بنام خدا

علی قاسم زاده ۴۰۱۱۰۶۳۳۹

۱- الف) نادرست است، مثال نقض: حالتی را بگیرید که یک مسیر با reward کمتر و طول یک داریم ولی مسیر دیگر ~~طول~~ طولش بیشتر از T است و reward بیشتری دارد، در این صورت شکل زیر را در نظر بگیرید:

$$S_{T+6} \xleftarrow{r=2} S_0 \xrightarrow{0} S_1 \xrightarrow{0} \cdots \xrightarrow{0} S_{T+4} \xrightarrow{r=10} S_{T+5}$$

در این صورت داریم که:

$\forall i,\ V_0(S_i)=0$

در value Iteration بعد از یک مرحله $V_1(S_0)=1$ و $V_1(S_{T+4})=10$ ، ... در مرحله i ام داریم که

$$\forall j \le i,\ V(S_{T+5-j})=2$$

در این شرایط $\pi(S_0)$ شرایط گفته شده را دارد چون T مرحله تغییری نکرده است ولی policy بهینه نیست.

ب) غلط است، Q-learning، بدون توجه به اینکه سیاست ما بهینه است یا خیر، حتی اگر sub-optimal هم باشد، converge می‌کند و جواب بهینه برای Q-value ها را به ما می‌دهد یعنی اگه action ها نابهینه باشند در نهایت مقادیر Q-value ها به مقادیر بهینه converge می‌کند.

ج) غلط است، زیرا داشتیم که

$$Q(s,a) \leftarrow (1-\alpha)\, Q(s,a) + \alpha[\text{sample}]$$

$$\xrightarrow{\alpha=1} Q(s,a) = \text{sample}$$

عملاً هیچ استفاده‌ای از policy قبلی نمی‌کنیم و همواره بر اساس sample جدیدی که گرفته‌ایم، آپدیت می‌کنیم ← اصلاً converge اتفاق نمی‌افتد و مقادیر بهینه Q-value ها بدست نمی‌آید.

د) درست است.

پیچیدگی هر iteration در:

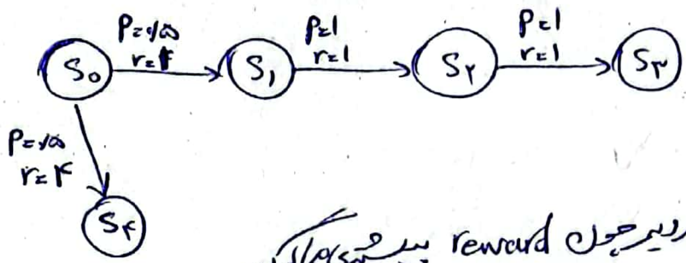
$$\text{Value-iteration} = O(|S|^2|A|)$$

$$\text{Policy-iteration} = O(|S|^2) \xrightarrow[\text{|A| قابل چشم‌پوشی است.}]{|A| \ll |S|} O(|S|^2|A|) = O(|S|^2)$$

⇐ اردر زمانی Value-iteration ، Policy iteration برابر می‌شود.

ه) نادرست است:

در نظر بگیرید که،



منطقی است که از S_0 به S_4 برویم چون reward بیشتری می‌گیریم.

حالا اگر تمام r ها را با 1 جمع کنیم، مسیر $S_0 \xrightarrow{*} S_3$ ، reward ای برابر ۶ خواهد داشت و مسیر $S_0 \rightarrow S_4$ ، reward ای برابر ۵ خواهد داشت ← Policy بهینه عوض می‌شود

۲- الف) در ابتدا داریم که :

$V_0(i,j) = 0 \leftarrow V_{i,j}$

برای خانه‌های صفر

$$V_{k+1}(s) = \max_{a} \sum_{s'} T(s,a,s')\left[R(s,a,s') + \gamma V_k(s')\right]$$

مرحله اول :

