Chap 19. Stochastic Linear Bandits.

V. 0.1

2022.03.20

§19.1 Setting.

- In each round t, the learner choose an action $A_t \in \mathcal{A}_t$.
- Receive a reward satisfying

$$Z_t = \langle \theta_*, A_t \rangle + \eta_t,$$

where θ_* is unknown and η_t is conditionally 1-subgaussian (Condition on $\mathcal{A}_1, A_1, Z_1, \mathcal{A}_2, A_2, Z_2, \dots, \mathcal{A}_t$, and A_t).

- Objective: minimize random regret and expected regret:

$$\hat{R}_n = \sum_{t=1}^{n} \max_{a \in \mathcal{A}_t} \langle \theta_*, a - A_t \rangle$$

○ (Expected Regret 中 Σ 是否穿过 $\mathbb{E}$ 有什么区别？).

$$R_n = \mathbb{E}[\hat{R}_n] = \mathbb{E}\left[\sum_{t=1}^{n} \max_{a \in \mathcal{A}_t} \langle \theta_*, a \rangle - \sum_{t=1}^{n} X_t\right].$$

§19.2. LinUCB.

- Upper Confidence Bound of reward $\langle \theta^*, a \rangle$.

$$UCB_t(a) = \max_{\theta \in C_t} \langle \theta, a \rangle.$$

where C_t is the confidence set of θ^* which we will determine later.

- The learner take A_t in round t as

$$A_t = \arg\max_{a \in \mathcal{A}_t} UCB_t(a).$$

- Estimate θ^* via ridge regression

$$\hat{\theta}_t = \underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \left(\sum_{s=1}^{t} \left(\langle \theta, A_s \rangle - z_s \right)^2 + \lambda \|\theta\|_2^2 \right).$$

The closed form solution of $\hat{\theta}_t$ is

$$\hat{\theta}_t = V_t^{-1} \cdot X_t^T \cdot Y_t ,$$

where $X_t \in \mathbb{R}^{t \times d}$, $X_t[s,:] = A_s^T$, $V_t \in \mathbb{R}^{d \times d}$, $V_t = (X_t^T X_t + \lambda I)$,
$Y_t \in \mathbb{R}^t$, $Y_t[s] = z_s$.

§19.3 Analysis.

§19.3.1 Regularity Assumption.

A1
- $\max\limits_{t \in [n]} \sup\limits_{a,b \in \mathcal{A}_t} \langle \theta^*, a-b \rangle \leq 1$. (Bounded instantaneous regret)

A2
- $\sup\limits_{a \in (\cup_{t=1}^{n} \mathcal{A}_t)} \|a\|_2 \leq L$.

§19.3.2 Regret.

Let $A_t^* = \underset{a \in \mathcal{A}_t}{\operatorname{argmax}} \langle \theta^*, a \rangle$ be the optimal arm in round t.

Define the instantaneous regret

$$r_t = \langle \theta^*, A_t^* - A_t \rangle.$$

Let $\tilde{\theta}_t \in C_t$ s.t. $\langle \tilde{\theta}_t, A_t \rangle = UCB_t(A_t)$. We start from intantaneous regret

$$
\begin{aligned}
r_t = \langle \theta^*, A_t^* - A_t \rangle &\leq UCB_t(A_t^*) - \langle \theta^*, A_t \rangle \\
&\leq UCB_t(A_t) - \langle \theta^*, A_t \rangle \\
&= \langle \tilde{\theta}_t - \theta^*, A_t \rangle \\
&= \| \tilde{\theta}_t - \theta^* \|_{V_{t-1}} \cdot \| A_t \|_{V_{t-1}^{-1}} . \\
&= \| \tilde{\theta}_t - \hat{\theta}_t + \hat{\theta}_t - \theta^* \|_{V_{t-1}} \cdot \| A_t \|_{V_{t-1}^{-1}} . \\
&\leq 2 \cdot \sqrt{\beta_t} \cdot \| A_t \|_{V_{t-1}^{-1}} .
\end{aligned}
$$

Assumption A1 shows that $r_t \leq 1$, which combined with $\beta_n \geq \max\{1, \beta_t\}$. leads to

$$
\begin{aligned}
r_t \leq 1 \wedge 2\sqrt{\beta_t} \cdot \| A_t \|_{V_{t-1}^{-1}} &\leq 2 \wedge 2\sqrt{\beta_t} \cdot \| A_t \|_{V_{t-1}^{-1}} \\
&\leq 2\sqrt{\beta_n} \cdot (1 \wedge \| A_t \|_{V_{t-1}^{-1}}).
\end{aligned}
$$

Then, by Cauchy inequality,

$$
\hat{R}_n = \sum_{t=1}^{n} r_t \leq \sqrt{n \cdot \sum_{t=1}^{n} r_t^2} = 2 \sqrt{n \cdot \beta_n \cdot \sum_{t=1}^{n} (1 \wedge \| A_t \|_{V_{t-1}^{-1}}^2)} .
$$

$\blacksquare$

§19.3.3 Technical Lemmas.

Lemma [Elliptical Potential Lemma]

$$
\sum_{t=1}^{n} (1 \wedge \| a_t \|_{V_{t-1}^{-1}}^2) \leq 2 \cdot \log \left(\frac{\det V_n}{\det V_0} \right) \leq 2d \cdot \log \left(\frac{\operatorname{trace} V_0 + n L^2}{d \cdot \det (V_0)^{\frac{1}{d}}} \right).
$$

Lemma. [Confidence Bound for Least Squares Estimators]

There exists a $\delta \in (0,1)$ such that $\theta^* \in C_t$ holds with w.p. $1-\delta$, where

$$C_t = \{ \theta \in \mathbb{R}^d : \| \theta - \hat{\theta}_{t-1} \|^2_{V_{t-1}} \leq \beta_t \}.$$

with nondecreasing sequence $(\beta_t)_{t=1}^n$.

Remark. 1. We will determine β_t in next chapter.

Remark 2. The confidence set is actually a confidence ellipsoid.

[Geometric meaning of quadratic form].

Lemma [Determinant-Trace Inequality]

Let $V_t = V_0 + \sum_{i=1}^{t} X_i X_i^T$, where $X_i \in \mathbb{R}^d$, $\forall i$. Then

$$\det(V_t) \leq \left(\frac{\operatorname{trace}(V_t)}{d} \right)^d.$$

Proof. AM-GM ineq.

∎