IE 498: Online Learning and Decision Making, Fall 2019 Oct 3rd

Lecture 12: Linear Bandits

Lecturer: Yuan Zhou *Scribe: Tiancheng Zhao, Xinyang Liu*

1 Problem Setting

In linear bandit problems, we have $\vec{\theta} \in R^d$ and $\vec{\theta}$ is normalized so that $\left\|\vec{\theta}\right\|_2 \leq 1$. There are n actions and the time horizon is T.

At time t:

1) Context vectors $\vec{x}_{t,i} \in R^d$ for arm i and $\|\vec{x}_{t,i}\|_2 \leq 1$ is observed by the player,

2) The player decides a_t and receives $r_t = \vec{\theta}^T \cdot \vec{x}_{t,a_t} + \epsilon_t$. ($\epsilon_t \sim N(0,1)$).

Define regret as $R_T \triangleq \mathbb{E} \sum_{t=1}^{T} [\max_{a_t} \{\vec{\theta}^T \cdot \vec{x}_{t,a_t}\} - r_t]$

Note: Such a scenario is called "oblivious adversary" when the adversary knows your strategy and picks all the contexts beforehand but he does not act adaptively.

Remark 1. *Relation to Contextual Bandits:* $\Pi = \{\pi_\theta(\vec{x}_1, ..., \vec{x}_n) = \arg\max_a \{\vec{\theta}^T \cdot \vec{x}_{t,a} \mid \vec{\theta} \in R^d\}\}$

Then, suppose at time t_0, we have made action $\vec{y}_t = \vec{x}_{t,a_t}$ $(t = 1, 2, 3, ..., t_0)$, observed rewards r_t. How can we estimate $\vec{\theta}$?

1.1 Maximum Likelihood Estimator

$$\hat{\vec{\theta}} = \arg\max_{\vec{\theta}} \sum_{t=1}^{t_0} ln P_{\epsilon_t}(r_t - \vec{\theta}^T \cdot y_t) \tag{*}$$

Theorem 1. (MLE theorem)

1) MLE estimator is asymptotically unbiased, $\hat{\vec{\theta}} \xrightarrow[t\to+\infty]{a.s.} \vec{\theta}$.

2) Variance of MLE is asymptotically less or equal to any unbiased estimator (under classical condition).

$$(*) : \hat{\vec{\theta}} = \arg\min_{\vec{\theta}} \sum_{t=1}^{t_0} (r_t - \vec{\theta}^T \cdot y_t)^2 \quad \text{(Ordinary Least Square)}$$

Let $S(\hat{\vec{\theta}}) = \sum_{t=1}^{t_0}(r_t - \vec{\theta}^T \cdot y_t)^2 = \sum_{t=1}^{t_0}[{r_t}^2 + \hat{\vec{\theta}}^T \cdot \vec{y}_t \cdot \vec{y}_t^T \cdot \hat{\vec{\theta}} - 2(\hat{\vec{\theta}}^T \cdot \vec{y}_t) \cdot r_t]$.

Assume $v_{t_0} \triangleq \sum_{t=1}^{t_0} \vec{y}_t \cdot \vec{y}_t^T$ invertible.

$$\text{Set } \frac{\mathrm{dS}(\hat{\vec{\theta}})}{\mathrm{d}\hat{\vec{\theta}}} = 0 \quad \Rightarrow \quad 2(\sum_{t=1}^{t_0} \tilde{y}_t \cdot \tilde{y}_t^T)\hat{\vec{\theta}} = 2\sum_{t=1}^{t_0} \tilde{y}_t \cdot r_t \quad \Rightarrow \quad \hat{\vec{\theta}} = (\sum_{t=1}^{t_0} \tilde{y}_t \cdot \tilde{y}_t^T)^{-1} \cdot (\sum_{t=1}^{t_0} \tilde{y}_t \cdot r_t)$$

1.2 Confidence Region

In this part, our goal is to find γ so that $Pr[|\vec{x}^T(\vec{\theta} - \hat{\vec{\theta}})| > \gamma] < \delta$ given context vector $\vec{x}$.

