Calendly — book your slots 1 hrs in advance ☆

Piazza post @42 (say a few words about your qns when booking)

Stochastic linear (contextual) bandit

Application: personalized recommendation

For $t = 1, 2, \dots T$:

observes $x_t \in \mathcal{X}$ (context)

takes action $a_t \in A = \{1 \dots k\}$

receive reward $r_t = f(x_t, a_t) + \underline{\varepsilon_t} \sim N(0, 1)$.

$\phi(x, a) = e_a = (0 \dots 1 \dots)$ deterministic fn unknown

$$f(x, a) = \langle \underbrace{\theta^*}_{\text{unknown}}, \underbrace{\phi(x, a)}_{\text{known}} \rangle$$

$$\mathbb{E}_{\varepsilon_t}[r_t] = f(x_t, a_t)$$

goal: maximize $\mathbb{E}\left[\sum_{t=1}^{T} r_t\right] = \mathbb{E}\left[\sum_{t=1}^{T} f(x_t, a_t)\right]$.

performance measure: pseudo-regret

$$R_T = \mathbb{E}\left[\cancel{\max_a \sum_{t=1}^{T} f(x_t, a)} \; \underline{\sum_{t=1}^{T} \max_a f(x_t, a)} - \sum_{t=1}^{T} f(x_t, a_t)\right]$$

$$\sum_{t=1}^{T} \max_a \left(f(x_t, a) + \varepsilon_t(a)\right) - \sum_{t=1}^{T} \left(f(x_t, a_t) + \varepsilon_t(a_t)\right)$$

Assume: $\|\theta^*\|_2 \le 1$. $\forall x, a$, $\|\phi(x, a)\|_2 \le 1$.

... good algorithm w/ small

Q1: How can we design a good alg ... pseudo regret ?

$$\left(x_s, a_s, \quad r_s = \langle \theta^*, \overset{\phi_s}{\boxed{\phi(x_s, a_s)}} \rangle + \varepsilon_s \right)_{s=1}^{t-1}$$

can we learn smth about θ^* ?

Q2 given $\hat{\theta}_t$. how to take actions ?

$$a_t = \underset{a \in A}{\arg\max} \underbrace{\langle \hat{\theta}_t, \phi(x_t, a) \rangle + \text{exploration}(a)}_{UCB_t(a)} .$$

$$UCB_t(a) \geq f(x_t, a)$$

Q1: $$\theta_t(\lambda) = \underset{\theta \in \mathbb{R}^d}{\arg\min} \sum_{s=1}^{t-1} \left(\langle \theta, \phi_s \rangle - r_s \right)^2 + \lambda \|\theta\|_2^2 .$$

Informally: $$\theta_t(0) = V_{t-1}^{-1} \left(\sum_{s=1}^{t-1} \phi_s r_s \right)$$

$$V_{t-1} = \sum^{t-1} \phi_s \phi_s^T \quad . \quad \phi_s^T \theta^*$$

$V_{t-1} \quad \sum_{s=1}$

$V_{t-1} \quad r_s = \langle \theta^*, \phi_s \rangle + \varepsilon_s$

$$\theta_t(0) = V_{t-1}^{-1}\left(\underline{\sum_{s=1}^{t-1} \phi_s \phi_s^T \theta^*} + \sum_{s=1}^{t-1} \phi_s \varepsilon_s \right)$$

$$= \theta^* + \underline{V_{t-1}^{-1} \sum_{s=1}^{t-1} \phi_s \varepsilon_s}.$$

$(*)$ $N(0, V_{t-1})$.

$N(0, V_{t-1}^{-1})$

ε_s

a_t

ϕ_t.

given new x, a, construct a confidence interval of $\langle \theta^*, \phi(x,a) \rangle$ based on $\theta_t(0)$?

$$\langle \underbrace{\theta_t(0)}_{x}, \underbrace{\phi(x,a)}_{a.} \rangle \sim N\left(\langle \theta^*, \phi(x,a) \rangle, \|\phi(x,a)\|^2_{V_{t-1}^{-1}}\right)$$

$$X \sim N(0^d, \Sigma_{d\times d})$$

$$a^T X \sim N(0, a^T \Sigma a)$$

By subgaussian tail property: w.p. $1-\delta$:

$$\left| \langle \theta_{t-1}^{(0)}, \phi(x,a) \rangle - \langle \theta^*, \phi(x,a) \rangle \right|$$

$$\le \|\phi(x,a)\|_{V_{t-1}^{-1}} \sqrt{\ln \frac{1}{\delta}}.$$

$$\Rightarrow UCB_t(a) = \langle \phi(x,a), \theta_t^{(0)} \rangle + \|\phi(x,a)\|_{V_{t-1}^{-1}} \sqrt{\ln \frac{1}{\delta}}.$$

problem: ① $(\phi_s)_{s=1}^{t-1}$ depend on $(\varepsilon_s)_{s=1}^{t-1}$,

(*) is wrong.

② V_{t-1} may not be full rank.

Fix: ① (Auer 2002), (Chu, Li, Reyzin,

") specialized binning

Schapire, 2011) :

procedure.

② setting $\lambda > 0$. in $\hat{\theta}_t(\lambda)$.

Q1 : $\hat{\theta}_t(\lambda) = V_{t-1}(\lambda)^{-1} \left(\sum_{s=1}^{t-1} \phi_s r_s \right)$

$(V(\lambda) = V + \lambda I)$.

Lemma : define $\beta_t(\delta) = 1 + \sqrt{2 \ln \frac{1}{\delta} + d \ln(1+\frac{t}{d})}$,

then, $\exists$ event E. $P(E) \geq 1-\delta$, and

on E, $(\bar{\ell}_{t-1}(a))_{a=1}^{K}$

$\| \hat{\theta}_t(1) - \theta^* \|_{V_{t-1}(1)} \leq \beta_t(\delta)$.

for all t.

$\phi(x, a) = e_a$

$$V_{t-1}(1) = \begin{pmatrix} m_{t-1}(1)+1 & & \\ & \ddots & \\ & & m_{t-1}(K)+1 \end{pmatrix}$$

$)^2 \leq O(K)$.

$\forall a, \left| l_{t-1}(a) - l(a) \right| \le \sqrt{\frac{1}{m_{t-1}(a)+1}}$

$\lesssim \sqrt{m_{t-1}(a)+1}$

Pf: $\hat{\theta}_t(1) - \theta^*$

$$= V_{t-1}^{-1}(1)\left(V_{t-1}\theta^* + \sum_{s=1}^{t-1} \phi_s \varepsilon_s \right) - \theta^*$$

$$\theta^* \to V_{t-1}^{-1}(1)\left(V_{t-1} + I\right)\theta^*$$

$$= -V_{t-1}^{-1}(1)\cdot\theta^* + V_{t-1}(1)^{-1} \sum_{s=1}^{t-1} \phi_s \varepsilon_s .$$

$$\left\| \hat{\theta}_t(1) - \theta^* \right\|_{V_{t-1}(1)}$$

$$\le \left\| V_{t-1}^{-1}(1)\, \theta^* \right\|_{V_{t-1}(1)} + \left\| V_{t-1}^{-1}(1) \sum_{s=1}^{t-1} \phi_s \varepsilon_s \right\|_{V_{t-1}(1)}$$

defn of Mahalanobis norm

$$= \left\| \theta^* \right\|_{V_{t-1}^{-1}(1)}^{\preceq I} + \left\| \sum_{s=1}^{t-1} \phi_s \varepsilon_s \right\|_{V_{t-1}^{-1}(1)} .$$

$\|\theta\| \ge \|\theta\|_B \quad B) \preceq A$

$\|\theta\|_A$

$$\leq 1 + \sqrt{2\ln\frac{1}{\delta} + d\ln\left(1+\frac{t}{d}\right)}.$$

self normalized bound.

