

به نام خدا



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



درس سیستم‌های هوشمند

پروژه پایانی

نام و نام خانوادگی : سیاوش شمس، محمدمهدی رحیمی، محمدرضا بختیاری

شماره دانشجویی : ۸۱۰۱۹۷۶۴۴، ۸۱۰۱۹۷۵۱۰، ۸۱۰۱۹۷۴۶۸

بهمن ۱۴۰۰

فهرست سوالات

مدلسازی	۳
پیاده سازی مسئله مسیریابی ربات	۴
سوال ۱	۴
سوال ۲	۵
سوال ۳	۶
سوال ۴	۷
سوال ۵	۸
سوال ۶	۱۰
سوال ۷	۱۳

۱. استیت: جمعیت خرچنگ ها

اکشن: شکار یا عدم شکار خرچنگ در آن ماه

احتمال انتقال بین استیت ها: احتمال کم یا زیاد شدن یا ثابت ماندن جمعیت بعد از انجام اکشن انتخاب شده

ریوارد: سود به دست آمده از فروش خرچنگ یا هزینه برای جبران جمعیت در صورت پایین تر آمدن جمعیت از حدی.

۲. استیت: ارزش سهام در ۷ مرحله که می توانیم در آن اکشن انجام دهیم

اکشن: خرید یا فروش سهام

احتمال انتقال بین استیت ها: احتمال افزایش یا کاهش ارزش سهام در استیت بعدی ریوارد: سود یا ضرر حاصل از تصمیم

۳. استیت: میزان تقاضا (d_i)

اکشن: تعیین میزان تولید (n_i تعداد کالا تولیدی در استیت i ام)

احتمال انتقال بین استیت ها: احتمال داشتن n_j مشتری در استیت بعدی

ریوارد: برای هر محصول تولیدی اضافه (خریده نشده) پاداش $r_i = -\alpha(n_i - d_i)$ و برای هر محصول نا کافی $r_i = -2\alpha(d_i - n_i)$ که α قیمت محصول است در نظر می گیریم.

۴. استیت: هر استیت را می توان به صورت دوتایی $s_i = (n_i, c_i)$ در نظر گرفت که n_i مکان های

منتظر سرویس که خود شامل نوع منطقه، بلندی آژیر و پیش فرض جمعیتی آن منطقه می باشد و c_i تعداد ماشین های موجود می باشد

اکشن: تعیین تعداد ماشین ارسالی به هر منطقه a_i

احتمال انتقال بین استیت ها: استیت بعدی شامل مکان های جدید آتش گرفته n_j و تعداد ماشین

موجود c_j در نظر گرفت که احتمال رفتن به هر استیت با توجه به اکشن انتخابی a_i و توزیعات

احتمالاتی ایجاد مکان آتش گرفته جدید و برگشت تعداد ماشین در نظر گرفت

$$p_{ij} = p(n_j | a_i, s_i) \cdot p(c_j | a_i, s_i)$$

ریوارد: در نظر گرفتن پاداش مثبت مثلا 1+ برای هر عملیات موفق و در نظر گرفتن پاداش 10-

برای هر عملیات نا موفق مربوط به ساختمان و 50- برای عملیات نا موفق مربوط به جان انسان، در

صورت به صدا در آمدن آژیر و نبودن ماشین هم پاداش 100- در نظر می گیریم.

پیاده سازی مسئله مسیریابی ربات

سوال ۱

توضیح الگوریتم: در الگوریتم مونت کارلو به صورت کاملاً تصادفی دنباله ای از حرکات را انجام می دهیم تا هنگامی که عامل به مقصد برسد و در طول این مدت ریوارد ها و اکشن ها و استیت ها را ذخیره می کنیم، بعد از رسیدن عامل به مقصد ارزش خانه ها را طبق رابطه بیان شده در الگوریتم زیر به روز رسانی می کنیم.

First-visit MC prediction, for estimating $V \approx v_\pi$

Input: a policy π to be evaluated

Initialize:

$V(s) \in \mathbb{R}$, arbitrarily, for all $s \in \mathcal{S}$

$Returns(s) \leftarrow$ an empty list, for all $s \in \mathcal{S}$

Loop forever (for each episode):

Generate an episode following π : $S_0, A_0, R_1, S_1, A_1, R_2, \dots, S_{T-1}, A_{T-1}, R_T$

$G \leftarrow 0$

Loop for each step of episode, $t = T-1, T-2, \dots, 0$:

$G \leftarrow \gamma G + R_{t+1}$

Unless S_t appears in S_0, S_1, \dots, S_{t-1} :

Append G to $Returns(S_t)$

$V(S_t) \leftarrow \text{average}(Returns(S_t))$

شکل ۱-۱ الگوریتم مونت کارلو برای پیدا کردن ارزش هر استیت

جدول ۱-۱: ارزش پیدا شده برای هر خانه بعد از ۱۰۰۰ تکرار

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	1000.00000	689.99734	576.64923	474.38957	423.36073	398.04702	0.00000	0.00000	64.55826	75.38411	84.74905	75.50303	85.08468	75.23522	70.97757
1	712.56629	667.18769	540.80466	487.44538	420.64184	399.00580	0.00000	0.00000	85.31433	95.42914	93.70961	93.64199	85.60985	90.09044	72.52186
2	547.94917	540.16339	500.09273	444.77979	413.29629	373.68002	0.00000	0.00000	105.58934	116.76874	107.00654	98.26861	91.65755	92.53576	78.56718
3	446.77352	455.92347	430.30620	415.07525	386.83986	341.63009	0.00000	0.00000	163.76508	145.35987	121.62225	106.24187	101.31201	91.51531	84.05532
4	393.56472	392.42902	401.61401	380.80029	353.04728	320.29668	265.68941	222.31889	194.30647	159.78458	136.29916	120.63623	104.52210	98.03010	80.74201
5	318.19904	358.92521	361.72055	351.85350	322.88506	302.91039	291.44664	248.93402	204.38744	175.69514	152.82439	130.55478	115.16895	100.53201	79.78805
6	310.43657	309.16476	332.98696	315.86375	297.34485	301.62203	282.12656	242.24376	219.15686	194.82612	155.09525	135.83082	114.56850	90.01031	74.92269
7	262.44054	292.20080	274.11055	298.44186	279.81946	280.93955	269.25100	245.59773	246.69663	232.51705	183.95157	137.12869	0.00000	0.00000	0.00000
8	218.58450	238.91988	255.56678	249.17147	245.75989	253.59458	267.84261	262.09369	247.94446	221.52758	210.73828	162.75750	0.00000	0.00000	0.00000
9	187.90455	217.48516	218.82446	223.52691	226.00186	222.16046	243.29256	239.58315	232.11828	216.80536	191.96837	166.06092	112.41186	69.03474	60.61808
10	179.83840	186.94350	191.99074	198.28715	199.31647	205.64268	208.75792	229.62239	193.13202	175.15896	190.93323	120.08474	102.01209	71.36536	66.93656
11	161.18839	162.53535	170.80896	171.94270	172.72342	0.00000	0.00000	154.36033	156.59084	162.43308	156.58095	125.57747	84.44759	79.99732	68.79599
12	141.48473	148.29580	154.30939	144.46354	136.28247	0.00000	0.00000	99.11628	112.29917	133.93760	108.08500	130.55150	98.16888	57.79401	60.20904
13	141.32559	148.34975	144.38728	134.30477	124.73289	0.00000	0.00000	88.18393	102.69277	111.41052	106.08304	95.98727	65.72454	45.41819	39.20068
14	131.93660	138.59494	137.53791	117.42295	113.41903	0.00000	0.00000	74.06987	71.10124	82.48348	102.48031	75.86057	74.77237	60.73434	0.00000

مقدار ضریب تخفیف برابر ۰.۹۹ در نظر گرفته شده است.

