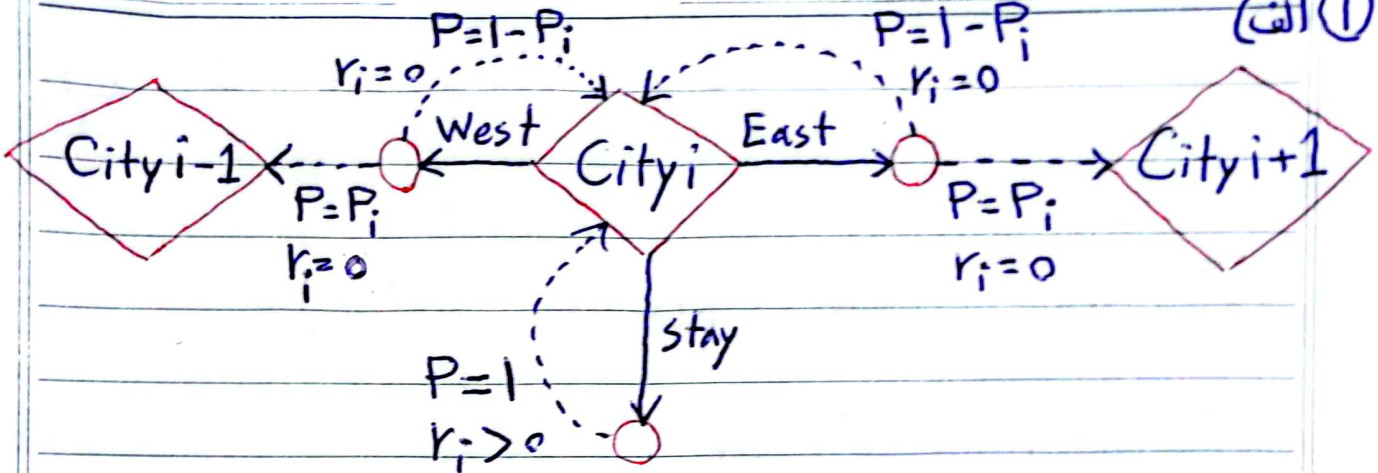
۹۹۵۲۱۲۷۱ تمرین دوازدهم هوش فرزان رحمانی

Subject:( ) Year: Month: Day:

① الف)



ب) $\gamma = \frac{1}{2}$, $P_i = 1$, $r_i = 1$

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

$T(s,\pi(s),s')$ → فقط یک انتخاب با احتمال ۱ داریم (P_i)

$R(s,\pi(s),s') = r_i$

$$V^{stay}(1) = 1\left(1 + \frac{1}{2}V^{stay}(1)\right) \rightarrow V^{stay}(1) = 1 + \frac{1}{2}V^{stay}(1)$$

$$\frac{1}{2}V^{stay}(1) = 1 \rightarrow \boxed{V^{stay}(1) = 2}$$

با همگرایی هم قابل حل است.

$$V_0^{stay}(1) = 0 \quad V_1^{stay}(1) = 1\left(1 + \frac{1}{2}\times 0\right) = 1 \quad V_2^{stay}(1) = 1\left(1 + \frac{1}{2}\right) = \frac{3}{2}$$

$$V_3^{stay}(1) = 1\left(1 + \frac{3}{4}\right) = \frac{7}{4} \rightarrow V^{stay}(1) = 1 + \frac{1}{2} + \frac{1}{4} + \ldots = \frac{1}{1-\frac{1}{2}} = 2$$

TANIN

Subject:( ) Year: Month: Day:

ب) $\gamma = \frac{1}{2}$, $P_i = 1$, $r_i = 1$

برای شرق و غرب

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

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

همچنین چون برای همهٔ آن‌ها پارامترها یکسان هستند لذا استیت‌ها متقارن هستند یعنی وضعیت مشابهی دارند.

$$V^*(1) = V^*(2) = V^*(3) = \dots = V^*(N)$$

I) $Q^*(1,stay) = 1 \times (1 + 0.5\,V^*(1)) = 1 + \frac{1}{2}V^*(1)$

II) $Q^*(1,east) = 1 \times (0 + 0.5\,V^*(2)) = \frac{1}{2}V^*(2) = \frac{1}{2}V^*(1)$

III) $Q^*(1,west) = 1 \times (0 + 0.5\,V^*(N)) = \frac{1}{2}V^*(N) = \frac{1}{2}V^*(1)$

$V^*(1) = \max_a Q^*(1,a) = Q^*(1,stay)$ → چون با 1 جمع می‌شود از بقیه بیشتر است.

