part 1

物理屋のための機械学習講義 8

。強化学習

0. 強化学習の問題設定

$s \in S \quad a \in A$

- 。環境 : $\mathcal{E} = \{S, A, \underline{r}(\{s_t\},\{a_t\}), \underline{p_T}(s_{t+1} | \{s_t\},\{a_t\})\}$
- 。エージェント : $ag = \{\underline{\pi}(a_t | \{s_t\},\{a_{t-1}\})\}$

$s_0 \quad \underline{R} = \sum_{t=0}^{T} r(\{s_t\},\{a_t\}) \qquad \{a_t\} = \{a_1, \cdots, a_t\}$

を最大化するような π を見つけたい

$\max_{\pi} \mathbb{E}_{\pi, p_T}[R | s_0]$ …… 目的関数

- エージェントが $r(\{s_t\},\{a_t\})$ と $p_T(s_{t+1} | \{s_t\},\{a_t\})$ を知ってるかどうか？
  - → 知ってる : プランニング問題
  - → 知らない : (狭義の)強化学習

$$\begin{cases} \underline{r}(\{s_t\},\{a_t\}) = r(s_t, a_t) \\ \underline{p_T}(a_{t+1} | \{s_t\},\{a_t\}) = p_T(a_{t+1} | s_t, a_t) \end{cases} \Rightarrow \text{Markov的な環境}$$

$\pi(a_t | \{s_t\},\{a_{t-1}\}) = \pi(a_t | s_t)$ Markov的方策

以降は簡単のため、環境がMarkovの場合に話を絞る。

0.1 Markov方策の十分性

☆ 環境がMarkov かつ 目的関数が各時刻の周辺確率で書けるならば、Markov方策を考えれば十分.

step1

$$\Pr(S_t = s, A_t = a \mid s_0, \pi^H) = \Pr(S_t = s, A_t = a \mid s_0, \pi^M) \quad \text{—①}$$

を示したい…
(帰納法を使う)

$$\pi^M_t(a|s) := \frac{\Pr(S_t = s, A_t = a \mid s_0, \pi^H)}{\Pr(S_t = s \mid s_0, \pi^H)} \quad \text{—②}$$

$t=0$ のとき. $\Pr(S_t = s_0 \mid s_0, \pi^H) = 1 = \Pr(S_t = s_0 \mid s_0, \pi^M)$

② で定義すれば ① を満たされる.

$t=k$ のときに ① が成立すると する.

$t=k+1$ のとき.

$$\Pr(S_{k+1} = s, A_{k+1} = a \mid s_0, \pi^H) = \pi^M_{k+1}(a|s) \underline{\Pr(S_{k+1} = s \mid s_0, \pi^H)}$$

$$\sum_{s', a'} P_T(s \mid a, s') \Pr(S_k = s', A_k = a' \mid s_0, \pi^H)$$

$$= \sum_{s', a'} P_T(s \mid a, s') \times \Pr(S_k = s', A_k = a' \mid s_0, \pi^M) \quad \left(\text{全てMarkovで書けている}\right)$$

$$:= \Pr(S_{k+1} = s, A_{k+1} = a \mid s_0, \pi^M) \; //$$

$f(\pi, s_0) = f(\{\Pr(S_t, a_t \mid \pi, s_0)\})$ で書けているなら.

ex) $\mathbb{E}_\pi[R \mid s_0]$ 　 Markov方策 π^M で十分 f を最大化可能

◦プランニング問題

$$\underline{R = \sum_t r(s_t, a_t)}$$

$$\underline{\mathbb{E}_\pi[R|s_0]}$$

安直にはこれを最大化したいが、代わりに以下を考える.

$$C = \sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \quad (0<\gamma<1)$$

(本当はγ=1を考えたいけど、γ<1の方が性質が良い)

$$V^\pi(s_0) = \mathbb{E}_\pi[C|s_0]$$ ….(状態)価値関数

$$V^*(s_0) = \max_\pi \mathbb{E}_\pi[C|s_0]$$ .. 最適価値関数.