First, we assume $v_{t_0} \triangleq \sum_{t=1}^{t_0} \vec{y}_t \cdot \vec{y}_t^T$ invertible.

$$\vec{x}^T(\hat{\vec{\theta}} - \vec{\theta}) = \vec{x}^T \cdot [{v_{t_0}}^{-1}(\sum_{t=1}^{t_0}(\vec{y}_t \cdot \vec{y}_t^T) \cdot \vec{\theta} + \sum_{t=1}^{t_0} \epsilon_t \cdot \vec{y}_t) - \vec{\theta}] = \sum_{t=1}^{t_0} \vec{x}^T \cdot {v_{t_0}}^{-1} \cdot \vec{y}_t \cdot \epsilon_t$$

Suppose $\{\epsilon_t\}$ is independent from $\{\vec{y}_t\}$, then $\vec{x}(\hat{\vec{\theta}} - \vec{\theta})$ is a centered *Gaussian* with variance $\left\|{v_{t_0}}^{-\frac{1}{2}}\vec{x}\right\|_2^2$. The variance is calculated as follows:

$$\sum_{t=1}^{t_0} \vec{x}^T \cdot {v_{t_0}}^{-1} \cdot \vec{y}_t \cdot \vec{y}_t^T \cdot {v_{t_0}}^{-1} \cdot \vec{x} = \vec{x}^T \cdot {v_{t_0}}^{-1}(\sum_{t=1}^{t_0} \vec{y}_t \cdot \vec{y}_t^T){v_{t_0}}^{-1} \cdot \vec{x} = \vec{x}^T \cdot {v_{t_0}}^{-1} \cdot \vec{x} = \left\|{v_{t_0}}^{-\frac{1}{2}}\vec{x}\right\|_2^2$$

Therefore, we set $\gamma = c \cdot \left\|{v_{t_0}}^{-\frac{1}{2}}\vec{x}\right\|_2 \sqrt{log(\frac{1}{\delta})}$.

Then, we consider singular v_{t_0}.

Consider $\hat{\vec{\theta}} = (I + v_{t_0})^{-1} \sum_{t=1}^{t_0} \epsilon_t \vec{y}_t$ which is Ridge Regression formula.

$$\vec{x}^T(\hat{\vec{\theta}} - \vec{\theta}) = \vec{x}^T[(I + v_{t_0})^{-1}(\vec{\theta} + v_{t_0} \cdot \vec{\theta} + \sum_{t=1}^{t_0} \epsilon_t \vec{y}_t - \vec{\theta}) - \vec{\theta}] = \vec{x}^T \cdot (I + v_{t_0})^{-1} \cdot \sum_{t=1}^{t_0} \epsilon_t \vec{y}_t \; - \; \vec{x}^T \cdot (I + v_{t_0})^{-1} \cdot \vec{\theta}$$

Note that:

1) $\vec{x}^T \cdot (I + v_{t_0})^{-1} \cdot \sum_{t=1}^{t_0} \epsilon_t \vec{y}_t$: centered *Gaussian* with variance smaller or equal to $\sum_{t=1}^{t_0} \vec{x}^T \cdot (I + v_{t_0})^{-1} \cdot \vec{x}$.

   Since $\sum_{t=1}^{t_0} \vec{x}^T \cdot (I + v_{t_0})^{-1} \cdot \vec{y}_t \cdot \vec{y}_t^T \cdot (I + v_{t_0})^{-1} \cdot \vec{x} \leq \sum_{t=1}^{t_0} \vec{x}^T \cdot (I + v_{t_0})^{-1} \cdot \vec{x}$ due to $v_{t_0} \leq I + v_{t_0}$.

2) $|\vec{x}^T \cdot (I + v_{t_0})^{-1} \cdot \vec{\theta}| \;\leq\; \left\|(I + v_{t_0})^{-\frac{1}{2}}\vec{x}\right\|_2 \left\|(I + v_{t_0})^{-\frac{1}{2}}\vec{\theta}\right\|_2 \;\leq\; \left\|(I + v_{t_0})^{-\frac{1}{2}}\vec{x}\right\|_2$

Therefore, we set $\gamma = c \cdot \left\|(I + v_{t_0})^{-\frac{1}{2}}\vec{x}\right\|_2 \sqrt{log(\frac{1}{\delta})}$.

2 Linear UCB algorithm

Based on the above discussion, we introduce the Linear UCB algorithm as the follows:

LinUCB: At time t, do the following:

1) set $U_{t-1} = I + \sum_{z=1}^{t-1} \vec{y_z}\vec{y_z}^{-1}$, $\hat{\vec{\theta}} = U_{t-1}^{-1} \sum_{z=1}^{t-1} \vec{y_z} r_z$

2) select $a_t = \arg\max_{i\in[n]}\{\hat{\vec{\theta}_t}\vec{x}_{t,i} + c \left\| U_{t-1}^{-1/2}\vec{x}_{t,i} \right\|_2 \sqrt{\log(T^2 n)}\}$

Analysis:

Let event $E = \{\forall t, i, \vec{x}_{t,i}^T |\hat{\vec{\theta}_t} - \vec{\theta}| \leq c\|U_{t-1}^{-1/2}\vec{x}_{t,i}\|_2\sqrt{\log(T^2n)}\}$, which is the event that our estimated return for each arm is within a confidence interval, and denote $CI_{t,i} = c\|U_{t-1}^{-1/2}\vec{x}_{t,i}\|_2\sqrt{\log(T^2n)}$.

