

99100VV

امیدوار صفاتی قادری

$$\text{States} = \{ R_1, R_2, R_3, B, D \}$$

Riddle 1 Beam Diffusion

$$\text{Actions} = \{ n, g \} \quad T(D, *, *) = 0 \quad T(B, *, *) = 0$$

سوال ۱

امیدواری پرداز با مرکت‌های بینه با شروع از s ، انجام نمکت، طبق مسأله $V_i(s)$ را تعریف کنیم

مول $Q_i(s, a) \nvdash s \in \text{States} \cdot V_0(s) = 0$ را تعریف کنیم

$$Q_i(s, a) = \sum_{s' \in S} T(s, a, s') [R(s, a, s') + \gamma V(s')]$$

\hookrightarrow discounting factor

$$V_i(s) = \max_{(s, a) \in A} Q_i(s, a)$$

حل هر مرحله Q_i را محاسبه کنیم و از روی آن V_i را محاسبه کنیم ۳ مقدار مرده قائم عالی است.

$i=1$

$$Q(R_1, g) = 0.3 \times [1000 + 0] + 0.7 \times [-1000 + 0] = -400$$

$$Q(R_1, n) = 0.5 \times [150 + 0] + 0.5 \times [-1000 + 0] = -425$$

$$Q(R_2, g) = 0.4 \times [1000 + 0] + 0.6 \times [-1000 + 0] = -200$$

$$Q(R_2, n) = 0.5 \times [400 + 0] + 0.5 \times [-1000 + 0] = -300$$

$$Q(R_3, g) = 0.9 \times [1000 + 0] + 0.1 \times [-1000 + 0] = 800$$

$$V(R_1) = \max(-400, -425) = -400$$

$$V(R_2) = \max(-200, -300) = -200$$

$$V(R_3) = \max(800) = 800$$

$i=2$

$$Q(R_1, g) = 0.3 \times [1000 + 0] + 0.7 \times [-1000 + 0] = -400$$

$$Q(R_1, n) = 0.5 \times [150 + (-200) \times 0.9] + 0.5 \times [-1000 + 0] = -515$$

$$Q(R_2, g) = 0.4 \times [1000 + 0] + 0.6 \times [-1000 + 0] = -200$$

$$Q(R_1, n) = 0.5(400 + (800 \times 0.9)) + 0.5 \times [-1000 + 0] = 60$$

$$Q(R_3, g) = 0.9 \times [1000 + 0] + 0.1 \times [-1000 + 0] = 800$$

$$V(R_1) = \max(-400, -515) = -400$$

$$V(R_2) = \max(-200, 60) = 60$$

$$V(R_3) = \max(800) = 800$$

$i=3$

$$Q(R_1, g) = 0.3 \times [1000 + 0] + 0.7 \times [-1000 + 0] = -400$$

$$Q(R_1, n) = 0.5(150 + (60 \times 0.9)) + 0.5[-1000 + 0] = -398$$

$$Q(R_2, g) = 0.4 \times [1000 + 0] + 0.6 \times [-1000 + 0] = -200$$

$$Q(R_2, n) = 0.5[400 + (800 \times 0.9)] + 0.5[-1000 + 0] = 60$$

$$Q(R_3, g) = 0.9 \times [1000 + 0] + 0.1 \times [-1000 + 0] = 800$$

$$V(R_1) = \max(-400, -398) = -398$$

$$V(R_2) = \max(-200, 60) = 60$$

$$V(R_3) = \max(800) = 800$$

حال برای بست آوردن یک بیان کننده Policy بهتر که این $\pi(s)$ باشد

یعنی policy به صورت تردیدی باشد

$$\pi(R_1) = n \quad \pi(R_2) = n \quad \pi(R_3) = g$$

عنی یک Policy برای نهمن این است که ۲باره میتوانی بروز بینی گیرد و هر حل کند و سپس گوس پوش

* در این MDP مخصوص $V(D), V(B)$ برای π است و من معنی transition درین State ها ندارم

و قدرتار نمایند $Q(s, a)$ در نظر گیری

بخش ۱) له بامض میت بودن پاداں هالاراونه منه بروارمی باسته فرض کنیه کو تاهه توین صید لزد
ه کے استیت یا یانی بصیرت $L_{5,000,000,000}$ ات باسته ایت کنیه که ارلن یی

$$\arg \min_{j=0} V_j(S_i) = L - L = 0 \text{ است هی لیکن بردار}$$

پایه: برای L چون استیت یا یانی است از همان ادله برقرار است دلای هیچ ساخت بردار

است چون اگر برای ۲ بردار را به L_0, L_1, L_2 کنیه میر کو تاهه قر آز کو تاهه توین صیر خوش شده
با تهای ناهنتر

