Subject: Year: Month: Day:

تمرین سوم یادگیری تقویتی — سیده ستاره خسروی ۹۹۴۱۱۴۲۵

سؤال ۱:

الف- ۱ صفحه و یک گوی، گوی را در وسط نگه داریم.

State: محل قرارگیری توپ، سرعت توپ

Action: زاویه شیب در جهات x و y

Reward: پاداش منفی متناسب با فاصله از مرکز ← β در فاصله شدا

پاداش منفی بزرگ برای وقتی که توپ بیفتد

ب- برای مهام:

State: ارزش خطی مهام، تحت خرید و فروش

Action: خرید و فروش و هیچ کاری نکردن

Reward: سود و زیان، پاداش منفی برای هزینه‌ای اضافی و تراکنش‌ها

سؤال 2:

1) X, left, 0, X, left, 0, X, left, 0, ...

2) X, right, +1, X, right, -1, Y, right, 4, End

3) $G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots +$

$G_0 = 1 + 0.5 \times (-1) + (0.5)^2 (4) = 1 - 0.5 + 1 = 2 - 0.5 = 1.5$

$0.25 \times 4 = 1$

King

Subject: Year: Month: Day:

سید مسعود حیدری ۹۹۴۱۱۴۵۵

$$V_{\pi_1} = \sum_a \pi_1(a|s) \sum_{s',r} p(s',r|s,a)[r + \gamma V_{\pi_1}(s')] \quad (4$$

$$V_{\pi_1}(s) = 4 + \underbrace{V_{\pi_1}(T)}_{0} = 4$$

$$V_{\pi_2}(X) = \frac{3}{4}(1 + \gamma V_{\pi_2}(X)) + \frac{1}{4}(1 + \gamma V_{\pi_2}(Y)) \quad (5$$

$$V_{\pi_2}(Y) = 4 + \underbrace{\gamma V_{\pi_2}(T)}_{0} = 4$$

$$V_{\pi_2}(X) = \frac{3}{4}(1 + 0.5\, V_{\pi_2}(X)) + \frac{1}{4}(1 + \underbrace{0.5 \times 4}_{2})$$

$$= \frac{3}{4}(1 + 0.5\, V_{\pi_2}(x) + 1)$$

$$= \frac{3}{4}(2 + 0.5\, V_{\pi_2}(x))$$

$$4 V_{\pi_2}(x) = 6 + 1.5 V_{\pi_2}(x)$$

$$2.5 V_{\pi_2}(x) = 6$$

$$V_{\pi_2}(x) = \frac{6 \times 2}{5} = 2.4$$

سئوال 3:

$\gamma = 0.9 \qquad R_1 = 5 \qquad G_t = \sum 10 \gamma^k$

$= \frac{10}{1-\gamma}, \frac{10}{1-0.9} = 100$

$G_1 = R_2 + \gamma G_2$

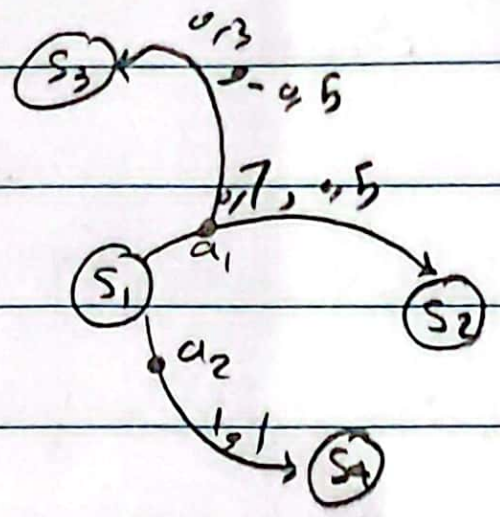
$= 10 + 0.9 \times 100 = 100$

$G_0 = R_1 + \gamma G_1 = 5 + 0.9 \times 100 = 95$ ]

Subject: Year: Month: Day:

سیده کتایون مزدی 99211357

سئوال 4:




$\pi(a_1|s_1) = 0.4$

$\pi(a_2|s_1) = 0.6$

$v_\pi(s_1) = 0.4[0.3[-0.5 + \gamma v(s_3)] + 0.7[0.5 + \gamma v(s_2)]] + 0.6[1 + \gamma v(s_4)]$

