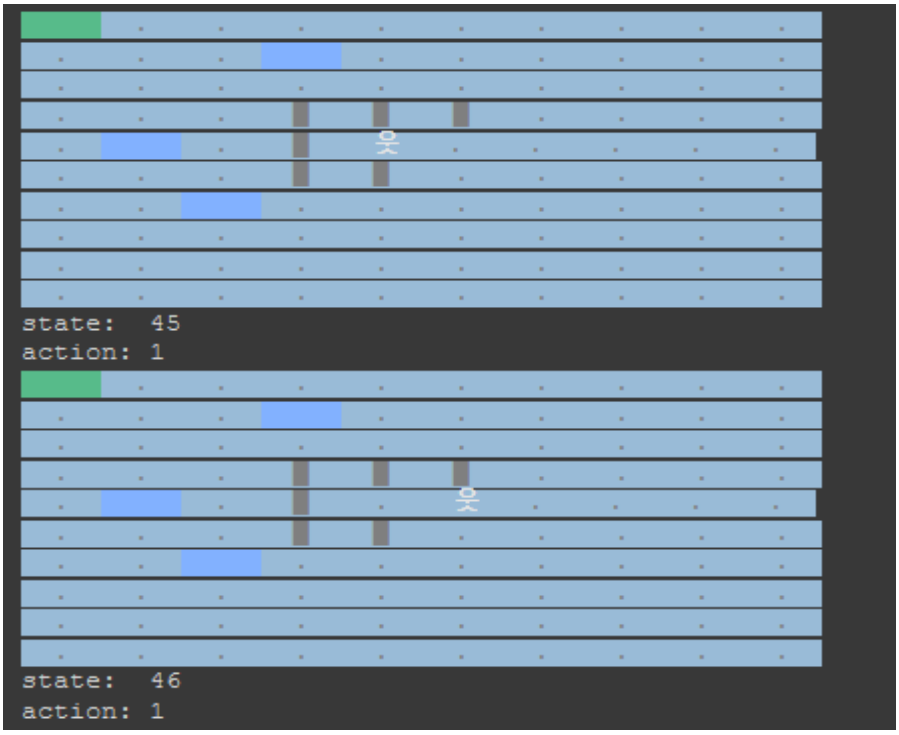
[1] Agent
[2] Learning rate
[3] Discount factor
[4] Greedy

```
[ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
[ 2.65025376e-02  5.40812929e-02  1.12198147e-01  0.00000000e+00]
[-6.18750000e+00  5.30187993e-01  3.11243256e-01  2.67702400e-02]
[ 5.50420118e-01  5.73324316e-01  5.44025030e-01  4.21065239e-01]
[ 5.79017953e-01  5.79163875e-01  5.67300759e-01  5.67159599e-01]
[ 5.85014015e-01  5.85014010e-01  5.73370508e-01  5.73371977e-01]
[ 5.90923248e-01  5.79163875e-01  5.79163845e-01  5.79163870e-01]
[ 5.85014015e-01  5.73372227e-01  5.73372192e-01  5.85013980e-01]
[ 5.79100770e-01  5.67149989e-01  5.67134067e-01  5.79163873e-01]
[ 5.73357269e-01  5.67153756e-01  5.55378606e-01  5.72746422e-01]
[ 0.00000000e+00  1.08885330e-01  0.00000000e+00  0.00000000e+00]
[ 2.67702400e-02  2.18795657e-01  3.96037565e-02  4.28834456e-02]
[ 4.01741405e-01  4.89983002e-01  0.00000000e+00  1.37181591e-01]
[ 4.03280977e-01  5.67039486e-01  4.33042869e-01  3.05654894e-01]
[ 5.73371329e-01  5.71931766e-01  5.38274059e-01  5.56236614e-01]
[ 5.79163779e-01  5.79159193e-01  5.63300908e-01  5.65681241e-01]
[ 5.85014010e-01  5.72996197e-01  5.72987796e-01  5.73371260e-01]
[ 5.79163875e-01  5.42919347e-01  5.01515718e-01  5.78975793e-01]
[ 5.47292652e-01  5.08718692e-01  3.53907780e-01  5.73329712e-01]
[ 5.67249361e-01  4.48513405e-01  3.64868114e-01  5.03641750e-01]
[ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
[ 1.48307645e-01  0.00000000e+00  0.00000000e+00  0.00000000e+00]
[ 0.00000000e+00  1.52288692e-01  0.00000000e+00  0.00000000e+00]
[ 4.16849665e-01  5.12682241e-01  0.00000000e+00  0.00000000e+00]
[ 5.67346729e-01  5.16369184e-01  5.34860144e-01  3.83033669e-01]
[ 5.71903760e-01  5.70212486e-01  5.41434850e-01  5.58953956e-01]
[ 5.79140234e-01  4.34668165e-01  5.38607139e-01  5.50621010e-01]
[ 5.48068146e-01  3.48494346e-01  3.04975665e-01  5.00592416e-01]
[ 4.95952240e-01  1.51647904e-01  2.07735872e-01  2.94130322e-01]
```

