AI - HW5

Subject:

Date

متین باقری ۴۰۲۱۰۵۷۲۷

۱) آ) درست، در این صورت کران بالایی برای utility function وجود خواهد داشت و می توان دید که value ها و policy نهایی converge می کنند.

ب) نادرست، هر دو این روش ها سیاست بهینه را پیدا می کنند.

ج) درست، پس از گذشت زمانی کافی، فارغ از action های انتخاب شده، Q-value های صحیح به دست می آیند و می توان از روی آنها سیاست بهینه را متوجه شد.

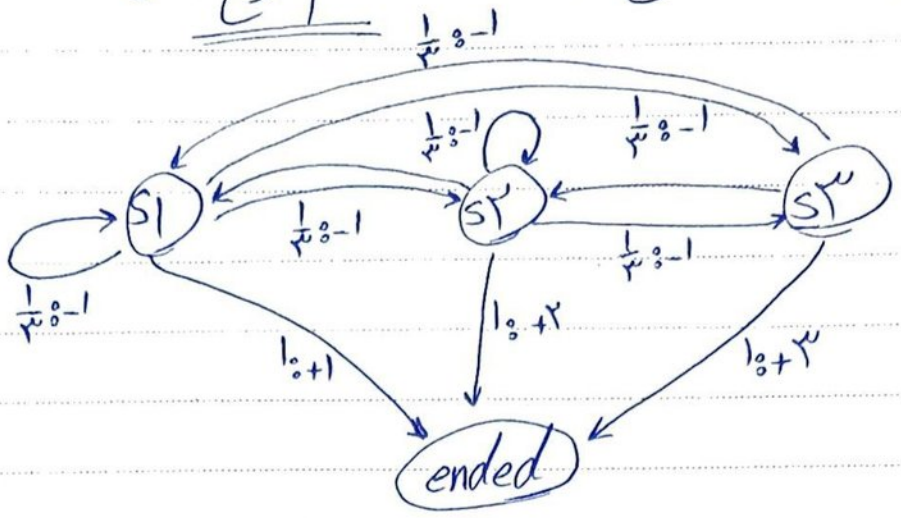
د) نادرست، صرفا به زمان بسیار بیشتری برای همگرا شدن Q ها نیاز دارد.

هـ) درست، چون نمودار P_1 بر اساس P_2 کاملاً صعودی است.

و) انتقال سریع تر پاداش در محاسبات و یادگیری سریع تر، وابستگی کمتر به مقادیر موجود Q

۲) الف) در هر دو قسمت ب و ج سیاست بهینه، اتمام سریع بازی است.

یادآوری: احتمال →



ب)

	S_1	S_2	S_3	ended
v_0	۰	۰	۰	۰
v_1	+1	۲	۳	۰
v_2	۱	۲	۳	۰
v_3	۱	۲	۳	۰

ج)

	S_1	S_2	S_3	ended
$v_0^{\pi_0}$	۰	۰	۰	۰
$v_1^{\pi_0}$	۱	۲	-۱	۰
$v_2^{\pi_0}$	۱	۲	-۰٫۴	۰
$v_3^{\pi_0}$	۱	۲	-۰٫۲۲	۰
$v_4^{\pi_0}$	۱	۲	-۰٫۱۴۳	۰

$v_{k+1}^{\pi_0}(S_3) = ۰٫۳ \times v_k(S_3) - ۰٫۱$

$v_4 = -۰٫۱۹۹$, $v_5 = -۰٫۱۴۹۸$

$v_4 = -۰٫۱۴۴۹۴$

$\pi_1(S_1) = \pi_1(S_2) =$ اتمام , $\pi_1(S_3) =$ اتمام $\begin{cases} \text{ماندن} \rightarrow -۱ + \frac{۰٫۹}{۳} \times ۲٫۱۵۷ = -۰٫۱۴۲۹ \\ \text{اتمام} \rightarrow ۳ \end{cases}$

PAPCO

v^{π} → S_1 ↓ ۱ , S_2 ↓ ۲ , S_3 ↓ ۳

Subject:
Date

سیاست بهینه در هر حالت quit است.

۳) آ) در افق زمانی ۱: ~~$\pi^*(s_1)=\pi^*(s_2)=climb$, $\pi^*(s_3,s_4,s_5,s_6)=quit$~~

ب)

	s_1	s_2	s_3	s_4	s_5
Q_0	۰	۰	۰	۰	۰
Q_1	۰/۱	۰/۲	۰/۴	۰/۵	۰/۷
Q_2	۳٫۵/۱	۴٫۵/۲	۳٫۵/۴	۴٫۵/۵	۱/۷
Q_3	۴٫۷۵/۱	۴٫۵/۲	۴٫۷۵/۴	۵٫۷۵/۵	۳٫۵/۷
Q_4	۵٫۱۲۵/۱	۵٫۲۵/۲	۵٫۱۲۵/۴	۵٫۷۵/۵	۴٫۷۵/۷
Q_5	۵٫۵/۱	۴٫۴۳۷۵/۲	۵٫۵/۴	۴٫۱۲۵/۵	۵٫۱۲۵/۷
Q_∞	۷/۱	۷/۲	۷/۴	۷/۵	۷/۷