$$= \max_{\pi_0} \mathbb{E}_{\pi_0}\Big[\underbrace{r(s_0,a_0) + \gamma \sum_{s_1} P_T(\quad ,s_0,a_0)\, V^*(s_1)}_{X(s,a)}\Big]$$

$$= \max_{a_0}\Big[r(s_0,a_0) + \gamma \sum_{s_1} P_T(s_1|\quad ,a_0)\, V^*(s_1)\Big]$$

$$= B_* V^*(s_0)$$

$$B^* v(s_1) = \max_{a_0}\Big[r(s_0,a_0) + \gamma \sum_{s_1} P_T(s_1|s_0,a_0)\, v(s_1)\Big]$$

… 最適ベルマン作用素

$$\pi^*(a|s) = \delta_{a,a'(s)} \qquad a'(s) = \operatorname*{argmax}_a X(s,a)$$

⇒ V^*が分かれば、π^*(目的関数を最大化する最適方策)がわかる.

⇒ V^*を考える上では、$\max_\pi \mathbb{E}_\pi$ が $\max_a$ に置きかわるので、定常的なMarkov方策を考えれば十分最大化できる!

(前節では、Markov方策の十分性は示したが、方策の時間依存性が残っていた)

○ ベルマン作用素の縮小性

$$v(s) - V^{\pi}(s) = d(s) \qquad V^{\pi} = B_{\pi} V^{\pi}$$
$$\mathbb{E}_{\pi}[\quad]$$

$$v'(s) = B_{\pi} v = \sum_a \pi(a|s)\{r(s,a) + \gamma \sum p_r(V^{\pi} + d)\}$$

$$= V^{\pi} + \gamma \sum_a \pi(a|s) p_r(s'|s,a) d(s')$$

$$v'(s) - V^{\pi}(s) = \gamma \sum_a \pi(a|s) p_r(s'|s,a) d(s')$$

$$|v' - V^{\pi}|(s) = |\gamma \sum_a \pi(a|s) p_r(\cdots) \quad d(s')|$$

$$\leqq \gamma \sum_a \pi(a|s) p_r(s'|s,a) |d(s')|$$

$$\leq \gamma \underbrace{\sum_{a,s'} \pi(a|s) p_r(\quad)}_{=1} \max_{s'} |d(s')|$$

$$\max_s |v'(s) - V^{\pi}(s)| \leq \gamma \max_{s'} |d(s')|$$

$$v - V^{\pi} = d(s)$$

$$v' = B_{\pi} v \qquad v' - V^{\pi} = |d'(s)| \leqq \gamma \max |d(s)|$$

$\pi \to \pi^*$ B^* についても同様

⇨ 縮小性から解の唯一性も言える。 プランニング問題の場合分かってる。

⇨ 縮小性があるので、ベルマン作用素が分かれば、価値関数も計算可

○ ベルマン方程式と最小作用の原理

$$V^* = \max_a \left[r(s,a) + \gamma \sum_{s'} P_T(s'|s,a) V^*(s') \right]$$

$\gamma = 1$. $P_T(s'|s,a) = \delta_{s', F(s,a)}$

$$s' = F(s,a) = s + f(a)\delta t$$

$S(x,a)$

$s = x$

$x' - x = f(a)\delta t$

$f(a) = \dot{x}$

$r(s,a) = L(x,a)$

$$\frac{\partial S(x,a)}{\partial a} \doteq 0 \Rightarrow \frac{\partial L}{\partial a} + \frac{\partial}{\partial a}\{V^*(x')\} = 0$$

$$\frac{\partial}{\partial a}\{V^*(x')\} = \frac{\partial V^*}{\partial x'} \frac{\partial f}{\partial a} \delta t$$

$$V^* = \max_\pi \mathbb{E}_\pi \left[\sum_t r_t | x \right]$$

$$\sum_t \frac{\partial L(x_t, a_t)}{\partial x_t} \frac{\partial f}{\partial a} \delta t$$

$r(x, \dot{x})$

$= \frac{1}{2} m \dot{x} - \frac{1}{2} k \dot{x}^2$

$$\int dt \frac{\partial L}{\partial x_t} \frac{\partial f}{\partial a}$$

$$\frac{\partial f}{\partial a} = 1$$

$a = \dot{x}$

$$\Rightarrow \frac{\partial L}{\partial \dot{x}} + \int dt \frac{\partial L}{\partial x} = 0 \Rightarrow \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) + \frac{\partial L}{\partial x} = 0$$