جدول ۱-۲: سیاست بهینه پیدا شده برای هر خانه بعد از ۱۰۰۰ تکرار

	± 0	± 1	± 2	± 3	± 4	± 5	± 6	± 7	± 8	± 9	± 10	± 11	± 12	± 13	± 14
0	-	W	W	W	SW	W	-	-	SE	S	SW	SW	SW	S	SW
1	N	NW	NW	NW	W	NW	-	-	SE	S	SW	SW	SW	S	SW
2	N	NW	NW	NW	NW	NW	-	-	S	SW	SW	SW	SW	SW	W
3	N	NW	NW	NW	NW	NW	-	-	SW	SW	SW	SW	SW	SW	SW
4	NE	N	NW	NW	NW	NW	NW	SW	SW	SW	SW	SW	SW	SW	SW
5	N	NE	N	NW	NW	NW	NW	W	W	SW	SW	SW	SW	W	W
6	NE	NE	N	NW	NW	NW	NW	NW	NW	SW	SW	SW	SW	NW	NW
7	N	NE	N	NW	NW	N	NW	NW	SW	SW	W	SW	-	-	-
8	NE	N	NE	N	NW	N	NW	NW	W	W	NW	W	-	-	-
9	NE	NE	N	NW	NE	NE	N	NW	NW	NW	NW	NW	W	W	SW
10	NE	NE	NE	NE	N	NE	N	NW	NW	NW	NW	NW	NW	NW	SW
11	NE	NE	NE	NE	NE	-	-	N	NW	NW	N	NW	SW	NW	W
12	NE	NE	NE	NE	N	-	-	NE	NE	N	NW	NW	W	W	NW
13	E	NE	N	NW	NW	-	-	NE	NE	N	NW	N	NW	NW	SW
14	NE	N	NW	NW	NW	-	-	NE	NE	N	NW	NW	NW	W	W

با توجه به نتیجه بالا می بینیم الگوریتم مونت کارلو با وجود تصادفی بودن گام ها و ۱۰۰۰ تکرار در نهایت به سیاست نسبتاً خوبی همگرا می شود.

سوال ۲

در این سوال الگوریتم تکرار تصمیم به صورت زیر تا همگرا شدن سیاست اجرا می شود.

Policy Iteration (using iterative policy evaluation) for estimating $\pi \approx \pi_*$

1. Initialization

$V(s) \in \mathbb{R}$ and $\pi(s) \in \mathcal{A}(s)$ arbitrarily for all $s \in \mathcal{S}$

2. Policy Evaluation

Loop:

$\Delta \leftarrow 0$

Loop for each $s \in \mathcal{S}$:

$v \leftarrow V(s)$

$V(s) \leftarrow \sum_{s',r} p(s',r|s,\pi(s)) [r + \gamma V(s')]$

$\Delta \leftarrow \max(\Delta, |v - V(s)|)$

until $\Delta < \theta$ (a small positive number determining the accuracy of estimation)

3. Policy Improvement

policy-stable \leftarrow true

For each $s \in \mathcal{S}$:

old-action $\leftarrow \pi(s)$

$\pi(s) \leftarrow \operatorname{argmax}_a \sum_{s',r} p(s',r|s,a) [r + \gamma V(s')]$

If *old-action* $\neq \pi(s)$, then *policy-stable* \leftarrow false

If *policy-stable*, then stop and return $V \approx v_*$ and $\pi \approx \pi_*$; else go to 2

شکل ۱-۲ الگوریتم تکرار سیاست برای پیدا کردن سیاست بهینه

در همه قسمت های ۲ و ۳ و ۴ ضریب تخفیف برابر ۱ در نظر گرفته شده تا بهترین تصمیم در استیت های دور از مقصد انتخاب شود.

جدول ۲-۱: ارزش پیدا شده برای هر خانه بعد از اجرای الگوریتم تکرار سیاست

	± 0	± 1	± 2	± 3	± 4	± 5	± 6	± 7	± 8	± 9	± 10	± 11	± 12	± 13	± 14
0	1000.00000	999.98664	999.97367	999.96040	999.94301	999.87005	0.00000	0.00000	999.74363	999.80630	999.81075	999.81101	999.81073	999.80907	999.80174
1	999.98664	999.98560	999.97324	999.96015	999.94085	999.84446	0.00000	0.00000	999.73094	999.81653	999.82273	999.82291	999.82186	999.81409	999.80384
2	999.97369	999.97328	999.97135	999.95934	999.94052	999.84261	0.00000	0.00000	999.74693	999.83037	999.83608	999.83571	999.82764	999.81604	999.80430
3	999.96077	999.96069	999.95987	999.95687	999.94087	999.87036	0.00000	0.00000	999.79406	999.84650	999.84973	999.84135	999.82923	999.81627	999.80436
4	999.94787	999.94784	999.94762	999.94621	999.94014	999.89853	999.85355	999.83527	999.84912	999.86414	999.85510	999.84247	999.82913	999.81594	999.80408
5	999.93496	999.93495	999.93487	999.93440	999.93170	999.92368	999.90671	999.89272	999.88069	999.86902	999.85584	999.84040	999.82469	999.81044	999.79956
6	999.92206	999.92206	999.92203	999.92184	999.92106	999.91785	999.90973	999.89636	999.88320	999.86996	999.85449	999.81377	999.76950	999.72597	999.73752
7	999.90916	999.90915	999.90914	999.90907	999.90877	999.90768	999.90412	999.89602	999.88320	999.86995	999.85303	999.78476	0.00000	0.00000	0.00000
8	999.89625	999.89625	999.89622	999.89612	999.89591	999.89539	999.89402	999.89023	999.88224	999.86969	999.85290	999.78408	0.00000	0.00000	0.00000
9	999.88335	999.88334	999.88324	999.88239	999.88071	999.87892	999.87821	999.87801	999.87551	999.86821	999.85382	999.81179	999.76693	999.72292	999.73433
10	999.87044	999.87042	999.87020	999.86779	999.84037	999.81098	999.80919	999.83613	999.86209	999.86139	999.85340	999.83815	999.82163	999.80723	999.79625
11	999.85754	999.85751	999.85712	999.85239	999.79685	0.00000	0.00000	999.79110	999.84444	999.84853	999.84737	999.83937	999.82589	999.81261	999.80063
12	999.84463	999.84460	999.84409	999.83765	999.75288	0.00000	0.00000	999.74493	999.82912	999.83510	999.83524	999.83347	999.82557	999.81279	999.80088
13	999.83179	999.83178	999.83128	999.82454	999.73830	0.00000	0.00000	999.72973	999.81556	999.82194	999.82233	999.82203	999.81973	999.81197	999.80070
14	999.81968	999.81999	999.81966	999.81486	999.75172	0.00000	0.00000	999.74272	999.80554	999.81020	999.81051	999.81043	999.80980	999.80699	999.79963

