

400104964

ثابت دو

ثابت سه

$$d(U(v_1), U(v_2)) = d(R + \gamma p v_1, R + \gamma p v_2) = \gamma d(p v_1, p v_2)$$

(١) انت

من در اینجا اثباتی نمایم که درنظر داشتم باید باشد:

$$\gamma d(p v_1, p v_2) = \gamma \|p v_1 - p v_2\|_\infty = \gamma \|p(v_1 - v_2)\|_\infty \leq \gamma \|p(1 \cdot \|v_1 - v_2\|_\infty)\|_\infty$$

$$= \gamma \|v_1 - v_2\|_\infty \underbrace{\|p\|}_1 = \gamma \|v_1 - v_2\|_\infty$$

در نتیجه (ج) در مجموع اثبات شده است که $U(v)$ مجموعه ای است که از v به v' می رساند.

$$\text{ب) من دلیل اثبات } U(\vec{v}^n) = \vec{v}^n /$$

$$\|U^n(v) - \vec{v}^n\|_\infty = \|U(U^{n-1}(v)) - U(\vec{v}^n)\|_\infty \leq \gamma \|U^{n-1}(v) - \vec{v}^n\|_\infty = \|U(U^{n-2}(v)) - U(\vec{v}^n)\|_\infty$$

$$\leq \gamma^2 \|U^{n-2}(v) - \vec{v}^n\|_\infty \leq \dots \leq \gamma^n \|v - \vec{v}^n\|_\infty$$

$$0 \leq \lim_{n \rightarrow \infty} \|U^n(v) - \vec{v}^n\|_\infty \leq \lim_{n \rightarrow \infty} \gamma^n \|v - \vec{v}^n\|_\infty \xrightarrow{\gamma < 1} 0 \Rightarrow \lim_{n \rightarrow \infty} U^n(v) = \vec{v}^n$$

در اینجا اثبات شده بود که $U^n(v) \rightarrow \vec{v}^n$ است. اینجا محدودیت مانند دارد:

$$\|TV_1 - TV_2\|_\infty = \|V_1 - V_2\|_\infty \leq \gamma \|V_1 - V_2\|_\infty \Rightarrow V_1 = V_2$$

$$\|V^n - U^{k-1}(v)\|_\infty \xrightarrow{\text{همچنین}} \leq \|V^n - U^k(v)\|_\infty + \|U^k(v) - U^{k-1}(v)\|_\infty <$$

$$= \|B(V^n) - B(U^{k-1}(v))\|_\infty + \epsilon \leq \gamma \|V^n - U^{k-1}(v)\|_\infty + \epsilon \Rightarrow \|V^n - U^{k-1}(v)\|_\infty < \frac{\epsilon}{1-\gamma}$$

$$\|V^n - U^k(v)\|_\infty = \|B(V^n) - B(U^{k-1}(v))\|_\infty \leq \gamma \|V^n - U^{k-1}(v)\|_\infty < \frac{\gamma \epsilon}{1-\gamma} < \frac{\epsilon}{1-\gamma}$$

$$V(23) = 10^* \quad V(18) = 10^* \quad V(17) = 10^* \quad V(22) = 10 \quad V(21) = 0 \quad \therefore Y=1 (\leftarrow) \quad (2)$$

$$V(22) = 10 + \frac{1}{2} (0 - 10) = 5^* \quad V(21) = 0 + \frac{1}{2} (-10 - 0) = -5^* \quad V(20) = -10^* \quad V(16) = -10^* \quad V(12) = -10^*$$

$$V(7) = -10^* \quad V(8) = -10^* \quad V(3) = -10^* \quad V(2) = -10^* \quad V(1) = -10^*$$

مقدار مکافای لوریم در اینجا اسید است

$$V(23) = 10^* \quad V(18) = 10^* \quad V(17) = 10^* \quad V(22) = 0^* \quad V(21) = -10^* \quad \therefore Y=1 (\leftarrow)$$

$$V(20) = -10^* \quad V(16) = -10^* \quad V(12) = -10^* \quad V(7) = -10^* \quad V(8) = -10^* \quad V(3) = -10^*$$

$$V(2) = -10^* \quad V(1) = -10^*$$

$$V_M^\pi(s) = E^\pi \left[\sum_{t=0}^{\infty} r_t^+ r_t^- | S_0 = s \right] = E^\pi \left[r_0^- + \sum_{t=1}^{\infty} r_t^+ r_t^- | S_0 = s \right] \quad (\leftarrow) \quad (3)$$

