Lecture 12
Page 1

CS 59000 - RL
Linear bandit
Agenda.
- Regret bound
- More structure?
- Thompson Sampling

Theorem: The regret of LinUCB satisfies,

$$\hat{R}_n \leq \sqrt{8 n \beta_n(\delta) \log\left(\frac{\det V_n(\lambda)}{\det(V)}\right)}$$

with probability at least $1-\delta$.

Before proving this theorem, let us study the following lemma.

Lemma: Let V be a positive definite matrix, $v = \text{trace}(V)$ and $x_1, \dots, x_n \in \mathbb{R}^d$ with $\|x_i\|_2 \leq L < \infty$ a sequence of vectors. Let $V_t(V) = \sum_{s=1}^{t} x_s x_s^T + V$. then:

$$\sum_{t=1}^{n} \left(1 \wedge \|x_t\|^2_{V^{-1}_{t-1}(V)}\right) \leq 2 \log\left(\frac{\det V_n(V)}{\det V}\right)$$

Proof: Using the fact that for any $u \geqslant 0$, $u \wedge 1 \leqslant 2\log(1+u)$

why?

we have

$$\sum_{t=1}^{n} \left(1 \wedge \|x_t\|^2_{V_t(V)^{-1}}\right) \leqslant 2 \sum_{t=1}^{n} \log\left(1 + \|x_t\|^2_{V_{t-1}(V)^{-1}}\right)$$

on the other hand, $V_t(V) = V_{t-1}(V) + x_t x_t^T$

$$= V_{t-1}(V)^{\frac{1}{2}} \left(I + V_{t-1}(V)^{-\frac{1}{2}} x_t x_t^T V_{t-1}(V)^{-\frac{1}{2}}\right) V_{t-1}(V)^{\frac{1}{2}}$$

Therefore:

$$\det\left(V_t(V)\right) = \det\left(V_{t-1}(V)\right) \det\left(I + V_{t-1}(V)^{-\frac{1}{2}} x x_t^T V_t(V)^{-1}\right)$$

$$= \det\left(V_{t-1}(V)\right) \left(1 + \|x_t\|^2_{V_{t-1}^{-1}(V)}\right)$$

Ergo

$$\det\left(V_t(V)\right) = \det(V) \underbrace{\prod_{t=1}^{n} \left(1 + \|x_t\|_{V_{t-1}(V)^{-1}}\right)}$$

Hence,

$$\sum_{t=1}^{n} \left(1 \wedge \|x_t\|^2_{V_{t-1}(V)^{-1}}\right) \leqslant$$

$$< 2 \log\left(\frac{\det(V_n(V))}{\det(V)}\right)$$

which is the statement of the lemma.

Page 3

Proof of the regret theorem:

Consider the per step regret of Lin UCB:

what is it? $\langle A_t^*, \theta^* \rangle - \langle A_t, \theta^* \rangle$

$A_t^* = \underset{a \in D_t}{\arg\max} \langle a, \theta^* \rangle$

Let's play a little bit.

$$\langle A_t^*, \theta^* \rangle - \langle A_t, \theta^* \rangle \leq \langle A_t, \tilde{\theta} \rangle - \langle A_t, \theta^* \rangle$$

$A_t = \underset{a \in D_t}{\arg\max} \langle a, \tilde{\theta}_t \rangle$

$\langle A_t^*, \theta^* \rangle \leq \langle A_t, \tilde{\theta} \rangle$

where optimism kicks in

$$= \langle A_t, \tilde{\theta}_t - \theta^* \rangle$$

$$\leq \|A_t\|_{V_{t-1}^{-1}(V)} \|\tilde{\theta}_t - \theta^*\|_{V_{t-1}(V)}$$

$\langle A_t, \tilde{\theta}_t - \theta^* \rangle = \underbrace{A_t^T V_{t-1}^{-\frac{1}{2}}(V)} \underbrace{V_{t-1}(V)^{\frac{1}{2}} (\tilde{\theta} - \theta^*)}$

$$\leq \|A_t\|_{V_{t-1}^{-1}(V)} \left(\|\tilde{\theta}_t - \hat{\theta}_{t-1} + \hat{\theta}_{t-1} - \theta^*\| \right)$$

$$\leq \|A_t\|_{V_{t-1}^{-1}(V)} \left(2\sqrt{\beta_{t-1}(\delta)} \right)$$

on the other hand we know that

$$|\langle A_t^*, \theta^* \rangle - \langle A_t, \theta^* \rangle| \leq 2$$

Page 4:

$$\langle A_t^*, \theta^* \rangle - \langle A_t, \theta^* \rangle \leq 2 \wedge \left(2 \|A\|_{V_{t-1}(V)} \sqrt{\beta_t(\delta)} \right)$$

when $\beta_n(\delta) \geq 1 \vee \beta_t(\delta)$ we have

why?