جدول ۲-۲: سیاست بهینه برای هر خانه بعد از اجرای الگوریتم تکرار سیاست

	± 0	± 1	± 2	± 3	± 4	± 5	± 6	± 7	± 8	± 9	± 10	± 11	± 12	± 13	± 14
0	-	W	W	W	W	W	-	-	SE	SE	SE	S	SW	SW	SW
1	N	NW	NW	NW	NW	NW	-	-	SE	SE	S	SW	SW	SW	SW
2	N	NW	NW	NW	NW	SW	-	-	SE	SE	S	SW	SW	SW	SW
3	N	NW	NW	NW	NW	W	-	-	SE	S	SW	SW	W	W	W
4	N	NW	NW	NW	NW	NW	SW	SW	SW	SW	SW	SW	W	NW	NW
5	N	NW	NW	NW	NW	NW	W	SW	SW	SW	SW	W	NW	NW	NW
6	N	NW	NW	NW	NW	NW	NW	W	W	W	W	NW	NW	NW	NW
7	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	-	-	-
8	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	SW	-	-	-
9	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	W	SW	SW	SW
10	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	SW	SW	SW
11	N	NW	NW	NW	NW	-	-	NE	N	NW	NW	NW	W	W	SW
12	N	NW	NW	NW	NW	-	-	NE	NE	N	NW	NW	NW	NW	W
13	N	NW	NW	NW	NW	-	-	NE	NE	NE	N	NW	NW	NW	NW
14	N	NW	NW	NW	NW	-	-	NE	NE	NE	N	NW	NW	NW	NW

*جهت های بهینه به صورت قطب نمایی نشان داده شده اند مثلا NW بیانگر بالا چپ می باشد.

سوال ۳

جدول ۳-۱: ارزش پیدا شده برای هر خانه بعد از اجرای الگوریتم تکرار سیاست

	± 0	± 1	± 2	± 3	± 4	± 5	± 6	± 7	± 8	± 9	± 10	± 11	± 12	± 13	± 14
0	1000.00000	1000.00000	1000.00000	1000.00000	999.99995	999.99934	0.00000	0.00000	999.99908	999.99972	999.99977	999.99977	999.99977	999.99977	999.99976
1	1000.00000	1000.00000	1000.00000	999.99999	999.99993	999.99907	0.00000	0.00000	999.99891	999.99978	999.99984	999.99985	999.99985	999.99984	999.99983
2	1000.00000	1000.00000	1000.00000	999.99999	999.99993	999.99907	0.00000	0.00000	999.99899	999.99984	999.99990	999.99991	999.99990	999.99989	999.99985
3	1000.00000	1000.00000	1000.00000	999.99995	999.99938	0.00000	0.00000	999.99934	999.99990	999.99994	999.99994	999.99994	999.99993	999.99990	999.99986
4	1000.00000	1000.00000	1000.00000	999.99997	999.99969	999.99939	999.99938	999.99967	999.99994	999.99997	999.99995	999.99995	999.99993	999.99989	999.99985
5	1000.00000	1000.00000	1000.00000	999.99999	999.99997	999.99996	999.99995	999.99997	999.99998	999.99997	999.99993	999.99989	999.99984	999.99981	999.99981
6	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	999.99999	999.99999	999.99999	999.99996	999.99966	999.99933	999.99899	999.99919
7	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	999.99999	999.99994	999.99994	999.99937	0.00000	0.00000	0.00000
8	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	999.99999	999.99999	999.99994	999.99937	0.00000	0.00000	0.00000
9	1000.00000	1000.00000	1000.00000	999.99999	999.99997	999.99996	999.99996	999.99997	999.99998	999.99999	999.99996	999.99967	999.99934	999.99902	999.99924
10	1000.00000	1000.00000	1000.00000	999.99997	999.99969	999.99939	999.99938	999.99967	999.99998	999.99997	999.99994	999.99937	0.00000	0.00000	0.00000
11	1000.00000	1000.00000	1000.00000	999.99995	999.99938	0.00000	0.00000	999.99936	999.99992	999.99997	999.99997	999.99996	999.99995	999.99993	999.99991
12	1000.00000	1000.00000	999.99999	999.99993	999.99907	0.00000	0.00000	999.99904	999.99989	999.99996	999.99996	999.99996	999.99995	999.99993	999.99992
13	1000.00000	1000.00000	999.99999	999.99993	999.99907	0.00000	0.00000	999.99901	999.99988	999.99995	999.99995	999.99995	999.99995	999.99993	999.99992
14	1000.00000	1000.00000	1000.00000	999.99995	999.99934	0.00000	0.00000	999.99925	999.99989	999.99993	999.99994	999.99994	999.99994	999.99993	999.99992

جدول ۳-۲: سیاست بهینه برای هر خانه بعد از اجرای الگوریتم تکرار سیاست

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	-	W	W	W	W	W	-	-	SE	SE	SE	S	SW	SW	SW
1	N	NW	NW	NW	NW	NW	-	-	SE	SE	SE	S	SW	SW	SW
2	N	NW	NW	SW	SW	SW	-	-	SE	SE	SE	S	SW	SW	SW
3	N	NW	SW	SW	SW	SW	-	-	SE	SE	S	SW	SW	SW	W
4	N	NW	SW	SW	SW	SW	SW	SE	SE	S	SW	SW	W	W	NW
5	N	NW	SW	SW	SW	SW	SW	SW	SW	SW	SW	W	NW	NW	NW
6	N	NW	W	W	SW	SW	SW	SW	SW	SW	SW	NW	NW	NW	NW
7	N	NW	NW	NW	W	W	W	W	W	W	W	NW	-	-	-
8	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	SW	-	-	-
9	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	SW	SW	SW	SW
10	N	NW	NW	NW	NW	NW	NE	NE	NE	N	NW	W	SW	SW	SW
11	N	NW	NW	NW	NW	-	-	NE	NE	N	NW	NW	W	SW	SW
12	N	NW	NW	NW	NW	-	-	NE	NE	NE	N	NW	NW	W	W
13	N	NW	SW	SW	SW	-	-	NE	NE	NE	N	NW	NW	NW	NW
14	N	NW	W	W	W	-	-	E	NE	NE	N	NW	NW	NW	NW