است ره ماقن می دیم

فرضی: برای $k+1$ نیز برقرار است هی نتلاک کن ناصندرنکه مادر سریان باشی $V_{k+1} = \max_{\alpha \in A} V_k(\alpha)$

فرضی خلف

حتم: برای $k+1$ نیز برقرار است

گام: فرض کنیه در V_{k+1} کنیه راسی مانند α باش که میز حد اشاره کنیه تا راسی یا یانی نه ارد دلیل صد ارمیت

دارد با خوبی ه این که V_{k+1} از V_k ایمهت حی مت لعن کنیه راسی مانند "د ذرهای های کبو

که در V_k ناصندرنکه طبق $\max_{\alpha \in A} \sum_{t=1}^{T_k} r_t(y_t, a_t, \alpha_t) V_k(a_t)$ می باشد مطلب که فرضی

سرکل، $V_0(t)$ و α_t terminals α_t های دلار که علیمی فرضی صیر کنکر کنیم) هی α کنیه ای دلار که علیمی فرضی صیر کنکر کنیم

این که ترمیل دارد می α که $\max_{\alpha \in A} V_{k+1}$ باست ترمیل دارد ه با مرعن خلف ایمان درسته است هی

رئیس که صیر حد کنکر مدل V_{k+1} ندارد در V_{k+1} صفر هسته خاله راسی دلخواه را گیرید مانند γ

که

که سیر با محدود هذکر ۱۱۱ را رس ترسیل دارد این سیر را در قله هذکر می بینیم مثلاً (نحوه داده شده)

اگر سیر (۱)۷، (۲)۷، (۳)۷ را گیری طبق روش در V_K صون (۱)۷ که سیر هذکر را کار داشته باشد.

$$T(v, a, v_{k+1}) = \max_{a \in A} \sum_{v' \in V} T(v, a, v') [\gamma V_k(v')]$$

و مطابق با V_{k+1} است

به ازای هر آنکه action و ناقص بخوبی V_{k+1} این بعبارت بیشتر از پیش از آن است. (نمای راهی که مردم رانند)

۵) (در اینجا از MDP حذف کنیم) می بینیم ابتدا های که محدود حداقل ۱۱۱ دارند همان V_{k+1} می باشند فراهم داشت

$$V_{k+1}(s) = \max_{a \in A} \sum_{s' \in S} T(s, a, s') \gamma V_k(s')$$

و کم مثله این تئوری می باشد $V_0(s) = 1$

$$V_{k+1}(s) = \max_{a \in A} \sum_{s' \in S} T(s, a, s') \gamma V_k(s')$$

باشه که فقط طبق فرمول $V_{k+1}(s) = \max_{a \in A} \sum_{s' \in S} |V_{k+1}(s') - V_k(s)| / \theta_k$ مقدار θ_k بصریت محاسبه می شود

$$\text{می بینیم } V_{k+2}(s) \leq V_{k+1}(s) + \gamma \theta_k$$

$$\theta_{k+1} \leq \gamma \theta_k \quad \text{با وجود قدرت } V_{k+1}(s) - \gamma \theta_k \leq V_{k+2}(s) \leq V_{k+1}(s) + \gamma \theta_k$$

$$\text{پس در هر مرحله از } 44 \text{ مرحله الگوریتم } \Delta \text{ می باشد که } \Delta = 0.001 \text{ می باشد (متداوله ای اینهاست)}$$

و اینکه محدود بینه لزوماً واسه با انتزاع محدود نیست و بینه تهار علیه از اینها محدود کننده دو قسم می باشد

که از اینها $a \in A$ ها بحد کم برداشت باشند اما آنکه محدود کردن خواهد بود و همچنان که محدود کردن خواهد شد پس از هر مرحله

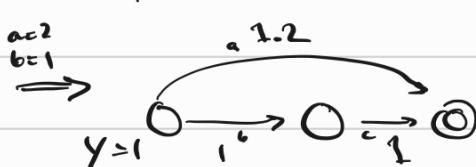
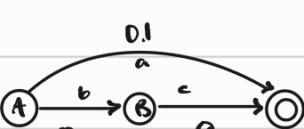
جیوه ۳) مدل زیر را در نظر بگیرید که حالتی باشد که هم می توانیم به مدل داریم میگذرد این را در نظر بگیرید

که با این مدل transition را بازشوند و اگر $\gamma = 1$ باشد آنکه داریم π بهینه موافق بینی اینهاست. به مدل مربوط MDP را در نظر بگیرید

$$\pi(a|s) = a$$

$$\pi(a|s) = (b, c)$$

action
نامهای کننده



در این سطح policy ممکن نیست مانند: همان دلیل می‌گذارد که نهاده این کر ارزش مسیرها باشد و با بایان مکارهای

بیشتر را بالاتری خواهد

خیلی (ع) می‌توان هر episode را محاصل تک در sequential در طبقه ترکیت یعنی آنکه در یک راه را می‌گذرد

$$G_t = G'_{t+1} + \gamma G'_{T_1} + \gamma^2 G'_{T_2} + \dots$$

که هر کدام از G' ها بعده را که هر کدام از G' ها بعده را

حال اگر $\gamma = 1$ مولید می‌شوند، G' داینامیکی \rightarrow sequential

بچتی (6) نے د حالتی مرتبہ کی کوئی طبیعی زیری صرف عادت داری

$$V^{\pi'}(s) = \max_{a \in \text{Actions}} \sum_{s' \in S} T(s, a, s') [R(s, a, s') + \gamma V^{\pi}(s')]$$

$$V^{\pi}(s) = \sum_{s' \in S} T(s, \pi(s), s') [R(s, \pi(s), s') + \gamma V^{\pi}(s')] , \quad \pi(s) \in A$$

$$\text{سے } V^{\pi}(s) \leq V^{\pi'}(s) \quad (\text{بجا ملے گرتنے نہ کر کے اسی لئے})$$

حال اگر برابریتہ ہے تو π معاملہ بھی ملن پڑے کہ اسی لئے

$$V^{\pi}(s) = \max_{s' \in S} \sum_{a \in A} T(s, a, s') [R(s, a, s') + \gamma V^{\pi}(s')]$$

وہی دلیل اسی معاملہ تکہ π نہ دارد سی π کے policy ہے اسی لئے

در آنچه اخیر می‌کنیم اساین برابر با صدیق بیان این است و مان طبق آن در آن Q -table

Episode 1

وزن هرا ایستاده کنی

$$Q((0,0), R) = Q((0,0), R) + 0.5 \left[0 + \max_{a \in \text{Actions}} Q((1,0), a) - Q((0,0), R) \right] = 0$$

جین w_i را تغییر نمایی گذاشت $= 10$ difference

$$Q((1,0), U) = Q((1,0), U) + 0.5 \left[0 + \max_{a \in \text{Actions}} Q((1,1), a) - Q((1,0), U) \right] = 0$$

حالت است ((1,0)) w_i را تغییر نمایی کنی

$$Q((1,1), U) = Q((1,1), U) + 0.5 \left[-100 + \max_{a \in \text{Actions}} Q((1,2), a) - Q((1,1), U) \right] = -50$$

مان طبق w_i را ایستاده کنی $diff = -100$

$$w_u = w_u + 0.5 \times (-100) \times 1 = -50$$

$$w_j = w_j + 0.5 \times (-100) \times 1 = -50$$

$$w_a = w_a + 0.5 \times (-100) \times 2 = -100$$

Episode 2

$$Q((0,0), R) = Q((0,0), R) + 0.5 \left(\underbrace{-100}_{0} + \max_{a \in \text{Actions}} Q((1,0), a) - Q((0,0), R) \right) = -125$$

$$Q((0,0), R) = -100$$

$$\max_{a \in \text{Actions}} Q((1,0), a) = -150, a=R$$

وزن هارا صورت صندوقی کنی $diff = -50$

$$w_u = w_u + 0.5 \times -50 \times 0 = -50$$

$$w_y = w_y + 0.5 \times -50 \times 0 = -50$$

$$w_a = w_a + 0.5 \times -50 \times 1 = -125$$

$$Q((1,0), R) = \underbrace{Q((1,0), R)}_{-175} + 0.5 [0 + \max_{a \in \text{Actions}} Q((2,0), a) - Q((1,0), R)]$$

$$= -200$$

$$Q((1,0), R) = -175$$

$$\max_{a \in \text{Actions}} Q((2,0), a) = -225$$

میرت زیر زن هارا آیینه کنی diff = 50

$$w_u = w_u + 0.5 \times -50 \times 1 = -75$$

$$w_y = w_y + 0.5 \times -50 \times 0 = -50$$

$$w_a = w_a + 0.5 \times -50 \times 1 = -150$$

$$Q((2,0), D) = \underbrace{Q((2,0), D)}_{-750} + 0.5 [100 + \max_{a \in \text{Actions}} Q((2,-1), a) - Q((2,0), D)] = -450$$