$\Rightarrow V_1(1,1) = 0$

$V_1(2,1) = 0$

به هر جایی که برود V_0 آن برابر صفر است

جهت‌ها : d/u/l/r

$$V_1(1,2) = \max(\overset{r}{0.8(-5)}, \overset{l}{0}, \overset{u}{0.1(-5)}, \overset{d}{0.1(-5)}) = 0$$

$$V_1(2,2) = \max(\overset{r}{0.8(5)}, \overset{l}{0}, \overset{u}{0.1 \times 5}, \overset{d}{0.1 \times 5}) = 4$$

۲,۱	۲,۲	۲,۳
0	۴	0
0	0	0
۱,۱	۱,۲	۱,۳

$V_1(1,3) = 0$

$V_1(2,3) = 0$

جابجایی نداریم ← احتمال هم که ضرب شه همیشه بین
به سمت صفر داریم ماکزیمم می‌گیریم

مرحله دوم :

$$V_2(1,1) = \max(\overset{r}{0}, \overset{l}{0}, \overset{u}{0}, \overset{d}{0}) = 0$$

$$V_2(2,1) = \max(\underset{r}{0.8(0.9 \times 4)}, \underset{l}{0.1 \times 0}, \underset{u}{0.1 \times (0.9 \times 4)}, \underset{d}{0.1 \times (0.9 \times 4)}) = 2.88$$

$$V_2(1,2) = \max(\underset{r}{0.8(-5) + 0.1 \times (0.9 \times 4)}, \underset{l}{0.1 \times (0.9 \times 4)}, \underset{u}{0.1(-5) + 0.8(0.9 \times 4)}, \underset{d}{0.1(-5)}) = 2.38$$

$$V_2(2,2) = \max(\underset{r}{0.8(5) + 0.1(0.9 \times 4)}, \underset{l}{0.1(0.9 \times 4)}, \underset{u}{0.8(0.9 \times 4) + 0.1(5)}, 0.1(5)) = 4.36$$

$V_2(1,3) = 0$

$V_2(2,3) = 0$

۲,۱	۲,۲	۲,۳
۲,۸۸	۴,۳۶	0
0	۲,۳۸	0
۱,۱	۱,۲	۱,۳

$$V_۳(۱,۱) = \max\Big(\overset{d}{۰٫۸(۰٫۹ \times ۲٫۳۸) + ۰٫۱ \times (۰٫۹ \times ۲٫۸۸)},\ \overset{r}{۰٫۱ \times (۰٫۹ \times ۲٫۸۸)},\ \overset{L}{۰٫۸(۲٫۸۸ \times ۰٫۹) + ۰٫۱ \times (۰٫۹ \times ۲٫۳۸)},\ \overset{u}{۰٫۱(۰٫۹ \times ۲٫۳۸)}\Big) = \underline{۲٫۲۸۷۸}$$

$$V_۳(۲,۱) = \max\Big(\overset{d}{۰٫۸(۰٫۹ \times ۴٫۳۶) + ۰٫۱(۰٫۹ \times ۲٫۸۸)},\ \overset{r}{۰٫۱(۰٫۹ \times ۲٫۸۸) + ۰٫۸(۰٫۹ \times ۲٫۸۸)},\ \overset{L}{۰٫۸ \times (۰٫۹ \times ۲٫۸۸) + ۰٫۱(۰٫۹ \times ۲٫۸۸) + ۰٫۱ \times (۰٫۹ \times ۴٫۳۶)},\ \overset{u}{۰٫۱ \times (۰٫۹ \times ۲٫۸۸) + ۰٫۱ \times (۰٫۹ \times ۴٫۳۶)}\Big) = \underline{۳٫۳۹۸۴}$$

$$V_۳(۱,۲) = \max\Big(\overset{r}{۰٫۸(-۵) + ۰٫۱(۲٫۳۸ \times ۰٫۹) + ۰٫۱ \times (۰٫۹ \times ۴٫۳۶)},\ \overset{L}{۰٫۱(۰٫۹ \times ۲٫۳۸) + ۰٫۱ \times (۰٫۹ \times ۴٫۳۶)},\ \underset{u}{۰٫۸ \times (۰٫۹ \times ۴٫۳۶) + ۰٫۱ \times (-۵)},\ \underset{d}{۰٫۸(۰٫۹ \times ۲٫۳۸) + ۰٫۱(-۵)}\Big) = ۲٫۶۳۹$$