$= 0.4[0.15 + 0.35] + 0.6[1]$

$= 0.08 + 0.6 = 0.68$

$q_\pi(s_1, a_2) = 1 \times [1 + \gamma v(s_4)] = 1$

$q_\pi(s_1, a_1) = 0.3 \times [-0.5 + \gamma v(s_3)] + 0.7[0.5 + \gamma v(s_2)]$

$= -0.15 + 0.35 = 0.2$

$v_\pi(s_1) = 0.4\, q_\pi(s_1, a_1) + 0.6\, q_\pi(s_1, a_2)$

سئوال 5:

$$q_*(s,a) = \sum_{s',r} p(s',r|s,a)[r + \gamma \max_{a'} q_*(s',a')]$$

$\pi^*(a|s) = \arg\max_a q^*(s,a)$ : برای

$\pi'^*(a|s) = \arg\max_a q'^*(s,a)$ : برای

$q'^*(s,a) = q^*(s,a) + f(s)$ داریم:

$\pi'^*(a|s) = \arg\max_a [q^*(s,a) + f(s)]$ پس برای $\pi'$ داریم:

الگوریتم تابعی از state است و نسبت به a ثابت است پس برای a ها برابر است. پس می‌توانیم max را ببریم بیرون.

با توجه به اینکه $f(s)$ نسبت به a ثابت است پس آن را بازی می‌کند نداریم:

$\arg\max_a [q^*(s,a) + f(s)] = \arg\max_a q^*(s,a) = \pi^*(a|s)$

King

Subject: Year: Month: Day:

۹۹۲۱۱۲۷۸

سید ستاره هزوی

نتیجه‌ی مشارکه: $\pi^*(a)$ و $\pi^*(b)$

$\gamma = 0,9$

سوال 6.

$$V_\pi(x) = \frac{1}{3}\left(\frac{4}{5}(1,6 + 0,9\,V_\pi(m)) + \frac{1}{5}(-5 + 0,9\times 0)\right)$$

$$+\frac{1}{3}(-2,7 + 0,9\times 3) + \frac{1}{3}\left(\frac{3}{4}(0,2 + 0,9\times 2) + \frac{1}{4}(-1,6 + 0,9\times 5)\right)$$

$$= \frac{4}{15}(1,6 + 0,9\,V_\pi(m)) + \frac{1}{3}(-1) + 0 + \frac{2}{4} + \frac{5,6}{12}$$

$$= 0,43 + 0,24\,V_\pi(m) - \frac{1}{3} + \frac{1}{2} + 0,47$$

$$V_\pi(x) - 0,24\,V_\pi(m) = 0,43 - \frac{1}{3} + \frac{1}{2} + 0,47$$

$$V_\pi(m) \simeq 1,4$$

$$V_*(m) = \max\left[\sum_{s',r} p(s',r|s,a)\,[r + \gamma V(m')]\right]$$

اگر فرض کنیم R بیشترین است

$$V_R(m) = \frac{4}{5}(1,6 + 0,9\,V_*(m)) + \frac{1}{5}(-5 + 0,9\times 0)$$

$$= 1,28 + 0,72\,V_*(m) - 1$$

$$0,28\,V_*(m) = 0,28 \rightarrow V^*(m) = 1$$

$$V_*(m) = \max\left[\frac{4}{5}(1,6 + 0,9\times 1) + \frac{1}{5}(-5 + 0,9\times 0),\right.$$

$$[-2,7 + 0,9\times 3],$$

$$\left.\frac{3}{4}(0,2 + 0,9\times 2) + \frac{1}{4}(-1,6 + 0,9\times 5)\right]$$

$$= \max[1, 0, 2,9] = 2,9$$

Subject: Year: Month: Day:

سید شهاب حزبی 99411455

سؤال ۲:

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

$$= \frac{1}{4}(0.9 \times 2.3) + \frac{1}{4}(0.9 \times 0.7) + \frac{1}{4}(0.9 \times 0.4) + \frac{1}{4}[0.9 \times (-0.9)]$$

$$= \frac{1}{4} \times 0.9 \times 3 = 0.675 \approx 0.7$$