By union bound, we have $Pr[E] \geq 1 - \frac{1}{T}$.

Remark 2. *The above inequality is not proved, since to prove it with union bound, we need y_t and ε_t to be independent. However, in the MLE estimation, y_t depend on previous ε_t. In other words, the decision we made on time t depends on previous noises. We'll continue using this inequality to give an intuition, and fix this issue in future lectures.*

Conditioned on E, $\forall t$, we have

$$\mathbb{E}[\text{Regret incurred at time} t] = \max_i\{\vec{x}_{t,i}^T\} - \vec{y_t}\vec{\theta} = \vec{x}_{t,i^*}^T\vec{\theta} - \vec{y_t}^T\vec{\theta},$$

where i^* is the maximizer.

$$\begin{aligned}
\vec{x}_{t,i^*}^T\vec{\theta} - \vec{y_t}^T\vec{\theta} &\leq \vec{x}_{t,i^*}^T\hat{\vec{\theta}}_t + |\vec{x}_{t,i^*}^T(\hat{\vec{\theta}}_t - \vec{\theta})| - \vec{y_t}\vec{\theta} \\
&\leq \vec{x}_{t,i^*}^T\vec{\theta}_t + CI_{t,i^*} - \vec{y_t}^T\vec{\theta} \\
&\leq \vec{x}_{t,a_t}\hat{\vec{\theta}} + CI_{t,a_t} - \vec{y_t}^T\vec{\theta} \\
&\leq 2CI_{t,a_t}
\end{aligned}$$

Lemma 2. (Elliptical potential lemma)

$$\sum_{t=1}^{T} \|U_{t-1}^{1/2}\vec{y_t}\|^2 \leq 2d\ln(\frac{T}{d} + 1),$$

And by Cauchy-Schwartz,

$$\sum_{t=1}^{T} \|U_{t-1}^{1/2}\vec{y_t}\|^2 \leq \sqrt{T2d\ln(\frac{T}{d} + 1)}$$

Proof. For every $t > 1$, $U_{t-1} \succeq I$ by definition, so we have:

$$\|U_{t-1}^{-1/2}\vec{y_t}\|_2 \leq \|\vec{y_t}\|_2 \leq 1.$$

Since by definition,

$$\begin{aligned}
U_t &= U_{t-1} + \vec{y_t}\vec{y_t}^T \\
&= U_{t-1}^{1/2}(I + U_{t-1}^{-1/2}\vec{y_t}\vec{y_t}^T U_{t-1}^{-1/2})U_{t-1}^{1/2},
\end{aligned}$$

we have:

$$\det(U_t) = \det(U_{t-1}) \det(I + U_{t-1}^{-1/2} \vec{y}_t \vec{y}_t^T U_{t-1}^{-1/2}) = \det(U_{t-1})(1 + \|U_{t-1}^{-1/2} \vec{y}_t\|_2^2).$$

Since $1 + x \geq e^x$ for $x \in [0, 1]$,

$$\begin{aligned}\det(U_t) &\geq \det(U_{t-1}) \exp(\frac{1}{2} \|U_{t-1}^{-1/2} \vec{y}_t\|_2^2) \\ &\geq \det(U_1) \exp(\frac{1}{2} \sum_{z=1}^{t} \|U_{z-1}^{-1/2} \vec{y}_z\|_2^2)\end{aligned}$$

$$\begin{aligned}\sum_{t=1}^{T} \|U_{t-1}^{-1/2} \vec{y}_t\|_2^2 &\leq 2 \ln \det(U_T) \\ \text{(due to AM-GM)} \\ &\leq 2 \ln(\mathrm{Tr}(U_T)/d)^d \\ &\leq 2 \ln(\frac{d+T}{d})^d \\ &\leq 2d \ln(\frac{T+d}{d})\end{aligned}$$

□

Plug in the lemma, we have

$$R_T \lesssim \mathbb{E} \sum_{t=1}^{T} CI_{t,a_i} \lesssim \mathbb{E} \sum_{t=1}^{T} \|U_{t-1}^{1/2} \vec{y}_t\|_2 \sqrt{\log(T^n)} \lesssim \sqrt{dT \log T \log(Tn)}$$