$$V_۳(۲,۲) = \max\Big(\overset{r}{۰٫۸(۵) + ۰٫۱(۰٫۹ \times ۲٫۳۸) + ۰٫۱(۰٫۹ \times ۴٫۳۶)},\ \overset{L}{۰٫۸(۰٫۹ \times ۲٫۸۸) + ۰٫۱ \times (۰٫۹ \times ۴٫۳۶) + ۰٫۱(۰٫۹ \times ۲٫۳۸)},\ \underset{u}{۰٫۸(۰٫۹ \times ۴٫۳۶) + ۰٫۱(۵) + ۰٫۱(۰٫۹ \times ۲٫۸۸)},\ \overset{d}{۰٫۸(۲٫۳۸ \times ۰٫۹) + ۰٫۱(۵) + ۰٫۱(۰٫۹ \times ۲٫۳۸)}\Big) = ۴٫۶۰۶۶$$

$$V_۳(۱,۳) = ۰$$

$$V_۳(۲,۳) = ۰$$

ب) باید از الگوریتم‌های RL استفاده کنیم.

model-based learning، در این روش عامل با مشاهده‌ی محیط مقدارهای $R(s,a,s')$ و $T(s,a,s')$ را به‌طور تخمینی بدست آورد. سپس بعد از بدست آوردن مقادیر، MDP ای که ساختیم را حل می‌کنیم.

passive RL، در این روش یک Policy ثابت تصادفی انتخاب می‌کنیم و هربار Q-value های دیده‌شده را آپدیت می‌کنیم و این فرایند را برای چند episode ادامه می‌دهیم و در نهایت مانند یک درخت expectimax، Policy بهینه را انتخاب می‌کنیم.

Active RL، علاوه بر مقادیری که در passive حساب می‌شد، Policy را هم در طول کار یاد می‌گیریم. در هر مرحله سیاست بهینه‌ای را که تا الان بدست آورده ایم را استفاده می‌کنیم. ε-greedy هم طوری است که با احتمال ε از Policy فعلی پیروی نمی‌کند و رندوم عمل می‌کند.

ج) $(1,1) - (1,2) - (1,3)$

$$sample = R(s, \pi(s), s') + \gamma V^{\pi}(s')$$

$$V^{\pi}(s) \leftarrow (1-\alpha)V^{\pi}(s) + \alpha[sample]$$

$$sample1 = 0 + 0.9(0) = 0 \longrightarrow V^{\pi}(1,1) = 0$$

$$sample2 = -5 + 0.9(0) = -5 \longrightarrow V^{\pi}(1,2) = 0 + 0.1(-5) = -0.5$$

در مسیر بعدی:

$(1,1) - (1,2) - (2,2) - (2,3)$

$$sample3 = 0 + 0.9(-0.5) = -0.45 \longrightarrow V^{\pi}(1,1) = 0 + 0.1(-0.45) = -0.045$$

$$sample4 = 0 + 0.9(0) = 0 \longrightarrow V^{\pi}(1,2) = 0.9(-0.5) + 0.1(0) = -0.45$$

$$sample5 = 5 + 0.9(0) = 5 \longrightarrow V^{\pi}(2,2) = 0.9(0) + 0.1(5) = 0.5$$

بقیه هم صفر هستند.