$$= E^\pi \left[r_0^- | S_0 = s \right] + \gamma E^\pi \left[\sum_{t=1}^{\infty} r_t^+ r_{t+1}^- | S_0 = s \right] = \sum_{a \in A} \pi(s, a) R(s, a) + \gamma \sum_{a \in A} \pi(s, a)$$

$$\sum_{s' \in S} T(s, a, s') E^\pi \left[\sum_{t=0}^{\infty} r_t^+ r_{t+1}^- | S_0 = s, A_0 = a, S_1 = s' \right] \quad \xrightarrow{\text{markov property}}$$

$$= \sum_{a \in A} \pi(s, a) R(s, a) + \gamma \sum_{a \in A} \pi(s, a) \sum_{s' \in S} T(s, a, s') E^\pi \left[\sum_{t=0}^{\infty} r_t^+ r_{t+1}^- | S_1 = s' \right]$$

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

باید راه را بر تنه های سعی کنیم که میتوانیم این طریق تا نهایت تله سیم را بگیریم و در مطابقت با این روش میتوانیم $V^\pi(s)$ را محاسبه کنیم.

با این روش میتوانیم $V^\pi(s)$ را محاسبه کنیم و این مقدار را میتوانیم در مطابقت با این روش میتوانیم $V_M^\pi(s)$ را محاسبه کنیم.

$$V_M^\pi(s) = V_{M'}^\pi(s) \quad \forall s \Rightarrow V_M^\pi = V_{M'}^\pi$$

$$V^\pi(s) \cdot E^\pi[G | S_0 = s] = E^\pi[G_t | S_t = s]$$

روزن تمهی خود مطالعه بخوبی انجام شود

(ج) π^* هي策略 M 的一个子集，且 π^* 是 M 中的一个策略，即 $\pi^* \in M$ (因为 M 是一个策略集)

π^* 是 M 中的一个策略 $\Rightarrow \forall \pi \in M: E^{\pi^*} \left[\sum_{t=0}^{\infty} \gamma^t R_t \right] \geq E^{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R_t \right]$

$$E^{\pi^*} \left[\sum_{t=0}^{\infty} \gamma^t \alpha R_t \right] = \alpha E^{\pi^*} \left[\sum_{t=0}^{\infty} \gamma^t R_t \right] \geq \alpha E^{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R_t \right] = E^{\pi} \left[\sum_{t=0}^{\infty} \gamma^t \alpha R_t \right] \quad \forall \pi \in M$$

$\Rightarrow \pi^*$ 是 M 中的一个策略

(ج) π^* 是 M 中的一个策略 $\Rightarrow \pi^*$ 是 M 中的一个策略

选择 π_1 和 π_2 使得 $E^{\pi_1} \left[\sum_{i=0}^{K_1} \gamma^i R_i \right] \geq E^{\pi_2} \left[\sum_{i=0}^{K_2} \gamma^i R_i \right]$

$$E^{\pi_1} \left[\sum_{i=0}^{K_1} \gamma^i R_i \right] = E^{\pi_2} \left[\sum_{i=0}^{K_2} \gamma^i R_i \right] \quad K_1 \neq K_2$$

添加一个阶段 c 到 π_1 和 π_2

$$E^{\pi_1} \left[\sum_{i=1}^{K_1} \gamma^i R_i + c \right] = E^{\pi_1} \left[\sum_{i=1}^{K_1} \gamma^i R_i \right] + c \sum_{i=1}^{K_1} \gamma^i = E^{\pi_2} \left[\sum_{i=1}^{K_1} \gamma^i R_i \right] + c \sum_{i=1}^{K_1} \gamma^i$$

$$E^{\pi_2} \left[\sum_{i=1}^{K_2} \gamma^i R_i \right] + c \sum_{i=1}^{K_2} \gamma^i = E^{\pi_2} \left[\sum_{i=1}^{K_2} \gamma^i R_i \right] + c \sum_{i=1}^{K_2} \gamma^i$$

选择 c 使得 $E^{\pi_1} \left[\sum_{i=1}^{K_1} \gamma^i R_i + c \right] \geq E^{\pi_2} \left[\sum_{i=1}^{K_2} \gamma^i R_i \right]$