ج) $\pi^*(s_1,s_2,s_3,s_4)=climb \quad , \pi^*(s_5)=quit$

د) $\frac{\gamma}{2}\times\left(max(2, Q(s_2,climb)+max(7, Q(s_5,climb)) > 5 \;?\; climb : quit\right)$

Subject :
Date

ج، آ) $v_{stay}(1) = \sum_{i=0}^{\infty} \gamma^i = 1 + \frac{1}{2} + \frac{1}{4} + \dots = 2$

ب) $v^*(1) = v_{stay}(1) = 2$

ج) از آنجاکه $\gamma = 1$ ، باید به نهری برویم که بیشترین r_i را دارد و سپس در آن شنا کنیم و تجارت کنیم. از آنجاکه p_i ها مثبت اند، انجام این کار ناممکن نیست.

د) i را کوچکترین مقداری در نظر بگیرید که به ازای آن $r_i > 0$ ، آنگاه بیشترین مقدار K که در آن ممکن است $v_K(1) = 0$ برابر است با $i-1$. اگر فرض شود به ازای هر i $r_i > 0$ آنگاه بیشترین مقدار ممکن برای K صفر است.

هـ) پس از یک مرحله iter. $v_K(i) = r_i > 0$ بنابراین بیشترین مقدار ممکن برای K صفر است.

و)

sarsa	$Q(1,S)$	$Q(1,E)$	$Q(2,W)$	$Q(2,S)$
initial	0	0	0	0
1, S, 4, 1	2	0	0	0
1, E, 0, 2	2	0	0	0
2, S, 6, 2	2	0	0	3
2, W, 0, 1	2	0	1	3
1, S, 4, 1	4	0	1	3

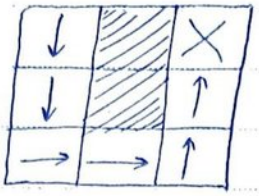
$\gamma = 1$, $\alpha = LR = 0.5$

ادامهٔ ۶) د) درصورتی که تنها مجاز به استفاده از رابطهٔ خطی باشیم خیر، در غیر این صورت بله

۶) آ)

$\frac{1}{4}$	$\frac{1}{2}$	$+1$
$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$
$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$

ب)



ج)

۲	۳	۴
۱	۲	۳
۰	۱	۲

$w_1 = w_2 = 1$

$v_{(s)} = f_1(s) + f_2(s) = x + y$

د) بله، می‌توان. اگر فاصلهٔ منهتن خانهٔ s با خانهٔ $+1$ را با $d_{(s)}$ نشان دهیم:

$v^*_{(s)} = 2^{-d(s)}$, $d_{(s)} = 2 - x + 2 - y = 4 - (f_1(s) + f_2(s))$

هـ) i. در قسمت ج نشان دادیم که برای A قابل حل است اما برای B و C به دلیل ساختار شبکه آنها قابل حل نیست.

ii. با این وزن‌ها برای A قابل حل است:

مشابه قسمت ج می‌شود: $w_{i,j} = i + j$

برای شبکهٔ B نیز قابل حل است: $w_{0,2} = 0$, $w_{0,1} = 1$, $w_{0,0} = 2$, $w_{1,0} = 3$

$w_{2,0} = 4$, $w_{2,1} = 5$, $w_{2,2} = 6$

برای شبکهٔ C نیز قابل حل است: $w_{0,0} = w_{0,2} = w_{2,0} = w_{2,2} = 0$, $w_{1,1} = 2$

$w_{0,1} = w_{1,0} = w_{2,1} = w_{1,2} = 1$

Subject :
Date

۴) هـ) iii. برای C قابل حل است: $w_1 = w_2 = -1,\ w_3 = 2$

برای A, B قابل حل نیست.

۵) آ) $W^*(y) = \sum_{s'} P(s'|y)\, U^*(s')$

ب) $Q^*(s,a) = R(s,a) + \sum_{s'} T(s,a,s')\gamma U^*(s')$

به کمک فرمت الف ← $U^*(s) = \max_a \left(R(s,a) + \gamma W^*(f(s,a)) \right.$

ج) $W^*(y) = \sum_{s'} p(s'|y) \left(\max_a \left(R(s',a) + \gamma W^*(f(s',a)) \right) \right)$

د) جای خالی اول:

$$\pi^{(i+1)}(s) = \arg\max_a \left(R(s+a) + \gamma W^{\pi^i}(f(s,a)) \right)$$

هـ) $W(y_t) \leftarrow (1-\alpha) W(y_t) + \alpha \left(R(s_{t+1}, a_{t+1}) + \gamma W(y_{t+1}) \right)$

د) جای خالی دوم: اگر مقدار $W^{\pi^i}(s)$ همگرا شده بود و اختلاف آن با $W^{\pi^{i-1}}(s)$ ناچیز و قابل چشم پوشی بود.