۳-۲) با استفاده از Q learning از خانه (0,0) شروع می‌کنیم و سپس هر بار که یک sample پیدا می‌کنیم با استفاده از معادلات بلمن بروزرسانی می‌کنیم!

$$V^*(s) = \max_a Q^*(s,a)$$

$$Q^*(s,a) = \sum_{s'} T(s,a,s')\left[R(s,a,s') + \gamma V^*(s')\right]$$

$$V^*(s) = \max_a \sum_{s'} T(s,a,s')\left[R(s,a,s') , \gamma V^*(s')\right]$$

چون T و R را نداریم، پس یادگیرنده سعی می‌کند تا sample هایی از محیط دریافت کند و مقادیر Q-value را یاد بگیرد.

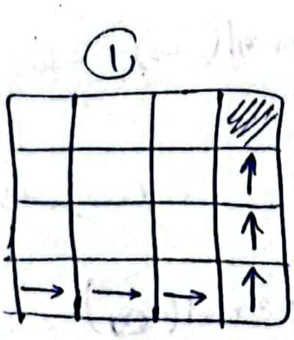
$$\text{sample} = R(s,a,s') + \gamma \max_{a'}(s',a')$$

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left(\text{sample} - Q(s,a)\right)$$

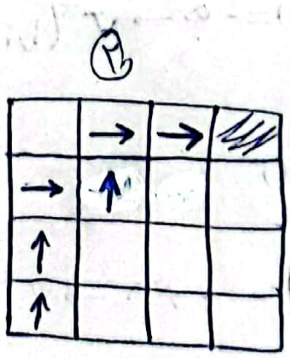
ابتدا همه‌ی Q ها را صفر می‌گذاریم سپس هر خانه معادل یک state است و هر جهت معادل action است. سپس در هر اپیزود مسیر را از شروع تا پایان می‌رویم و در این زمان نیز عملیات learning را انجام می‌دهیم.

ب) دو تا مسیر می‌گیریم!

1



2



$$Q((0,0),\underset{\downarrow\ \text{right}}{r}) = \overset{(1-\alpha)Q(s,a)}{0} + \overset{\alpha}{0.1}(\underset{R(s,a,s')}{-1} + \underset{\gamma \max_{a'} Q(s',a')}{0}) = -0.1$$

$$Q((0,1),r) = 0 + 0.1(-1+0) = -0.1$$

بقیه صفر

$$Q((0,2),r) = 0 + 0.1(-1+0) = -0.1$$

$$Q((0,3),u) = 0 + 0.1(-1+0) = -0.1$$

$$Q((1,3),u) = 0 + 0.1(-1+0) = -0.1$$

$$Q((2,3),u) = 0 + 0.1(10) = 1$$

در مسیر دوم،

$$Q((0,0),u) = 0 + 0.1(-1+0) = -0.1$$

$$Q((1,0),u) = 0 + 0.1(-1+0) = -0.1$$

$$Q((2,0),r) = 0 + 0.1(-1+0) = -0.1$$

$$Q((2,1),u) = 0 + 0.1(-1+0) = -0.1$$

$$Q((3,1),r) = 0 + 0.1(-1+0) = -0.1$$

$$Q((3,2),r) = 0 + 0.1(10) = 1$$

ج) نقش اکتشاف، در این الگوریتم exploration این نقش دارد که ما محیط را به طور کامل مشاهده کنیم. ممکن است پاداش های احتمالی زیادی وجود داشته باشند که آنها را نمی‌بینیم و با exploration آنها را می‌بینیم ← در پیدا کردن بهینه‌ترین جواب کاربرد داره

نقش استفاده از اطلاعات، در این الگوریتم، این نقش exploration را دارد
ما آنقدر سیاستی که تا الان بدست آمده است استفاده کنیم تا سیاستمان ناگهانی
تغییر نکند و در نهایت converge کند، اگه استفاده از اطلاعات کم باشد
میل شدت به اکتشاف‌مان وابسته خواهد بود و حرکات تصادفی زیاد می‌شود و
مدت زمان زیادی طول می‌کشد تا الگوریتم converge کند. ← استفاده از اطلاعات
گذشته در converge کردن الگوریتم مؤثر است.