سوال ۴

جدول ۴-۱: ارزش پیدا شده برای هر خانه بعد از اجرای الگوریتم تکرار سیاست

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	100.00000	98.66451	97.36922	96.07459	94.74850	92.98418	0.00000	0.00000	83.33772	83.88716	83.90633	83.84964	83.30194	82.37156	81.28412
1	98.66452	98.56059	97.32788	96.06482	94.72870	92.78145	0.00000	0.00000	84.38049	85.10047	85.09835	84.57806	83.65413	82.49794	81.36464
2	97.36944	97.32818	97.13938	95.98616	94.70776	92.75972	0.00000	0.00000	85.78050	86.45599	85.93596	84.98644	83.79748	82.55017	81.37929
3	96.07750	96.06904	95.99037	95.73024	94.60703	92.93512	0.00000	0.00000	87.87042	87.30778	86.32693	85.10751	83.83841	82.55850	81.38126
4	94.78680	94.78386	94.76417	94.65005	94.31361	93.02048	91.25023	89.79024	88.70463	87.66893	86.41666	85.13446	83.84080	82.55654	81.37899
5	93.49635	93.49553	93.48843	93.45496	93.30181	92.89383	91.65871	90.32516	89.00961	87.72356	86.43100	85.11930	83.80282	82.50759	81.33799
6	92.20599	92.20571	92.20354	92.19099	92.14169	91.95312	91.49760	90.32422	89.02611	87.73176	86.42071	84.90940	83.33661	81.73919	80.74435
7	90.91566	90.91556	90.91478	90.91064	90.89150	90.82537	90.60536	90.11192	88.99004	87.72197	86.40628	84.67993	0.00000	0.00000	0.00000
8	89.62533	89.62527	89.62480	89.62275	89.61526	89.58798	89.50458	89.25720	88.73414	87.65361	86.38465	84.65846	0.00000	0.00000	0.00000
9	88.33499	88.33490	88.33403	88.32685	88.31140	88.28871	88.25255	88.16302	87.90303	87.36113	86.29836	84.84067	83.15895	81.53154	80.50490
10	87.04466	87.04449	87.04254	87.02242	86.80527	86.57000	86.53973	86.71081	86.81366	86.55162	85.98679	84.93707	83.59745	82.26717	81.08685
11	85.75435	85.75410	85.75075	85.70960	85.23886	0.00000	0.00000	85.03495	85.47493	85.47981	85.20142	84.61970	83.59357	82.30315	81.12304
12	84.46446	84.46437	84.45994	84.40241	83.63327	0.00000	0.00000	83.39630	84.13409	84.17763	84.14903	83.85378	83.25916	82.26323	81.11232
13	83.18017	83.18271	83.17923	83.11748	82.33172	0.00000	0.00000	82.06307	82.83664	82.89199	82.88530	82.82820	82.51704	81.91801	81.03950
14	81.97153	82.00502	82.00300	81.95950	81.35443	0.00000	0.00000	81.07378	81.67304	81.71254	81.71289	81.69434	81.60845	81.27244	80.69670

جدول ۴-۲: سیاست بهینه برای هر خانه بعد از اجرای الگوریتم تکرار سیاست

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	-	W	W	W	W	W	-	-	SE	S	SW	SW	SW	SW	SW
1	N	NW	NW	NW	NW	NW	-	-	SE	S	SW	SW	SW	SW	SW
2	N	NW	NW	NW	NW	NW	-	-	S	SW	SW	SW	SW	SW	SW
3	N	NW	NW	NW	NW	NW	-	-	SW	SW	SW	SW	SW	SW	W
4	N	NW	NW	NW	NW	NW	W	SW	SW	SW	SW	W	W	NW	NW
5	N	NW	NW	NW	NW	NW	NW	W	W	SW	SW	W	NW	NW	NW
6	N	NW	NW	NW	NW	NW	NW	NW	W	W	NW	NW	NW	NW	NW
7	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	-	-	-	-
8	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	-	-	-
9	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	SW	SW	SW
10	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	W	W	SW
11	N	NW	NW	NW	NW	-	-	NE	N	NW	NW	NW	NW	NW	W
12	N	NW	NW	NW	NW	-	-	NE	NE	N	NW	NW	NW	NW	NW
13	N	NW	NW	NW	NW	-	-	NE	NE	N	NW	NW	NW	NW	NW
14	N	N	NW	NW	NW	-	-	NE	NE	N	NW	NW	NW	NW	NW

سوال ۵

اگر به نتایج قسمت های ۲،۳،۴ دقت کنیم می بینیم که پاداش های قسمت ۴ به سیاست بهتری رسیده اند چون در این حالت سعی می شود کوتاه ترین مسیر انتخاب شود و کمترین برخورد با مانع را داشته باشیم.

جدول ۵-۱: ارزش پیدا شده برای هر خانه بعد از اجرای الگوریتم تکرار سیاست با ضریب تخفیف ۰.۹۹

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	100.00000	97.34042	94.79433	92.28494	89.78252	86.93692	0.00000	0.00000	70.72203	71.29188	71.28318	71.09709	69.77574	68.02101	66.12538
1	97.34043	97.13849	94.71524	92.26869	89.76363	86.76315	0.00000	0.00000	72.67048	73.36710	73.29368	71.97019	70.19845	68.16050	66.23294
2	94.79451	94.71549	94.35259	92.11937	89.72655	86.73837	0.00000	0.00000	75.10693	75.70900	74.30784	72.46587	70.35574	68.21890	66.24951
3	92.28737	92.27221	92.12289	91.63183	89.52540	86.85271	0.00000	0.00000	78.23219	76.69559	74.77281	72.59570	70.40097	68.22824	66.25171
4	89.81483	89.81024	89.77433	89.56205	88.95494	86.78016	83.97957	81.28210	79.13808	77.10914	74.86363	72.62478	70.40386	68.22653	66.24954
5	87.37455	87.37395	87.36177	87.30174	87.02818	86.31926	84.25648	81.83653	79.47117	77.15770	74.87807	72.61124	70.36770	68.17833	66.20875
6	84.96583	84.96622	84.96320	84.94156	84.85513	84.52739	83.74642	81.79327	79.47615	77.16379	74.86856	72.42603	69.92768	67.41003	65.59496
7	82.58850	82.58920	82.58839	82.58198	82.54934	82.43566	82.06076	81.22511	79.36583	77.13948	74.85166	72.22231	0.00000	0.00000	0.00000
8	80.24202	80.24278	80.24241	80.23990	80.22854	80.18321	80.04229	79.62695	78.75088	76.97130	74.80229	72.19064	0.00000	0.00000	0.00000
9	77.92576	77.92648	77.92579	77.91940	77.90448	77.87752	77.81849	77.66052	77.22045	76.31900	74.59390	72.30446	69.63137	67.07288	65.22859
10	75.63923	75.63988	75.63822	75.61993	75.42922	75.22079	75.18110	75.29508	75.29605	74.84651	73.92218	72.23852	70.02865	67.80701	65.81746
11	73.38207	73.38265	73.37969	73.34074	72.89766	0.00000	0.00000	72.58724	72.99311	72.97455	72.50584	71.56490	69.93143	67.82717	65.84946
12	71.15456	71.15542	71.15149	71.09574	70.32867	0.00000	0.00000	69.98303	70.71604	70.75145	70.68742	70.19982	69.24897	67.67594	65.82248
13	68.96504	68.97015	68.96754	68.90647	68.12091	0.00000	0.00000	67.74536	68.51353	68.56719	68.55285	68.44649	67.94147	66.99432	65.62086
14	66.92788	66.98388	66.98263	66.93987	66.31437	0.00000	0.00000	65.92649	66.54525	66.58438	66.58269	66.54915	66.40010	65.86350	64.96029