(ج) $K_1 = K_2$ 时， π^* 是一个终止状态 s^* ，即 $\pi^*(s^*) = 1$

$K_1 = K_2 = K$ 或 $K_1 = K_2 = \infty$ 时， π^* 是一个非终止状态 s^*

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

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

$$\gamma^2 \sum_{a \in A} \pi(s, a) \sum_{s' \in S} T(s, a, s') \sum_{a'' \in A} \pi(s', a'') R(s'', a'') = \dots$$

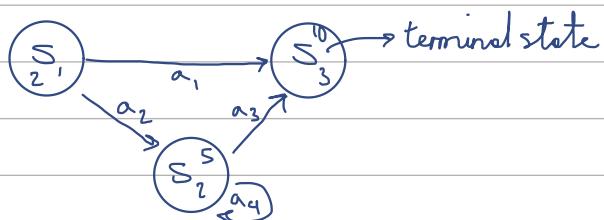
从上式可以看出 $V^\pi(s)$ 是一个递归的方程，即 $V^\pi(s) = \sum_{a \in A} \pi(s, a) \sum_{s' \in S} T(s, a, s') V^\pi(s')$ (递归方程)

$$\forall s: V_2^{\pi^*}(s) = V_1^{\pi^*}(s), R_1(s,a) = R_2(s,a) \quad \text{مقدار مرتبت راسيم من راصب سيم رسيون تابع لـ} \quad \text{در سمع}$$

$$\exists s, a: a \notin \pi^*(s), a \in \pi(s) \Rightarrow R_2(s,a) = R_1(s,a) - c \quad \text{در سمع باقى مقدار مرتبت راسيم صادر می شود}$$

$$\forall \pi \in \Pi_{\pi^*} \quad V_2^{\pi}(s) = E[\sum_{t=0}^{\infty} \gamma^t R_+^{(2)} | S_t = s] = E[\sum_{t=0}^{\infty} \gamma^t R_+^{(1)} | S_t = s] - \gamma^k c \leq V_1^{\pi}(s) \leq V_1^{\pi^*}(s) = V_2^{\pi^*}(s)$$

در سمع باقى مقدار مرتبت راسيم صادر می شود



ادعیه ۴: این مقدار مرتبت راسيم صادر می شود.

$$V^{\pi}(s) = \sum_{a \in A} \pi(s,a) p(s,a,s') [R(s,a) + \gamma V^{\pi}(s')] \quad \text{در معادله مقدار مرتبت راسيم صادر می شود}$$

$$\pi(s_1) = a_1, \quad \pi(s_2) = a_3 \rightarrow V^{\pi}(s_3) = 10, \quad V^{\pi}(s_2) = 5 + 10 = 15 \quad V^{\pi}(s_1) = 2 + 10 = 12$$

$$V^{\pi}(s_1) = 0 \quad V^{\pi}(s_2) = \frac{0-2}{1} = -2 \quad V^{\pi}(s_3) = \frac{0-2}{1} + \frac{-2-5}{1} = -7 \quad \text{در معادله داده شده مقدار مرتبت راسيم صادر می شود}$$

اعلیٰ این نتیجه از نظر مقدار مرتبت راسيم صادر می شود، هر برآرد هاست این است که مقدار مرتبت راسيم صادر می شود.

$$V_+^{\pi}(s) = E[G_+ | S_+ = s, \pi] = E[\sum_{k=0}^{\infty} \gamma^k R_{t+k} | S_+ = s, \pi] = \quad (\text{ب})$$

$$\sum_{a \in A} \pi(s,a) [R(s,a) + \gamma \sum_{s' \in S} \tau(s,a,s') E[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | S_+ = s, A_+ = a, S_{t+1} = s', \pi]]$$

$$= \sum_{a \in A} \pi(s,a) R(s,a) + \gamma \sum_{a \in A} \pi(s,a) \sum_{s' \in S} \tau(s,a,s') \sum_{a' \in A} \pi(s',a') R(s',a')$$

$$\gamma^2 \sum_{a \in A} \pi(s,a) \sum_{s' \in S} \tau(s,a,s') \sum_{a \in A} \pi(s',a') \sum_{s'' \in S} \tau(s',a',s'') \sum_{a'' \in A} \pi(s'',a'') R(s'',a'')$$