برای ε - greedy، ابتدا مقدار ε را زیاد می‌گذاریم چون در ابتدا لازم است که
در محیط جستجو کنیم و اطلاعات بدست آوریم، ولی رفته رفته ε را باید کم کنیم
حرکت رندوم بزنیم
تا تأثیر exploitation بیشتر شود و agent ما سیاست بهینه خود را
پیدا کند. اینگونه شناختی که از محیط بدست آورده ایم را استفاده می‌کنیم
همچنین الگوریتم زودتر همگرا شود برای این هر یک مقدار برای کمترین مقدار
ε و یک مقدار برای نرخ کم شدن ε در نظر می‌گیریم.

۴- الف)

$$V^*(s) = \max_a Q^*(s,a) = Q^*(s,\pi^*(s))$$

$$\Rightarrow V^*(s) - Q^*(s,\pi(s)) = Q^*(s,\pi^*(s)) - Q^*(s,\pi(s))$$

نرم بینهایت معادل ماکزیمم است ← به ازای هر s و a داریم که:

$$\|\tilde{Q}(s,a) - Q^*(s,a)\|_\infty \le \varepsilon \Rightarrow$$

$$\Rightarrow \begin{cases} \tilde{Q}(s,a) - Q^*(s,a) \le \varepsilon \\ Q^*(s,a) - \tilde{Q}(s,a) \le \varepsilon \end{cases}$$

$$\xrightarrow{a \text{ را } \pi(s) \text{ قرار دهیم}} V^*(s) - Q^*(s,\pi(s)) = Q^*(s,\pi^*(s)) - Q^*(s,\pi(s))$$

$$\le Q^*(s,\pi^*(s)) + \varepsilon - \tilde{Q}(s,\pi(s))$$

حالا چون $\tilde{Q}(s,\pi(s)) \ge \tilde{Q}(s,\pi^*(s))$ داریم که:

$$Q^*(s,\pi^*(s)) + \varepsilon - \tilde{Q}(s,\pi(s)) \le \underbrace{Q^*(s,\pi^*(s)) + \varepsilon - \tilde{Q}(s,\pi^*(s))}_{\text{که حداکثر مساوی } \varepsilon} \le 2\varepsilon$$

$$\Rightarrow Q^*(s,\pi^*(s)) + \varepsilon - \tilde{Q}(s,\pi^*(s)) \le 2\varepsilon$$

$$\Rightarrow V^*(s) - Q^*(s,\pi(s)) \le Q^*(s,\pi^*(s)) + \varepsilon - \tilde{Q}(s,\pi^*(s)) \le 2\varepsilon$$

$$\Rightarrow V^* - Q^*(s,\pi(s)) \le 2\varepsilon$$

ب) از نامساوی بخش قبل داریم که

$$\xrightarrow{B} V^*(s) - V_\pi(s) \le Q^*(s,\pi(s)) + 2\varepsilon - V_\pi(s) = 2\varepsilon + Q^*(s,\pi(s)) - \tilde{Q}(s,\pi(s))$$

$$Q^*(s,\pi(s)) - \tilde{Q}(s,\pi(s)) =$$

$$\sum_{s'} T(s,\pi(s),s')\left[\cancel{R(s,\pi(s),s')} + \gamma V^*(s')\right] - \sum_{s'} T(s,\pi(s),s')\left[\cancel{R(s,\pi(s),s')} + \gamma V_\pi(s')\right]$$

$$= \gamma \sum_{s'} T(s,\pi(s),s')\left[V^*(s') - V_\pi(s')\right] \le \gamma \max_{s'}\left(V^*(s') - V_\pi(s')\right) \times \underbrace{\sum_{s'} T(s,\pi(s),s')}_{1}$$

$$= \gamma \max_{s'}\left(V^*(s') - V_\pi(s')\right)$$

در نامساوی اول این قسمت می‌دانیم که به ازای هر s برقرار است ⇐ در امی که ~~ا~~ آن را ماکزیمم می‌کند،

$$\max_s\left(V^*(s) - V_\pi(s)\right) \le 2\varepsilon + \gamma \max_{s'}\left(V^*(s') - V_\pi(s')\right)$$

$$\Rightarrow \max_s\left(V^*(s) - V_\pi(s)\right) \le \frac{2\varepsilon}{1-\gamma}$$

$$\Rightarrow \forall s,\ V^*(s) - V_\pi(s) \le \max_s\left(V^*(s) - V_\pi(s)\right) \le \frac{2\varepsilon}{1-\gamma} \Rightarrow \boxed{V_\pi(s) \ge V^*(s) - \frac{2\varepsilon}{1-\gamma}}$$

ج) برای بدست آوردن $V^*(s_1)$ و $V^*(s_2)$ ، از s_1 باید به s_2 برویم تا پاداش مثبتی بگیریم.
برای s_2 هم فقط می‌توانیم در خودش بمانیم حالا طبق معادله داریم:

$$V^*(s) = \max_a Q^*(s,a)$$

T ها مشخص و deterministic هستند و معادل احتمال وقوع قطعی یک action هستند
یا عدم وقوع قطعی

a ← ماندن
b ← رفتن

$$V^*(s_2) = \max\Big(\overset{1}{T(s_2,a,s_2)}\big(R(s_2,a,s_2) + \gamma V^*(s_2)\big) + \underset{0}{T(s_2,a,s_1)}\big(R(s_2,a,s_1) + \gamma V^*(s_1)\big),$$
$$\overset{1}{T(s_2,b,s_2)}\big(R(s_2,b,s_2) + \gamma V^*(s_2)\big) + \underset{0}{T(s_2,b,s_1)}\big(R(s_2,b,s_1) + \gamma V^*(s_1)\big)\Big)$$

$$= 24 + \gamma V^*(s_2) \Rightarrow \boxed{V^*(s_2) = \frac{24}{1-\gamma}}$$

$$V^*(s_1) = \max\Big(\overset{0}{T(s_1,a,s_2)}\big(R(s_1,a,s_2) + \gamma V^*(s_2)\big) + \underset{1}{T(s_1,a,s_1)}\big(R(s_1,s,s_1) + \gamma V^*(s_1)\big),$$
$$\overset{1}{T(s_1,b,s_2)}\big(R(s_1,b,s_2) + \gamma V^*(s_2)\big) + \underset{0}{T(s_1,b,s_1)}\big(R(s_1,b,s_1) + \gamma V^*(s_1)\big)\Big)$$

حالا واضح است که $V^*(s_1) \neq R(s_1,a,s_1) + \gamma V^*(s_1)$ چرا که در آن صورت $\leftarrow V^*(s_1) = 0$

$$V^*(s_1) = R(s_1,b,s_2) + \gamma V^*(s_2) \rightarrow V^*(s_1) = 24 + \frac{24\gamma}{1-\gamma} = \frac{24}{1-\gamma}$$

حال مقدار $Q^*(s_1,a)$ را حساب می‌کنیم

$$Q^*(s_1,a) = T(s_1,a,s_2)\big(R(s_1,a,s_2)+\gamma V^*(s_2)\big) + T(s_1,a,s_1)\big(R(s_1,a,s_1)+\gamma V^*(s_1)\big) = \frac{2\gamma\epsilon}{1-\gamma}$$

$$Q^*(s_1,b) = T(s_1,b,s_2)\big(R(s_1,b,s_2)+\gamma V^*(s_2)\big) + T(s_1,b,s_1)\big(R(s_1,b,s_1)+\gamma V^*(s_1)\big) = 2\epsilon+\gamma V^*(s_1) = \frac{2\epsilon}{1-\gamma}$$

د) ادعا می‌کنیم که اگر $\tilde{Q}(s_1,a)=\tilde{Q}(s_1,b)=\frac{(1+\gamma)\epsilon}{1-\gamma}$ آنگاه شرایط مسئله برقرار است.