جدول ۵-۲: سیاست بهینه برای هر خانه بعد از اجرای الگوریتم تکرار سیاست با ضریب تخفیف ۰.۹۹

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	-	W	W	W	W	W	-	-	SE	S	SW	SW	SW	SW	SW
1	N	NW	NW	NW	NW	NW	-	-	SE	S	SW	SW	SW	SW	SW
2	N	NW	NW	NW	NW	NW	-	-	S	SW	SW	SW	SW	SW	SW
3	N	NW	NW	NW	NW	NW	-	-	SW	SW	SW	SW	SW	SW	W
4	N	NW	NW	NW	NW	NW	NW	SW	SW	SW	SW	SW	W	W	NW
5	N	NW	NW	NW	NW	NW	NW	W	W	SW	SW	W	NW	NW	NW
6	N	NW	NW	NW	NW	NW	NW	NW	NW	W	W	NW	NW	NW	NW
7	NE	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	-	-	-
8	NE	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	-	-	-
9	NE	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	W	SW	SW
10	NE	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	W	SW	SW
11	NE	N	NW	NW	NW	-	-	NE	N	NW	NW	NW	NW	NW	W
12	NE	N	NW	NW	NW	-	-	NE	N	NW	NW	NW	NW	NW	NW
13	NE	N	NW	NW	NW	-	-	NE	NE	N	NW	NW	NW	NW	NW
14	NE	N	NW	NW	NW	-	-	NE	NE	N	NW	NW	NW	NW	NW

جدول ۵-۳: ارزش پیدا شده برای هر خانه بعد اجرای الگوریتم تکرار سیاست با ضریب تخفیف ۰.۹

	± 0	± 1	± 2	± 3	± 4	± 5	± 6	± 7	± 8	± 9	± 10	± 11	± 12	± 13	± 14
0	100.00000	85.89694	73.92471	63.46596	54.30828	46.14573	0.00000	0.00000	13.22650	13.80823	13.76546	13.45247	11.28347	8.98159	6.82484
1	85.89694	85.04423	73.63263	63.44327	54.31517	46.09781	0.00000	0.00000	16.29845	16.96323	16.79684	14.29995	11.67453	9.09140	6.91511
2	73.92476	73.63271	72.23503	62.92530	54.22914	46.07635	0.00000	0.00000	20.42241	20.91268	17.87884	14.79977	11.80235	9.13431	6.92627
3	63.46673	63.44441	62.92645	61.23500	53.57095	45.99378	0.00000	0.00000	25.75748	22.00355	18.38153	14.91289	11.83755	9.14079	6.92769
4	54.31983	54.33193	54.24640	53.58464	51.76119	45.31469	38.57543	31.85579	26.75273	22.47226	18.46410	14.93597	11.83961	9.13924	6.92590
5	46.32337	46.34469	46.32439	46.18108	45.43385	43.58169	38.18483	32.44398	27.13719	22.51572	18.47381	14.92401	11.80765	9.09649	6.88991
6	39.33030	39.35068	39.34711	39.30554	39.11416	38.32594	36.51987	31.99056	27.10012	22.51036	18.46371	14.75566	11.38117	8.33388	6.29134
7	33.20696	33.22524	33.22520	33.21538	33.15449	32.92602	32.12924	30.41527	26.59616	22.42168	18.43314	14.56862	0.00000	0.00000	0.00000
8	27.84410	27.86021	27.86095	27.85853	27.84167	27.76385	27.50907	26.72691	25.13112	21.89644	18.30703	14.51709	0.00000	0.00000	0.00000
9	23.14699	23.16164	23.16215	23.15741	23.14457	23.11338	23.02165	22.75764	22.01336	20.55102	17.78890	14.53436	11.03803	7.97970	5.94941
10	19.03331	19.04598	19.04564	19.03116	18.87111	18.69564	18.65335	18.70828	18.59244	17.90479	16.57345	14.19934	11.40530	8.70853	6.51353
11	15.43002	15.44073	15.43924	15.40664	14.98949	0.00000	0.00000	14.71046	15.08709	14.96319	14.32433	13.12129	11.08536	8.70957	6.53665
12	12.27409	12.28368	12.28127	12.23312	11.47164	0.00000	0.00000	11.19028	11.92300	11.94552	11.79555	11.20535	10.12570	8.38711	6.50029
13	9.51721	9.52905	9.52761	9.47446	8.69880	0.00000	0.00000	8.42636	9.18223	9.22765	9.20817	9.04118	8.49885	7.54130	6.17380
14	7.21227	7.26950	7.26896	7.23228	6.62435	0.00000	0.00000	6.36714	6.96803	7.00213	6.99856	6.96100	6.77822	6.26679	5.44468

جدول ۵-۴: سیاست بهینه برای هر خانه بعد از اجرای الگوریتم تکرار سیاست با ضریب تخفیف ۰.۹

	± 0	± 1	± 2	± 3	± 4	± 5	± 6	± 7	± 8	± 9	± 10	± 11	± 12	± 13	± 14
0	-	W	W	W	W	SW	-	-	SE	S	SW	SW	SW	SW	SW
1	N	NW	NW	NW	NW	W	-	-	SE	S	SW	SW	SW	SW	SW
2	N	NW	NW	NW	NW	NW	-	-	S	SW	SW	SW	SW	SW	SW
3	N	NW	NW	NW	NW	NW	-	-	SW	SW	SW	SW	SW	SW	W
4	N	NW	NW	NW	NW	NW	NW	W	SW	SW	SW	SW	W	W	NW
5	NE	N	NW	NW	NW	NW	NW	NW	W	W	W	W	NW	NW	NW
6	NE	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW
7	NE	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	-	-	-
8	NE	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	-	-	-
9	NE	NE	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	W	SW	SW
10	NE	NE	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	W	SW	SW
11	NE	N	NW	NW	NW	-	-	N	NW	NW	NW	NW	NW	NW	W
12	NE	N	NW	NW	NW	-	-	NE	N	NW	NW	NW	NW	NW	NW
13	NE	N	NW	NW	NW	-	-	NE	NE	N	NW	NW	NW	NW	NW
14	NE	N	NW	NW	NW	-	-	NE	NE	N	NW	NW	NW	NW	NW

جدول ۵-۵: ارزش پیدا شده برای هر خانه بعد از اجرای الگوریتم تکرار سیاست با ضریب تخفیف ۰.۶

	± 0	± 1	± 2	± 3	± 4	± 5	± 6	± 7	± 8	± 9	± 10	± 11	± 12	± 13	± 14
0	100.00000	52.53057	27.45651	13.81754	6.40815	1.94141	0.00000	0.00000	-2.98854	-2.48593	-2.46585	-2.46520	-2.46576	-2.47110	-2.48229
1	52.53057	51.37497	27.22050	13.88330	6.44574	1.77347	0.00000	0.00000	-3.18824	-2.46762	-2.43775	-2.43761	-2.44661	-2.46790	-2.48146
2	27.45654	27.22055	25.87454	13.54051	6.42141	1.76912	0.00000	0.00000	-3.11675	-2.40847	-2.38381	-2.39934	-2.44053	-2.46674	-2.48133
3	13.81832	13.88447	13.54169	12.47148	6.13284	1.94016	0.00000	0.00000	-2.56571	-2.28187	-2.30985	-2.38972	-2.43907	-2.46664	-2.48134
4	6.42661	6.47347	6.44906	6.15351	5.40295	1.93454	-0.53448	-1.70862	-2.06114	-2.13957	-2.29555	-2.38838	-2.43981	-2.46771	-2.48211
5	2.38886	2.41722	2.41315	2.38139	2.16230	1.66816	-0.12270	-1.21857	-1.80943	-2.12096	-2.29483	-2.39945	-2.46220	-2.49459	-2.50256
6	0.17892	0.19465	0.19481	0.18846	0.15931	0.01075	-0.29493	-1.21835	-1.79968	-2.11708	-2.30212	-2.63497	-2.93930	-3.21378	-3.00346
7	-1.03194	-1.02320	-1.02286	-1.02351	-1.02977	-1.05280	-1.14829	-1.33169	-1.80918	-2.11851	-2.30977	-2.88086	0.00000	0.00000	0.00000
8	-1.69543	-1.69052	-1.69034	-1.69064	-1.69180	-1.69724	-1.71395	-1.77299	-1.88040	-2.12859	-2.31154	-2.88492	0.00000	0.00000	0.00000
9	-2.05899	-2.05632	-2.05643	-2.06035	-2.06830	-2.07680	-2.08098	-2.08508	-2.11317	-2.17180	-2.31210	-2.64422	-2.94553	-3.22006	-3.00717
10	-2.25828	-2.25689	-2.25741	-2.26888	-2.50548	-2.75087	-2.75553	-2.51724	-2.28658	-2.29293	-2.33008	-2.41065	-2.47385	-2.50130	-2.50656
11	-2.36755	-2.36682	-2.36779	-2.39055	-2.86736	0.00000	0.00000	-2.87704	-2.40613	-2.38476	-2.38923	-2.41044	-2.45222	-2.47489	-2.48627
12	-2.42744	-2.42707	-2.42837	-2.45894	-3.17870	0.00000	0.00000	-3.18713	-2.46812	-2.43823	-2.43741	-2.44073	-2.45257	-2.47410	-2.48552
13	-2.46026	-2.46007	-2.46134	-2.49243	-3.21583	0.00000	0.00000	-3.22084	-2.49780	-2.46698	-2.46577	-2.46600	-2.46826	-2.47481	-2.48564
14	-2.47791	-2.47766	-2.47850	-2.49923	-3.00230	0.00000	0.00000	-3.00527	-2.50238	-2.48169	-2.48086	-2.48087	-2.48115	-2.48264	-2.48619

جدول ۵-۶: سیاست بهینه برای هر خانه بعد از اجرای الگوریتم تکرار سیاست با ضریب تخفیف ۰.۶

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	-	W	W	W	SW	SW	-	-	SE	SE	SE	S	SW	SW	SW
1	N	NW	NW	NW	W	W	-	-	SE	SE	S	SW	SW	SW	SW
2	N	NW	NW	NW	NW	NW	-	-	SE	S	SW	SW	SW	SW	SW
3	N	NW	NW	NW	NW	NW	-	-	SW	SW	SW	SW	SW	W	W
4	NE	N	NW	NW	NW	NW	NW	SW	SW	SW	SW	SW	W	NW	NW
5	NE	N	NW	NW	NW	NW	NW	W	SW	SW	SW	W	NW	NW	NW
6	NE	N	NW	NW	NW	NW	NW	NW	W	W	W	NW	NW	NW	NW
7	NE	NE	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	-	-	-
8	NE	NE	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	-	-	-
9	NE	NE	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	SW	SW	SW
10	NE	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	SW	SW	SW
11	NE	N	NW	NW	NW	-	-	NE	N	NW	NW	NW	W	W	SW
12	NE	N	NW	NW	NW	-	-	NE	NE	N	NW	NW	NW	NW	W
13	NE	N	NW	NW	NW	-	-	NE	NE	NE	N	NW	NW	NW	NW
14	NE	N	NW	NW	NW	-	-	NE	NE	NE	N	NW	NW	NW	NW

با توجه به نتایج بالا می بینیم که هر چه مقدار ضریب تخفیف به ۱ نزدیک تر باشد آینده نگری ربات بیشتر و در نتیجه ربات به پاداش به دست آمده در دراز مدت اهمیت بیشتری می دهد، و در نتیجه جهت استیت های دور از نقطه پایان بهینه تر انتخاب می شوند طوری که کمترین برخورد و کمترین مسیر را طی کند. اما وقتی ضرایب تخفیف کاهش می یابند ربات به پاداش دار مدت کمتر اهمیت می دهد و ممکن است در بعضی از استیت ها اکشن غیر بهینه را انتخاب کند.

سوال ۶

Value Iteration, for estimating $\pi \approx \pi_*$

Algorithm parameter: a small threshold $\theta > 0$ determining accuracy of estimation

Initialize $V(s)$, for all $s \in \mathcal{S}^+$, arbitrarily except that $V(\text{terminal}) = 0$

Loop:

| $\Delta \leftarrow 0$

| Loop for each $s \in \mathcal{S}$:

| $v \leftarrow V(s)$

| $V(s) \leftarrow \max_a \sum_{s',r} p(s',r|s,a)[r + \gamma V(s')]$

| $\Delta \leftarrow \max(\Delta, |v - V(s)|)$

until $\Delta < \theta$

Output a deterministic policy, $\pi \approx \pi_*$, such that

$$\pi(s) = \arg \max_a \sum_{s',r} p(s',r|s,a)[r + \gamma V(s')]$$

شکل ۶-۱ الگوریتم تکرار ارزش برای پیدا کردن سیاست بهینه

جدول ۶-۱: ارزش پیدا شده برای هر خانه بعد از اجرای الگوریتم تکرار ارزش (هزینه متوسط)