باتوجه به مقادیر بالا

I) $Q^*(1,stay) = 1 + \frac{1}{2}Q^*(1,stay) \rightarrow Q^*(1,stay) = 2 = V^*(1)$

II) $Q^*(1,east) = \frac{1}{2}V^*(1) = \frac{2}{2} = 1 \rightarrow Q^*(1,east) = 1$

($V^*(1) = Q^*(1,stay)$)

III) $Q^*(1,west) = \frac{1}{2}V^*(1) = \frac{2}{2} = 1 \rightarrow Q^*(1,west) = 1$

TANIN

$$V^{stay}(1) = V^*(1) = \max_a Q^*(1,a) = Q^*(1,stay) = 2$$

Subject:( ) Year: Month: Day:

$\gamma = 1$, $P_i > 0$, $r_i > 0$ (ت

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

$$\pi^*(i) = \arg\max \{ \underbrace{1(r_i + V^*(i))}_{\text{stay}} , \underbrace{P_i V^*(i+1) + (1-P_i) V^*(i)}_{\text{east}} ,$$

$$\underbrace{P_i V^*(i-1) + (1-P_i) V^*(i)}_{\text{west}} \}$$

فرمول سیاست بهینه

$$\pi^*(i) = \arg\max \{ \underbrace{V^*(i) + r_i}_{\text{stay}} , \underbrace{V^*(i) + P_i (V^*(i+1) - V^*(i))}_{\text{east}} ,$$

$$\underbrace{V^*(i) + P_i (V^*(i-1) - V^*(i))}_{\text{west}} \}$$

وابسته به مقدار r_i ، $P_i(V^*(i+1) - V^*(i))$ ، $P_i(V^*(i-1) - V^*(i))$

r_i پاداش شهر فعلی

$V^*(i+1) - V^*(i)$ اختلاف ارزش شهر جدید با شهر فعلی

می باشد.

بهینه

بسته به مقادیر بالا سیاست می تواند متفاوت باشد مثلاً اگر r_i ها

شهرها برابر باشد آنگاه $\forall i \ \pi^*(i) = stay$ می شود چرا که

ارزش شهرها برابر است و با ماندن در یک شهر سود مشابهی می گیریم.

$V^*(i+1) - V^*(i) = 0$

اما اگر r_i ها تفاوت داشته باشند سیاست بهینه این می شود که

TANIN

ابتدا به شهری برویم که بیشترین r_i را دارد سپس آنها بمانیم

و تجارت کنیم.

چون r_i و P_i دقیق داده نشده اینطور نمی توان سیاست ثابتی

را بیان کرد و برای تعیین سیاست باید از فرمول به دست آمده و

مقدارهای r_i و P_i هر شهر استفاده کنیم. مقدارهای $V^*(s)$

نیز وابسته به r_i و P_i هستند.

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

$$\pi^*(i) = \arg\max \left\{ \underbrace{V^*(i) + r_i}_{\text{stay}}, \underbrace{P_i\left(V^*(i+1) - V^*(i)\right) + V^*(i)}_{\text{east}}, \underbrace{V^*(i) + P_i\left(V^*(i-1) - V^*(i)\right)}_{\text{west}} \right\}$$

Subject:( ) (ب Year: Month: Day: $\gamma = 0.5$ $\alpha = 0.5$ - learning rate

(s, a, r, s')	Q(1, stay)	Q(1, east)	Q(2, west)	Q(2, stay)
initial	0	0	0	0
(1, stay, 4, 1)	2	0	0	0
(1, east, 0, 2)	2	0	0	0
(2, stay, 6, 2)	2	0	0	3
(2, west, 0, 1)	2	0	0.5	3
(1, stay, 4, 1)	3.5	0	0.5	3

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

از فرمول روبرو استفاده می‌کنیم.

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

$(1, stay, 4, 1)$: $sample = 4 + \frac{1}{2}0 = 4$

$Q(2, stay) = \frac{1}{2} \times 0 + \frac{1}{2} \times 4 = 2$

$(1, east, 0, 2)$: $sample = 0 + \frac{1}{2}\max_{a'} Q(2, a') = 0 + 0 = 0$

$Q(1, east) = \frac{1}{2} \times 0 + \frac{1}{2} \times 0 = 0$

$(2, stay, 6, 2)$: $sample = 6 + \frac{1}{2}0 = 6$

$Q(2, stay) = \frac{1}{2} \times 0 + \frac{1}{2} \times 6 = 3$

$(2, west, 0, 1)$: $sample = 0 + \max_{a'}(1, a') = 0 + \frac{1}{2}2 = 1$

$Q(2, west) = \frac{1}{2} \times 0 + \frac{1}{2} \times 1 = 0.5$

$(1, stay, 4, 1)$: $sample = 4 + \frac{1}{2} \times 2 = 5$

$Q(1, stay) = \frac{1}{2} \times 2 + \frac{1}{2} \times 5 = 3.5$

TANIN