$$= \sum_{a \in A} p(A_0 = a | S_0 = s) R(s,a) + \gamma \sum_{a \in A} p(A_0 = a | S_0 = s) \sum_{s' \in S} p(S_1 = s' | A_0 = a, S_0 = s)$$

$$\sum_{a' \in A} (A_1 = a' | S_1 = s) R(s',a') + \dots = E[\sum_{t=0}^{\infty} \gamma^t R_+ | S_0 = s, \pi] = E[G | S_0 = s, \pi] = V^{\pi}(s)$$

$$V^\pi(S_+) = V^\pi(S_+) = E^\pi \left[\sum_{k=0}^{\infty} \gamma R_{k+1} | S_+ = s_+ \right] \xrightarrow[\text{sequence of states is deterministic}] {\gamma=1} E^\pi \left[\sum_{k=0}^{\infty} R_{k+1} \right] = E^\pi[R_+] + E \left[\sum_{k=0}^{\infty} R_{k+1} \right] = E^\pi[R_+] + E \left[\sum_{k=0}^{\infty} R_{k+1} | S_+ = s_+ \right]$$

$$= E^\pi[R_+] + V^\pi(S_{t+1}) = E^\pi[R_+] + V^\pi(S_{t+1}) \xrightarrow{E^\pi[R_+] < 0} V^\pi(S_{t+1})$$

$$G_+^{(n)} = R_{t+1} + \gamma R_{t+2} + \dots + \gamma^{n-1} R_{t+n} + \gamma^n V(S_{t+n})$$

$$V(S_+) = V(S_+) + \alpha (G_+^{(n)} - V(S_+))$$

در حالت $n=1$ از اینجا $V(S_0) = 0.5 + \alpha \times 0 = 0.5$ / اگر برای امتیاز داریم:

$$G_0^{(1)} = 0 + V(S_1) = 0.5 \quad V(S_1) = 0.5 + \alpha \times 0 = 0.5 \quad G_1^{(1)} = 1 \quad V(S_2) = 0.5 + 0.1 \times 0.5 = 0.55$$

در سه دراین حالت مقادیر S_2 یعنی E نمیتواند باشد.

در حالت $n=2$ داریم: $V(S_0)$ بین قصیر و $V(S_1), V(S_2)$ امیدت منزه شود.

و در حالت $n=3$ مقادیر S_3 امیدت منزه شود

ب) در این متن α -variance tradeoff برقرار است. اندیکاتور α صیغه $\alpha = \frac{1}{1 + \text{Variance}}$ تابع متأثرات جبری است.

لذا دفعه زیاد منزه شود. اندیکاتور α صیغه $\alpha = \frac{1}{1 + \text{Variance}}$ دفعه زیاد منزه شود.

زیاد منزه / در عمل هر دفعه بعده افزایش ضمای RMS منزه شود

(c) با اضافه کردن تعادل استحکامی مانند بقیه پارامترها، ماتماتیکا حالت بیشتری برای طبقه بندی داریم آن

بعواصم حالت های جبری هم بوسیله دفعه اضافی داریم بسیار منزه داریم افزایش منزه شود.

با افزایش تعادل ایندیکاتور α ماتماتیکا داریم ر مقادیر ارزش های ماتماتیکا اصلی تریکسر منزه شوند (LOLN)

محض داریم و به صور این طبقه منزه شود

اصحاف درین سهاد تبارها از مدل به اشاره طرف دسی باشد رسیده RMSE را همین دهد در عین این صورت من برای این

سبک اورستیشن مدل را تابع مطابد بصرد با داده های صبیر کرد و این را انتساب می کند.

$$E_0(S) = 0 \quad E_{t+1}(S) = \gamma \lambda E_{t+1}(S) + 1(S_{t+1} = S)$$

$$E_t(S) = \gamma \lambda E_{t-1}(S) + 1$$

از ابعادی به عامل به همان حالت داشته باشند مردد داریم:

$$= 1 + \gamma \lambda + (\gamma \lambda)^2 E_{t-2} = 1 + \gamma \lambda + (\gamma \lambda)^2 + (\gamma \lambda)^3 E_{t-3} = \dots = \sum_{n=0}^{t-1} (\gamma \lambda)^n$$

$$E_{t+1}(S) = \sum_{n=0}^{\infty} (\gamma \lambda)^n = \frac{1}{1 - \gamma \lambda} = \frac{1}{1 - 0.2} = 1.25$$

ام رسم داریم: