IE 498: Online Learning and Decision Making, Fall 2019 10/8/2019

Lecture 14: Lower Bounds for Linear Bandits

Lecturer: Yuan Zhou *Scribe: Ali Bibak*

Recap

In previous section, we discussed linear contextual bandits, and showed that a regret of $O(\sqrt{dT\text{poly}\log(nT)})$ or $O(d\sqrt{T\text{poly}\log T})$ is achievable, where the latter does not depend on n. The choice of which bound to use, therefore, depends on how large n is. For example, as we discussed in the previous lecture, in the infinitely-many arms case the former regret bound is not helpful due to its dependence on n.

The objective of this lecture will be to investigate if the provided upper bounds are in fact tight. In this lecture, we answer this question (almost) affirmatively by providing bounds that are short of the polylogarithmic factor of T. That is, a lower bound of $\Omega(\sqrt{dT\log n})$ and $\Omega(d\sqrt{T})$.

1 First Lower Bound

As we have seen in previous lectures, KL divergence is often a reliable tool when proving lower bounds. Hence we briefly recall the definition of KL divergence:

Definition 1 (Kullback-Leibler divergence)**.** *For continuous distributions P and Q, the Kullback-Leibler (KL) divergence between them is defined as*

$$\mathcal{D}_{KL}(P \parallel Q) := \int_{dP} \left(\ln \frac{dQ}{dP}\right)\, dP.$$

Since we are working with linear bandits, it suffices to consider Gaussian distributions. We proceed by finding the KL divergence of two Gaussian distributions with the same variance, which is immediate from definition.

Fact 1. *We have that*

$$\begin{aligned}\mathcal{D}_{KL}\left(\mathcal{N}(\mu_1,\sigma^2) \parallel \mathcal{N}(\mu_2,\sigma^2)\right) &= \int_{-\infty}^{+\infty} \left(\frac{1}{\sqrt{2\pi\sigma^2}} - \exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right) dx\right).\frac{(x-\mu_1)^2-(x-\mu_2)^2}{2\sigma^2} \\ &= \frac{1}{2\sigma^2}(\mu_1-\mu_2)^2\end{aligned}$$

In the next theorem, we prove a lower bound of $\Omega(d\sqrt{T})$ for linear bandits. This lower bound corresponds to the upper bound of $O(d\sqrt{T\text{poly}\log T})$ derived in the pervious lecture.

Theorem 2. *For T and d with $T \geq d^2$, let action set $\mathcal{A}_t = \mathcal{A} = \left\{\pm\sqrt{\frac{d}{T}}\right\}^d$. For any policy π, there exists $\vec{\theta} \in \left\{\pm\sqrt{\frac{1}{d}}\right\}^d$ such that*

$$\mathcal{R}^{\pi}_{T,\vec{\theta}} \gtrsim d\sqrt{T}.$$

Proof. Let $P_{t,\vec{\theta}}$ be the probability distribution of $\{\vec{y}_1, r_1, \vec{y}_2, r_2, \ldots, \vec{y}_t, r_t\}$ when hidden vector is $\vec{\theta}$. Using Fact 1 we have that

$$\mathcal{D}_{KL}(P_{t,\vec{\theta}} \parallel P_{t,\vec{\theta'}}) = \sum_{z=1}^{t} \mathbb{E}\, \mathcal{D}_{KL}\left(D\left(r_z \mid \vec{y}_z, \vec{\theta}\right) \parallel D\left(r_z \mid \vec{y}_z, \vec{\theta'}\right)\right) = \frac{1}{2}\sum_{z=1}^{t} \mathbb{E}\left[\left(\vec{y}_z^{\intercal}(\vec{\theta} - \vec{\theta'}\right)^2\right].$$

Let $\vec{\theta}^{\oplus i} := \left(\vec{\theta}_1, \vec{\theta}_2, \ldots, \vec{\theta}_{i-1}, -\vec{\theta}_i, \vec{\theta}_{i+1}, \ldots, \vec{\theta}_d\right)$. That is, the vector $\vec{\theta}$ whose i-th coordinate is negated.

Note that for all $\vec{y} \in \mathcal{A}$ and $i \in \{1, 2, \ldots, d\}$, by using a simple calculation we have that $\left(\vec{y}^{\intercal}\left(\vec{\theta} - \vec{\theta}^{\oplus i}\right)\right)^2 = \frac{4}{T}$. Therefore, $\mathcal{D}_{KL}(P_{t,\vec{\theta}} \parallel P_{t,\vec{\theta}^{\oplus i}}) = \frac{2t}{T}$ for all $\vec{\theta} \in \left\{\pm\sqrt{\frac{1}{d}}\right\}^d$ and $i \in \{1, 2, \ldots, d\}$.