شکل۳-۱- بخشی از جدول Q بعد از آموزش



شکل۳-۲- نمودار پاداش بر حسب هر تکرار بازی

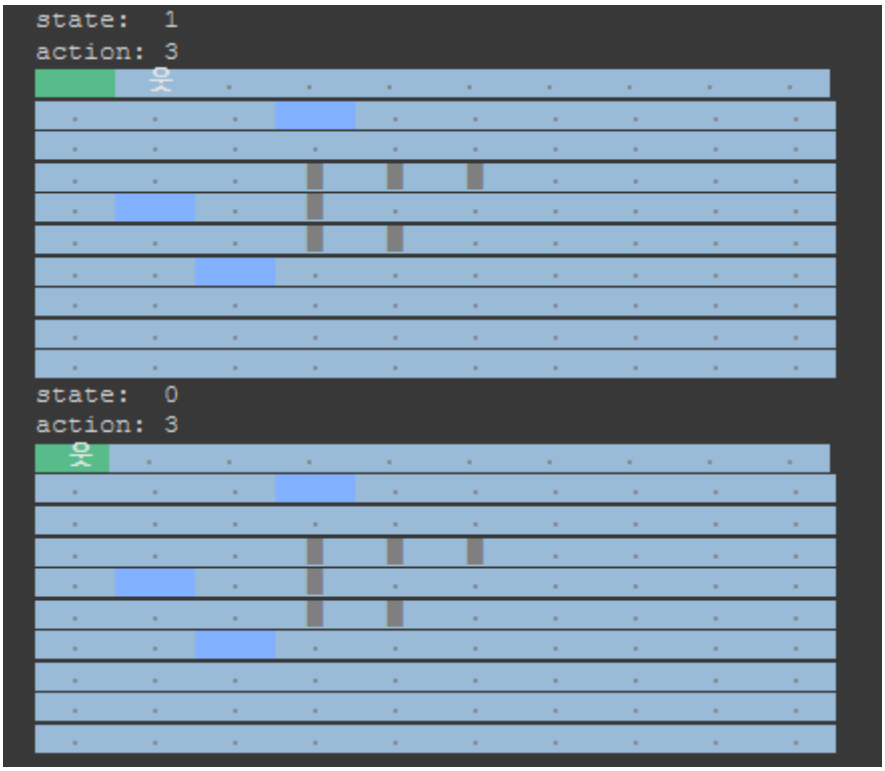


شکل۳-۳- دو تکرار اولیه بازی

.

.

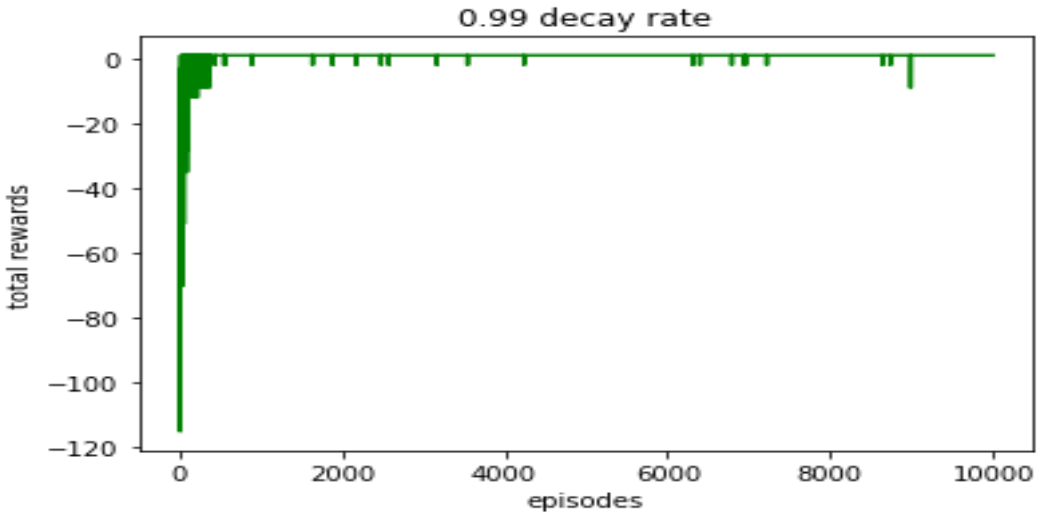
.