$$Q((2,0), D) = -750$$

$$\max_{a \in \text{Actions}} Q((2,-1), a) = -250$$

مکان هارا میرت زیر آبیت کنی diff = 600

$$w_u = w_u + 0.5 \times 600 \times 2 = 525$$

$$w_y = w_y + 0.5 \times 600 \times 0 = -50$$

$$w_a = w_a + 0.5 \times 600 \times 4 = 1050$$

Episode 3

3100

$$Q((0,0), R) = \underbrace{Q((0,0), R)}_{1050} + 0.5 \left[0 + \max_{a \in \text{Actions}} (Q((0,1), a) - Q((0,0), R)) \right] = 2600$$

$$Q((0,0), R) = 1050$$

$$\max_{a \in \text{Actions}} Q((0,1), a) = 4150, a=1$$

درین جانب صورت زیست آئینه رفتونه diff = 3100

$$w_n = w_n + 0.5 \times 3100 \times 0 = 525$$

$$wy = wy + 0.5 \times 3100 \times 6 = -90$$

$$wa = wa + 0.5 \times 3100 \times 1 = 9600$$

2175

$$Q((0,1), L) = Q((0,1), L) + 0.5 \left[100 + \max_{a \in \text{Actions}} (Q((-1,1), a) - Q((0,1), L)) \right] = 88375$$

$$Q((0,1), L) = 7750$$

$$\max_{a \in \text{Actions}} Q((-1,1), a) = 2825, a=0$$

درین جانب صورت زیست آئینه diff = 2175

$$w_n = w_n + 0.5 \times 2175 \times 0 = 525$$

$$wy = wy + 0.5 \times 2175 \times 1 = 1037.5$$

$$wa = wa + 0.5 \times 2175 \times 3 = 7412.5$$

من میگویم که درین نظرها من را بخوبی در اینجا معرفی کنید. $w = (w_n, wy, wa)$

این پیش بازه حرمتی است هیچ حکم کرده و رازیادی کند با ارزش ترین اتفاقات را زیاد کند

فکری نداشته باشید فربودون نیز هابهم را بسنجید.

جایگاه (اتاق)
↑
متغیر

$$S = \{(l, c) \mid l \in \{\text{bedroom, bathroom, living room}\}, c \in N\}$$

ویژه

5

در این داده‌ات می‌توان حرکت ایمنی خواست
که باید عملی کر Value iteration

+ (حواله نشانه اگر $c=1$ می‌تواند بهم بپری جهی می‌تواند بین فضاهای اخیر)

$$\begin{cases} c=1 \quad V((l, 1)) = \gamma V((\text{bathroom}, 5)) \\ c \geq 2 \quad V((l, c)) = \max_{a \in \text{Actions}} \sum_{l'} T(l, a, l') [R(l, a, l') + \gamma V(a', c-1)] \end{cases}$$

(b)

$$\begin{cases} c=2 \quad Q((l, 2), a) = \sum_{l'} T(l, a, l') [R(l, a, l') + \gamma \max_{a' \in \text{Actions}} Q((\text{bathroom}, 5), a')] \\ c \geq 2 \quad Q((l, c), a) = \sum_{l'} T(l, a, l') [R(l, a, l') + \gamma \max_{a' \in \text{Actions}} Q((l, c-1), a')] \end{cases}$$