	± 0	± 1	± 2	± 3	± 4	± 5	± 6	± 7	± 8	± 9	± 10	± 11	± 12	± 13	± 14
0	1000.00000	999.98664	999.97367	999.96040	999.94301	999.87005	0.00000	0.00000	999.74368	999.80634	999.81079	999.81105	999.81077	999.80912	999.80180
1	999.98664	999.98560	999.97324	999.96015	999.94085	999.84446	0.00000	0.00000	999.73097	999.81655	999.82276	999.82293	999.82189	999.81413	999.80390
2	999.97369	999.97328	999.97135	999.95934	999.94052	999.84261	0.00000	0.00000	999.74694	999.83039	999.83609	999.83572	999.82767	999.81608	999.80436
3	999.96077	999.96069	999.95987	999.95687	999.94087	999.87036	0.00000	0.00000	999.79407	999.84651	999.84974	999.84136	999.82925	999.81631	999.80442
4	999.94787	999.94784	999.94762	999.94621	999.94014	999.89853	999.85355	999.83527	999.84913	999.86414	999.85510	999.84248	999.82916	999.81598	999.80414
5	999.93496	999.93495	999.93487	999.93440	999.93170	999.92368	999.90671	999.89272	999.88069	999.86902	999.85585	999.84042	999.82472	999.81049	999.79962
6	999.92206	999.92206	999.92203	999.92184	999.92106	999.91785	999.90973	999.89636	999.88320	999.86997	999.85451	999.81379	999.76953	999.72602	999.73758
7	999.90916	999.90915	999.90914	999.90907	999.90877	999.90768	999.90412	999.89602	999.88320	999.86995	999.85304	999.78479	0.00000	0.00000	0.00000
8	999.89625	999.89625	999.89622	999.89612	999.89591	999.89539	999.89402	999.89023	999.88224	999.86969	999.85292	999.78413	0.00000	0.00000	0.00000
9	999.88335	999.88334	999.88324	999.88239	999.88071	999.87892	999.87820	999.87801	999.87552	999.86822	999.85385	999.81186	999.76728	999.72297	999.73431
10	999.87044	999.87042	999.87020	999.86779	999.84037	999.81098	999.80918	999.83613	999.86209	999.86139	999.85342	999.83820	999.82169	999.80726	999.79623
11	999.85753	999.85750	999.85712	999.85238	999.79685	0.00000	0.00000	999.79109	999.84444	999.84853	999.84738	999.83940	999.82593	999.81263	999.80061
12	999.84463	999.84459	999.84408	999.83764	999.75287	0.00000	0.00000	999.74492	999.82911	999.83509	999.83523	999.83348	999.82559	999.81280	999.80085
13	999.83177	999.83176	999.83125	999.82452	999.73828	0.00000	0.00000	999.72970	999.81553	999.82191	999.82230	999.82200	999.81972	999.81196	999.80067
14	999.81963	999.81995	999.81961	999.81482	999.75168	0.00000	0.00000	999.74265	999.80547	999.81014	999.81044	999.81037	999.80974	999.80693	999.79956

جدول ۶-۲: سیاست بهینه پیدا شده برای هر خانه بعد از اجرای الگوریتم تکرار ارزش (هزینه متوسط)

	± 0	± 1	± 2	± 3	± 4	± 5	± 6	± 7	± 8	± 9	± 10	± 11	± 12	± 13	± 14
0	-	W	W	W	W	W	-	-	SE	SE	SE	S	SW	SW	SW
1	N	NW	NW	NW	NW	NW	-	-	SE	SE	S	SW	SW	SW	SW
2	N	NW	NW	NW	NW	NW	-	-	SE	SE	S	SW	SW	SW	SW
3	N	NW	NW	NW	NW	W	-	-	SE	S	SW	SW	SW	W	W
4	N	NW	NW	NW	NW	NW	SW	SW	SW	SW	SW	SW	W	NW	NW
5	N	NW	NW	NW	NW	NW	W	SW	SW	SW	SW	W	NW	NW	NW
6	N	NW	NW	NW	NW	NW	NW	W	W	W	W	NW	NW	NW	NW
7	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	-	-	-
8	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	SW	-	-	-
9	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	W	SW	SW	SW
10	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	SW	SW	SW
11	N	NW	NW	NW	NW	-	-	NE	N	NW	NW	NW	W	W	SW
12	N	NW	NW	NW	NW	-	-	NE	NE	N	NW	NW	NW	NW	W
13	N	NW	NW	NW	NW	-	-	NE	NE	NE	N	NW	NW	NW	NW
14	N	NW	NW	NW	NW	-	-	NE	NE	NE	N	NW	NW	NW	NW

جدول ۶-۳: ارزش پیدا شده برای هر خانه بعد از اجرای الگوریتم تکرار ارزش (هزینه کم)

	± 0	± 1	± 2	± 3	± 4	± 5	± 6	± 7	± 8	± 9	± 10	± 11	± 12	± 13	± 14
0	1000.00000	1000.00000	1000.00000	1000.00000	999.99995	999.99934	0.00000	0.00000	999.99913	999.99976	999.99980	999.99980	999.99980	999.99977	999.99970
1	1000.00000	1000.00000	1000.00000	999.99999	999.99993	999.99907	0.00000	0.00000	999.99893	999.99980	999.99986	999.99986	999.99985	999.99982	999.99972
2	1000.00000	1000.00000	1000.00000	999.99999	999.99993	999.99907	0.00000	0.00000	999.99900	999.99985	999.99991	999.99991	999.99990	999.99984	999.99973
3	1000.00000	1000.00000	1000.00000	1000.00000	999.99995	999.99938	0.00000	0.00000	999.99934	999.99990	999.99994	999.99994	999.99991	999.99984	999.99973
4	1000.00000	1000.00000	1000.00000	1000.00000	999.99997	999.99969	999.99939	999.99938	999.99967	999.99994	999.99996	999.99995	999.99991	999.99984	999.99973
5	1000.00000	1000.00000	1000.00000	1000.00000	999.99999	999.99997	999.99996	999.99995	999.99996	999.99998	999.99997	999.99992	999.99986	999.99978	999.99968
6	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	999.99999	999.99999	999.99999	999.99999	999.99995	999.99965	999.99930	999.99906
7	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	999.99999	999.99999	999.99999	999.99994	999.99936	0.00000	0.00000
8	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	999.99999	999.99999	999.99999	999.99999	999.99999	999.99994	999.99936	0.00000	0.00000
9	1000.00000	1000.00000	1000.00000	999.99999	999.99997	999.99996	999.99995	999.99997	999.99998	999.99998	999.99995	999.99965	999.99929	999.99889	999.99899
10	1000.00000	1000.00000	1000.00000	999.99997	999.99969	999.99939	999.99938	999.99968	999.99995	999.99997	999.99996	999.99992	999.99984	999.99974	999.99961
11	1000.00000	999.99999	999.99999	999.99994	999.99938	0.00000	0.00000	999.99935	999.99991	999.99996	999.99996	999.99994	999.99989	999.99979	999.99965
12	999.99999	999.99999	999.99998	999.99991	999.99906	0.00000	0.00000	999.99901	999.99986	999.99992	999.99993	999.99992	999.99988	999.99979	999.99966
13	999.99997	999.99996	999.99996	999.99989	999.99902	0.00000	0.00000	999.99894	999.99980	999.99987	999.99987	999.99987	999.99985	999.99978	999.99964
14	999.99993	999.99993	999.99992	999.99987	999.99923	0.00000	0.00000	999.99910	999.99973	999.99978	999.99978	999.99978	999.99977	999.99972	999.99960

جدول ۴-۶: سیاست بهینه پیدا شده برای هر خانه بعد از اجرای الگوریتم تکرار ارزش (هزینه کم)