شکل۳-۴- دو تکرار نهایی بازی

برای سریع تر همگرا شدن الگوریتم چند ایده را بررسی می کنیم:

۱. بررسی تاثیر ε



شکل۳-۵- نمودار پاداش بر حسب هر تکرار بازی با نرخ کاهشی 0.99

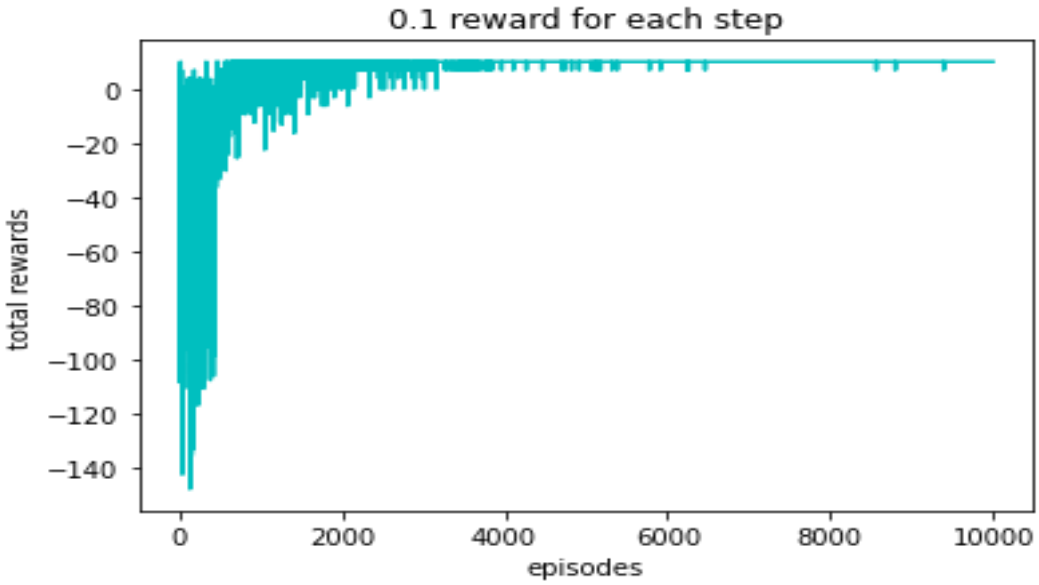
با توجه به نمودار بالا می بینیم که الگوریتم ما خیلی سریعتر به رفتار بهینه همگرا می شود، در نتیجه کاهش ε باعث سریعتر همگرا شدن الگوریتم شده است، البته این روش در محیط های پیچیده تر مناسب نیست زیرا ممکن است محیط به خوبی جست و جو نشود و بهینه ترین حالت پیدا نشود.

۲. تغییر پاداش نهایی

شکل۳-۶- نمودار پاداش بر حسب هر تکرار بازی با پاداش نهایی برابر ۱۰

با مقایسه نمودار شکل ۳-۶ با ۳-۲ می بینیم که تغییر پاداش نهایی تاثیری زیادی در سرعت همگرایی نداشت.

۳. در نظر گرفتن پاداش 0.1- برای هر گام



شکل۳-۷- نمودار پاداش بر حسب هر تکرار بازی با پاداش -۰.۱ برای هر گام

با مقایسه نمودار شکل ۳-۷ و نمودار شکل ۳-۲ می بینیم که در نظر گرفتن پاداش کوچک برای گام های بدون مجازات به سرعت همگرایی کمک می کند.

*نکته: سایر تغییرات از جمله زیاد کردن مجازات برخورد با سخره یا آب تاثیری در سرعت همگرایی نداشت.

پس نتیجه ای که می گیریم این است که نرخ کاهشی ε بیشترین تاثیر در سرعت همگرایی را دارد و در نظر گرفتن پاداش 0.1- به دلیل در نظر گرفتن مجازات برای تعداد گام باعث سریعتر همگرا شدن به جواب می شوند.