$$\gamma = 0.8$$

$$Q((\text{bathroom}, c), \text{tolivingroom}) = 2.2$$

$$Q((\text{bathroom}, c), \text{tobedroom}) = 3.1$$

$$Q((\text{bedroom}, c), \text{tolivingroom}) = 2.5$$

$$Q((\text{bedroom}, c), \text{tobath room}) = 2$$

$$Q((\text{diningroom}, c), \text{tobath room}) = 1$$

$$Q((\text{livingroom}, c), \text{tobedroom}) = 2$$

$$Q((*, 2), a) = 0$$

$$Q((\text{bathroom}, c), \text{tolivingroom}) = 3.8$$

$$Q((\text{bathroom}, c), \text{tobedroom}) \approx 4.28$$

$$Q((\text{bedroom}, c), \text{tolivingroom}) \approx 4.188$$

$$Q((\text{bedroom}, c), \text{tobath room}) \approx 4.48$$

$$Q((\text{diningroom}, c), \text{tobath room}) \approx 4.48$$

$$Q((\text{diningroom}, c), \text{tobedroom}) \approx 3.24$$

$$A((*, 2), a) = 2.48 + R_a$$

$$Y=0.2$$

$$Q((\text{bathroom}, c), \text{tolivingroom}) = 2.2$$

$$Q((\text{bathroom}, c), \text{tobedroom}) \approx 3.1$$

$$Q((\text{bedroom}, c), \text{tolivingroom}) \approx 2.5$$

$$Q((\text{bedroom}, c), \text{tobath room}) \approx 2$$

$$Q((\text{diningroom}, c), \text{tobath room}) \approx 1$$

$$Q((\text{diningroom}, c), \text{tobedroom}) \approx 2$$

$$A((*, 2), a) = 0$$

next iteration \rightarrow

$$Q((\text{bathroom}, c), \text{tolivingroom}) = 2.6$$

$$Q((\text{bathroom}, c), \text{tobedroom}) \approx 3.6$$

$$Q((\text{bedroom}, c), \text{tolivingroom}) \approx 3$$

$$Q((\text{bedroom}, c), \text{tobath room}) \approx 2.6$$

$$Q((\text{diningroom}, c), \text{tobath room}) \approx 1.5$$

$$Q((\text{diningroom}, c), \text{tobedroom}) \approx 2.6$$

$$A((*, 2), a) = 0.6 + R_a$$

۲) همان طور که راجع بود در مالیک *discourteous* گزارش نباید مانند *disrespectful* باشد.

هر دسته از این اصطلاحات مراحل صدور نظر کم رکته در سینما می باشند اما اینها معمولی هستند.

را از لومت دهم

سؤال ۵ طبق فرمول کو Q کی خوبی

$$r_t^{\pi_i}(s) = \overset{a_i}{Q}_t^{\pi_i}(s, \pi_i(s))$$

* اسپریافی حافظت میں پذیری درد $E[x+y] = E[x] + E[y]$

$$\begin{aligned}
 & Q_t^{\pi_1} - Q_t^{\pi_2} = E \left[\left(\sum_{\substack{u_t \sim \pi_1 \\ u_{t+1} \sim \pi_2}} T(u_t, \pi_1(u_t), u_{t+1}) [R(u_t, \pi_1(u_t), x_{t+1}) + \gamma Q_{t+1}^{\pi_1}(u_{t+1}, \right. \right. \\
 & \left. \left. \pi_1(u_{t+1})) \right] - \left(\sum_{\substack{u_t \sim \pi_1 \\ u_{t+1} \sim \pi_2}} T(u_t, \pi_1(u_t), x_{t+1}) [R(u_t, \pi_2(u_t), u_{t+1}) + \gamma Q_{t+1}^{\pi_2}(u_{t+1}, \right. \right. \\
 & \left. \left. \pi_2(u_{t+1})) \right] \right] = E \left[\sum_{\substack{u_t \sim \pi_1 \\ u_{t+1} \sim \pi_2}} T(u_t, \pi_1(u_t), u_{t+1}) [Q_{t+1}^{\pi_1}(u_{t+1}, \pi_1(u_{t+1})) \right. \\
 & \left. - Q_{t+1}^{\pi_2}(u_{t+1}, \pi_2(u_{t+1})) \right] \\
 & \stackrel{\sum T=1}{=} E \left[Q_{t+1}^{\pi_1}(u_{t+1}, \pi_1(u_{t+1})) - Q_{t+1}^{\pi_2}(u_{t+1}, \pi_2(u_{t+1})) \right] \\
 \Leftrightarrow & E \left[\underset{u_t \sim \pi_2}{V_t^{\pi_1}(u_t) - V_t^{\pi_2}(u_t)} \right] = E \left[\underset{\substack{\longleftarrow \\ Q_t^{\pi_1}(u_t, \pi_1(u_t)) - Q_t^{\pi_2}(u_t, \pi_2(u_t))}}{Q_{t+1}^{\pi_1}(u_{t+1}, \pi_1(u_{t+1})) - Q_{t+1}^{\pi_2}(u_{t+1}, \pi_2(u_{t+1}))} \right] \\
 & - \left(Q_t^{\pi_2}(u_t, \pi_2(u_t)) - Q_t^{\pi_1}(u_t, \pi_2(u_t)) \right) \\
 & \underbrace{Q_{t+1}^{\pi_1}(u_{t+1}, \pi_2(u_{t+1})) - Q_{t+1}^{\pi_2}(u_{t+1}, \pi_2(u_{t+1}))}_{Q_{t+1}^{\pi_1} + Q_{t+2}^{\pi_2} + \dots} \text{ مانندلا}
 \end{aligned}$$

$$\sum E(Q_{t+1}^{\Delta})$$