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	-	W	W	W	W	W	-	-	SE	SE	SE	S	SW	SW	SW
1	N	NW	NW	NW	NW	NW	-	-	SE	SE	SE	S	SW	SW	SW
2	N	NW	NW	SW	SW	SW	-	-	SE	SE	S	SW	SW	SW	SW
3	N	NW	SW	SW	SW	SW	-	-	SE	SE	S	SW	SW	W	W
4	N	NW	SW	SW	SW	SW	SW	SE	SE	S	SW	SW	W	NW	NW
5	N	NW	W	SW	SW	SW	SW	SW	SW	SW	SW	W	NW	NW	NW
6	N	NW	NW	W	SW	SW	SW	SW	SW	SW	SW	NW	NW	NW	NW
7	N	NW	NW	NW	W	W	W	W	W	W	W	NW	-	-	-
8	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	SW	-	-	-
9	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	SW	SW	SW	SW
10	N	NW	NW	NW	NW	NW	NE	NE	NE	N	NW	W	SW	SW	SW
11	N	NW	NW	NW	NW	-	-	NE	NE	N	NW	NW	W	W	SW
12	N	NW	NW	NW	NW	-	-	NE	NE	NE	N	NW	NW	NW	W
13	N	NW	NW	NW	NW	-	-	NE	NE	NE	N	NW	NW	NW	NW
14	N	NW	NW	NW	NW	-	-	NE	NE	NE	N	NW	NW	NW	NW

جدول ۵-۶: ارزش پیدا شده برای هر خانه بعد از اجرای الگوریتم تکرار ارزش (هزینه زیاد)

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	100.00000	98.66451	97.36922	96.07459	94.74850	92.98418	0.00000	0.00000	83.33771	83.88715	83.90633	83.84965	83.30198	82.37171	81.28450
1	98.66452	98.56059	97.32788	96.06482	94.72870	92.78145	0.00000	0.00000	84.38049	85.10047	85.09835	84.57808	83.65419	82.49813	81.36506
2	97.36944	97.32818	97.13938	95.98616	94.70776	92.75972	0.00000	0.00000	85.78050	86.45599	85.93596	84.98646	83.79755	82.55037	81.37973
3	96.07750	96.06904	95.99037	95.73024	94.60703	92.93512	0.00000	0.00000	87.87042	87.30778	86.32694	85.10753	83.83848	82.55872	81.38170
4	94.78680	94.78386	94.76417	94.65005	94.31361	93.02048	91.25023	89.79024	88.70463	87.66893	86.41666	85.13448	83.84087	82.55675	81.37944
5	93.49635	93.49553	93.48843	93.45496	93.30181	92.89383	91.65871	90.32516	89.00961	87.72356	86.43101	85.11933	83.80289	82.50780	81.33844
6	92.20599	92.20571	92.20354	92.19099	92.14169	91.95312	91.49760	90.32422	89.02611	87.73176	86.42071	84.90942	83.33668	81.73940	80.74481
7	90.91566	90.91556	90.91478	90.91064	90.89150	90.82537	90.60536	90.11192	88.99004	87.72197	86.40628	84.67995	0.00000	0.00000	0.00000
8	89.62533	89.62527	89.62480	89.62275	89.61526	89.58798	89.50458	89.25720	88.73414	87.65361	86.38465	84.65847	0.00000	0.00000	0.00000
9	88.33499	88.33490	88.33403	88.32685	88.31140	88.28871	88.25255	88.16302	87.90303	87.36113	86.29836	84.84067	83.15892	81.53146	80.50470
10	87.04466	87.04449	87.04254	87.02242	86.80527	86.57000	86.53973	86.71081	86.81366	86.55161	85.98678	84.93706	83.59742	82.26708	81.08664
11	85.75434	85.75409	85.75074	85.70959	85.23886	0.00000	0.00000	85.03494	85.47492	85.47980	85.20141	84.61969	83.59354	82.30306	81.12284
12	84.46444	84.46435	84.45993	84.40239	83.63325	0.00000	0.00000	83.39626	84.13406	84.17759	84.14899	83.85374	83.25910	82.26312	81.11211
13	83.18014	83.18268	83.17919	83.11744	82.33167	0.00000	0.00000	82.06298	82.83655	82.89190	82.88521	82.82810	82.51693	81.91784	81.03925
14	81.97145	82.00494	82.00292	81.95942	81.35434	0.00000	0.00000	81.07359	81.67286	81.71235	81.71270	81.69415	81.60824	81.27218	80.69634

جدول ۶-۶: سیاست بهینه پیدا شده برای هر خانه بعد از اجرای الگوریتم تکرار ارزش (هزینه زیاد)

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	-	W	W	W	W	W	-	-	SE	S	SW	SW	SW	SW	SW
1	N	NW	NW	NW	NW	NW	-	-	SE	S	SW	SW	SW	SW	SW
2	N	NW	NW	NW	NW	NW	-	-	S	SW	SW	SW	SW	SW	SW
3	N	NW	NW	NW	NW	NW	-	-	SW	SW	SW	SW	SW	SW	W
4	N	NW	NW	NW	NW	NW	W	SW	SW	SW	SW	SW	W	W	NW
5	N	NW	NW	NW	NW	NW	NW	W	W	SW	SW	W	NW	NW	NW
6	N	NW	NW	NW	NW	NW	NW	NW	NW	W	W	NW	NW	NW	NW
7	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	-	-	-
8	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	-	-	-
9	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	SW	SW	SW
10	N	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	NW	W	W	SW
11	N	NW	NW	NW	NW	-	-	NE	N	NW	NW	NW	NW	NW	W
12	N	NW	NW	NW	NW	-	-	NE	NE	N	NW	NW	NW	NW	NW
13	N	NW	NW	NW	NW	-	-	NE	NE	N	NW	NW	NW	NW	NW
14	N	N	NW	NW	NW	-	-	NE	NE	N	NW	NW	NW	NW	NW

در همه حالت های بالا ضریب تخفیف برابر ۱ در نظر گرفته شده، با مقایسه نتایج حاصل از الگوریتم تکرار ارزش می بینیم که سیاست های پیدا شده کاملاً مشابه سیاست بهینه به دست آمده از الگوریتم تکرار سیاست می باشد اما مقدار ارزش استیت ها تفاوت اندکی دارد.

سوال ۷

با مقایسه قسمت های ۲ و ۳ و ۴ می بینیم که در قسمت ۲ و ۳ به دلیل ضریب تخفیف ۱ و جریمه کم در نظر گرفته شده برای حرکت مقادیر خانه ها بسیار به هم نزدیک هستند، اما در قسمت ۴ به دلیل هزینه ۱۰۰ برابر حرکت نسبت به قسمت ۱ ارزش خانه ها هر چه دورتر از هدف باشد کاهش می یابد.

همچنین اگر به سیاست بهینه به دست آمده هر قسمت توجه کنیم متوجه می شویم در قسمت ۲ سعی می شود تا از مانع ها فاصله بگیریم و مسیر مان کوتاه باشد اما در بعضی از استیت ها به دلیل کوچک بودن هزینه حرکت به کوتاه بودن مسیر زیاد توجه نمی شود و تصمیم انتخاب شده به نزدیک ترین مکان به مقصد منجر نمی شود. در قسمت ۳ به دلیل بی هزینه بودن حرکت فقط سعی می شود عامل از مانع ها دور حتی اگر به قیمت دور شدن از مقصد باشد. ولی در سوال ۴ با معقول در نظر گرفتن پاداش ها و هزینه ها، سیاست به دست آمده مصالحه بین کوتاه بودن فاصله و کمترین برخورد با موانع است و معقولانه تر به نظر می رسد.