$\lambda = 1.$

$\boxtimes$

self-normalized bound:

there exists an event E. $P(E) \geq 1-\delta$,

in which

$$\left\| \sum_{s=1}^{t} \phi_s \xi_s \right\|_{V_t^{-1}(\lambda)} \leq \sqrt{2\ln\frac{1}{\delta} + d\ln\left(1+\frac{t}{d\lambda}\right)}$$

(Abbasi-Yadkori, Pal, Szepesvari, 2011)

Q2:

$$\Theta_t = \left\{\theta : \left\| \hat{\theta}_t(1) - \theta \right\|_{V_{t-1}(1)} \leq \beta_t(\delta) \right\}$$

on event E. $\theta^* \in \Theta_t$.

at time step t. construct UCB for

$\langle \theta^*, \phi(x_t, a) \rangle$ for all a.

define on E $\geq \xi(x_t, a)$

$$\underbrace{UCB_t(a)}_{} = \max_{\theta \in \Theta_t} \langle \theta, \phi(x_t, a) \rangle$$

$$= \max_{\|\gamma\|_{V_{t-1}(1)} \leq \beta_t(\delta)} \langle \hat{\theta}_t(1) + \gamma, \phi(x_t, a) \rangle$$

$$= \underbrace{\langle \hat{\theta}_t(1), \phi(x_t, a) \rangle}_{\text{predicted reward}} + \underbrace{\beta_t(\delta) \cdot \|\phi(x_t, a)\|_{V_{t-1}^{-1}(1)}}_{\text{uncertainty for this } (x_t, a)}$$

MAB: $\sqrt{K} \cdot \sqrt{\dfrac{\ln T}{m_{t-1}(a)+1}}$.

Alg: $a_t = \underset{a \in A}{\operatorname{argmax}}\ UCB_t(a)$.

(LinUCB, OFUL)

Analysis:

Thm: for LinUCB:

$$R_T \le O\left(T\delta + \underbrace{\beta_T(\delta)}_{\sqrt{d \ln T}} \sqrt{d \cdot T \ln\left(1+\frac{T}{d}\right)}\right)$$

setting $\delta = \frac{1}{T}$ $\quad R_T \le \tilde{O}(d \sqrt{T})$

(For MAB: $\phi(x,a) = e_a \in \mathbb{R}^K \rightarrow K\sqrt{T}$

UCB: $\sqrt{KT}$.)

Pf: on event $\bar{E}$: contributions to

$$R_T \le 2T \cdot \delta .$$

on event E:

$$\rho_t = \max_a f(x_t, a) - f(x_t, a_t)$$

$$= \max_a UCB_t(a) - f(x_t, a_t)$$

$$= UCB_t(a_t) - f(x_t, a_t)$$

$$= \langle \hat{\theta}_t, \phi_t \rangle + \beta_t(\delta) \|\phi_t\|_{V_{t-1}^{-1}}$$

$$- \langle \theta^*, \phi_t \rangle \quad \le \beta_t$$

$$\|\hat{\theta}_t - \theta^*\|_{V_{t-1}(1)} \|\phi_t\|_{V_{t-1}^{-1}(1)}$$

$$= 2\beta_t(\delta) \cdot \|\phi_t\|_{V_{t-1}^{-1}(1)}$$

$$\le 2\beta_t(\delta)\|\phi_t\|_{V_t(1)^{-1}} \cdot 2.$$

$$\left(V_t(1) \lesssim 2 \cdot V_{t-1}(1) \right)$$

$$\sum_{t=1}^{T} \rho_t \le 4\beta_t(\delta) \cdot \sum_{t=1}^{T} \|\phi_t\|_{V_t(1)^{-1}}$$

C-S

$$\le 4\beta_t(\delta) \cdot \sqrt{T \cdot \sum_{t=1}^{T} \|\phi_t\|^2_{V_t(1)^{-1}}}$$

Elliptic potential

$$\le 4\beta_t(\delta)\sqrt{T d \ln\left(1+\frac{T}{d}\right)}$$

Next: reinforcement learning.