$$\left|\tilde{Q}(s_1,a)-Q^*(s_1,a)\right| = \left|\frac{(1-\gamma)\epsilon}{1-\gamma}\right| = \epsilon \le \epsilon$$

$$\left|\tilde{Q}(s_1,b)-Q^*(s_1,b)\right| = \left|\frac{(1+\gamma)\epsilon}{1-\gamma}-\frac{2\epsilon}{1-\gamma}\right| = \left|\frac{(1-\gamma)\epsilon}{1-\gamma}\right| = \epsilon \le \epsilon$$

که نتیجه می‌دهد که خطای $\tilde{Q}$ حداکثر ϵ است. حال چون

$$\tilde{Q}(s_1,a)=\tilde{Q}(s_1,b) \longrightarrow \pi(s_1)=\arg\max_a \tilde{Q}(s_1,a)$$

می‌توان دید که در هر مرحله action ، a را انتخاب کرد

$$V_\pi(s_1)=R(s_1,a,s_1)+\gamma V_\pi(s_1) \Rightarrow V_\pi(s_1)=0$$

چون عامل همواره عمل ماندن را انجام می‌دهد

$$V_\pi(s_1)-V^*(s_1)=0-\frac{2\epsilon}{1-\gamma}=-\frac{2\epsilon}{1-\gamma}$$

۵- الف) در مرحله اول پاداش صفر می‌گیریم $r_0 = 0$ و در بقیه مراحل پاداش یک می‌گیریم ← $r_t = 1$ ⇐

$$\sum_{t=0}^{\infty} \gamma^t r_t = \sum_{t=1}^{\infty} \gamma^t = \frac{1}{1-\gamma} - 1 = \frac{\gamma}{1-\gamma}$$

ب) در مرحله اول پاداش $\frac{\gamma^2}{1-\gamma}$ می‌گیریم و در بقیه مراحل 0 ←

$$r_0 = \frac{\gamma^2}{1-\gamma}, \; r_t = 0 \quad \forall i \geq 1 \Rightarrow \sum_{t=0}^{\infty} \gamma^t r_t = \frac{\gamma^2}{1-\gamma}$$

حالا چون $\gamma < 1$ ← $\frac{\gamma}{1-\gamma} > \frac{\gamma^2}{1-\gamma}$ ← عمل بهینه a_1 است ← با a_1 به s_1 می‌رویم.

ج)

$$V_{n+1}(s) = \max_a \sum_{s'} T(s, a, s') \left[R(s, a, s') + \gamma V_n(s') \right]$$

برای حالت s_0، تنها مقدارهای $T(s_0, a_1, s_1) = T(s_0, a_2, s_2) = 1$ ناصفر اند.

$$\Rightarrow V_{n+1}(s_0) = \max\left(R(s_0, a_1, s_1) + \gamma V_n(s_1), R(s_0, a_2, s_2) + \gamma V_n(s_2) \right)$$

$$= \max\left(\sum_{t=1}^{n} \gamma^t, \frac{\gamma^2}{1-\gamma} \right)$$

حالا حالتی بهینه است که $\sum_{t=1}^{n} \gamma^t > \frac{\gamma^2}{1-\gamma}$ برای اینکه action a_1 را انتخاب کند و کار بهینه را کرده باشد.

$$V_{n^*}(s_0) = \sum_{t=1}^{n^*} \gamma^t \Rightarrow \sum_{t=1}^{n^*} \gamma^t \ge \frac{\gamma^2}{1-\gamma} \Rightarrow \gamma\left(\frac{1-\gamma^{n^*}}{1-\gamma}\right) \ge \frac{\gamma^2}{1-\gamma} \longrightarrow$$

$$1-\gamma^{n^*} \ge \gamma \longrightarrow 1-\gamma \ge \gamma^{n^*} \xrightarrow{\log} n^* \log\gamma \le \log(1-\gamma)$$

کوچکتر از

$$\xrightarrow{\log\gamma < 0} \boxed{n^* \ge \frac{\log(1-\gamma)}{\log(\gamma)}}$$

حکم سوال ثابت شد.