For all $i \in \{1, 2, \ldots, d\}$ and sign variable $b \in \{\pm 1\}$, let $E_{i,b}$ be the event that

$$\left|\left\{t \in \{1, 2, \ldots, T\} : \operatorname{sgn}\left((\vec{y}_t)_i\right) \neq b\right\}\right| \geq \frac{T}{2}.$$

That is, the sign of i-th coordinate of at least half of $\vec{y}_i$'s does not agree with b. Noting that at least half of $\vec{y}_i$'s are in agreement with b in the sign of their i-th coordinate if and only if at most half of them are in such an agreement with $-b$, we have that

$$\begin{aligned}
\Pr[E_{i,b} \mid \vec{\theta}] + \Pr[E_{i,-b} \mid \vec{\theta}^{\oplus i}] &= \Pr[E_{i,b} \mid \vec{\theta}] + \Pr[\overline{E_{i,b}}, \vec{\theta}^{\oplus i}] \\
&\geq 1 - \left|\Pr[E_{i,b} \mid \vec{\theta}] - \Pr[E_{i,b} \mid \vec{\theta}^{\oplus i}]\right| \\
&\geq \frac{1}{2}\exp(-\mathcal{D}_{KL}(P_{T,\vec{\theta}} \parallel P_{T,\vec{\theta}^{\oplus i}})) \\
&\geq \frac{1}{2}e^{-2}.
\end{aligned}$$

Now, for simplicity, let $q_{i,\theta} := \Pr[E_{i,\vec{\theta}_i} \mid \vec{\theta}]$. Note that we have just showed that $q_{i,\vec{\theta}} + q_{i,\vec{\theta}^{\oplus i}} \geq \frac{1}{2}e^{-2}$. We are now ready to calculate the regret. First, notice that by definition of our events, we have that

$$\mathcal{R}^{\pi}_{T,\vec{\theta}} \geq \sum_{i=1}^{d} q_{i,\vec{\theta}} \cdot \frac{T}{2} \cdot 2\sqrt{\frac{1}{T}} = \sqrt{T}\sum_{i=1}^{d} q_{i,\vec{\theta}}. \tag{1}$$

Next, we calculate the average regret over all possible $\vec{\theta}$. This average regret equals

$$\begin{aligned}
\frac{1}{2^d}\sum_{\vec{\theta}} \mathcal{R}^{\pi}_{T,\vec{\theta}} &\geq \frac{\sqrt{T}}{2^d}\sum_{\vec{\theta}}\sum_{i=1}^{d} q_{i,\vec{\theta}} && \text{From (1)} \\
&= \frac{\sqrt{T}}{2^d}\sum_{i=1}^{d}\sum_{\vec{\theta}} q_{i,\vec{\theta}} && \text{(Changing the order of summation)} \\
&= \frac{\sqrt{T}}{2^d}\sum_{i=1}^{d}\sum_{\vec{\theta}} \frac{q_{i,\vec{\theta}} + q_{i,\vec{\theta}^{\oplus i}}}{2} && \text{(Pairing } \vec{\theta} \text{ and } \vec{\theta^{\oplus i}} \text{ together)} \\
&\geq \frac{\sqrt{T}}{2^d}\sum_{i=1}^{d} \frac{1}{4}e^{-2}.2^d && \text{(From } q_{i,\vec{\theta}} + q_{i,\vec{\theta}^{\oplus i}} \geq \frac{1}{2}e^{-2}\text{)} \\
&\geq d\sqrt{T}\frac{e^{-2}}{4}.
\end{aligned}$$

Since there exists an instance that is at least the average, the theorem statement follows.

□

2 Second Lower Bound

Next, we prove a lower bound of $\Omega(\sqrt{dT \log n})$ for linear bandits. This lower bound corresponds to the upper bound of $O(\sqrt{dT \text{poly} \log(nT)})$ derived in the previous lecture.

Corollary 3. *For all $n = 2^k$ ($k \in \{1, 2, \ldots, d\}$) and all policies π, there exists a d-dimensional linear bandit instance I such that*

$$\mathcal{R}^{\pi}_{T,I} \gtrsim \sqrt{dTk} = \sqrt{dT \log n}.$$

Proof. The idea here is to break the instance into smaller k dimensional instances. Suppose we have $\beta = \frac{d}{k}$ many k-dimensional, n-arm, horizon $\frac{T}{\beta}$, instances $I_1, I_2, \ldots, I_\beta$ such that $\mathcal{R}^{pi}_{\frac{T}{\beta}, I_j} \gtrsim k\sqrt{\frac{T}{\beta}}$. Construct $I = I(I_1, I_2, \ldots, I_\beta)$ as follows

- Divide d dimensions into β blocks of size k
- Divide T into β consecutive periods, each having $\frac{T}{\beta}$ time steps
- Let the hidden vector be $\vec{\theta} = (\vec{\theta}_1, \vec{\theta}_2, \ldots, \vec{\theta}_\beta)$, where $\vec{\theta}_j \in I_j$.

Feature vectors at time $\tau = (i-1)\frac{T}{\beta} + t$ offer arms at time t from I_j. That is, for all $i \in \{1, 2, \ldots, n\}$, $\mathcal{X}_{i,\tau} = \left(\mathbf{0}^\intercal, \ldots, \mathbf{0}^\intercal, \mathcal{X}^{(j)}_{i,t}, \mathbf{0}^\intercal, \ldots, \mathbf{0}^\intercal\right)$, where the non-zero entries are located in the j-th block.

Hence for all policies π, there exists policies $\pi_1, \pi_2, \ldots, \pi_\beta$ such that $\mathcal{R}^{\pi}_{T,I} = \sum_{j=1}^{\beta} \mathcal{R}^{\pi_j}_{\frac{T}{\beta}, I_j}$.

Using Theorem 2, we can find instances $I_1, I_2, \ldots, I_\beta$ such that $\mathcal{R}^{\pi}_{T,I} = \sum_{j=1}^{\beta} \mathcal{R}^{\pi_j}_{\frac{T}{\beta}, I_j} \gtrsim \sum_{j=1}^{\beta} k\sqrt{\frac{T}{\beta}} = k\sqrt{T\beta} = k\sqrt{\frac{Td}{k}} = \sqrt{dTk} = \sqrt{dT \log n}$.

□