$$\langle A_t^*, \theta^* \rangle - \langle A, \theta^* \rangle \leq 2\sqrt{\beta_n(\delta)} \left(1 \wedge \|A_t\|_{V_t(V)^{-1}} \right)$$

Finally, using Jenson's inequality we have

$$\hat{R}_n = \sum_{t=1}^{n} \langle A_t^*, \theta^* \rangle - \langle A_t, \theta^* \rangle$$

$$\leq \sqrt{n \sum_{t=1}^{n} \left(\langle A_t^*, \theta^* \rangle - \langle A_t, \theta^* \rangle \right)^2}$$

$$\leq 2\sqrt{n \beta_n \sum_{t=1}^{n} \left(1 \wedge \|A_t\|_{V_t(V)^{-1}} \right)}$$

$$\leq 2\sqrt{2n \beta_n \log \left(\frac{\det(V_n(V))}{\det V} \right)}$$

Remark, setting $V = \lambda I$, where $\det(V) = \lambda^d$ and the fact that $\det(V_n(\lambda)) \leq \left(\frac{\lambda + nL^2}{d} \right)^d$

we have $\hat{R}_n \leq 2\sqrt{2n\beta_n \left(d \log \left(\frac{\lambda + nL^2/d}{\lambda} \right) \right)}$

Remark:

$$\hat{R}_n \leq 2\sqrt{2nd\,\beta_n \log\left(\frac{\lambda + \frac{nL^2}{d}}{\lambda}\right)}$$

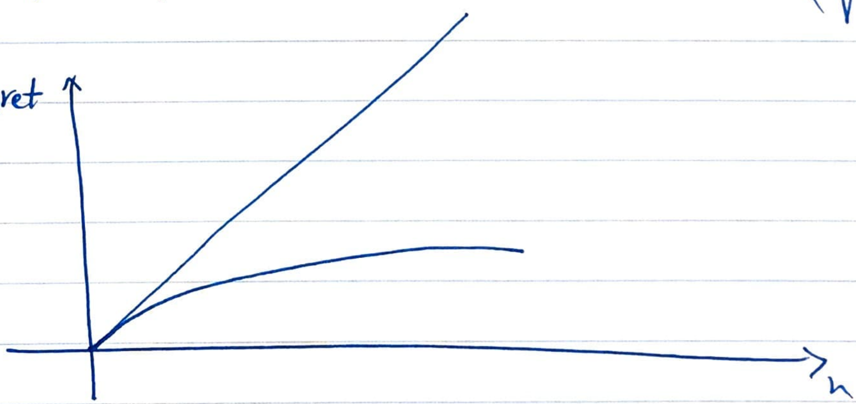
$$\leq 2\sqrt{2nd \log\left(\frac{\lambda + \frac{nL^2}{d}}{\lambda}\right)}\left(\sqrt{\lambda}\, S + \sqrt{d \log\left(\frac{1 + \frac{nL^2}{d\lambda}}{\delta^{2/d}}\right)}\right)$$

where we used $\sqrt{\beta_n(\delta)} = \sqrt{\lambda}\, S + \sqrt{2\log\left(\frac{1}{\delta}\right) + d\log\left(\frac{\lambda + \frac{nL^2}{d}}{\lambda}\right)}$

Remark

$$\hat{R}_n = \tilde{O}\left(d\sqrt{n}\right)$$

your per step regret vanishes with rate $\tilde{O}\left(\frac{d}{\sqrt{n}}\right)$



$\Omega\left(d\sqrt{n}\right) \rightarrow$ lower bound is $\sqrt{n}$

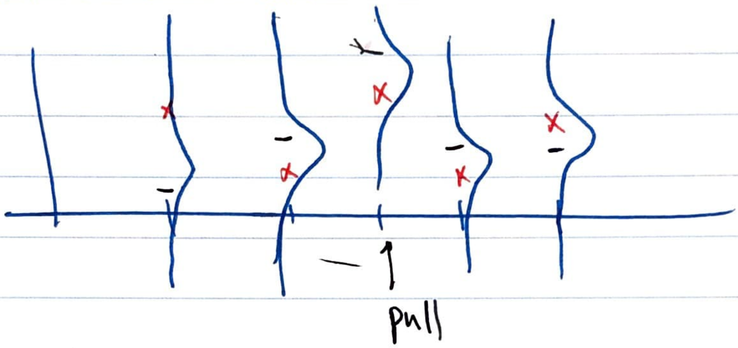
$\Theta\left(d\sqrt{n}\right)$ or $\theta\left(d\sqrt{n}\right)$, Algorithm matches the lower bound.

Page 6

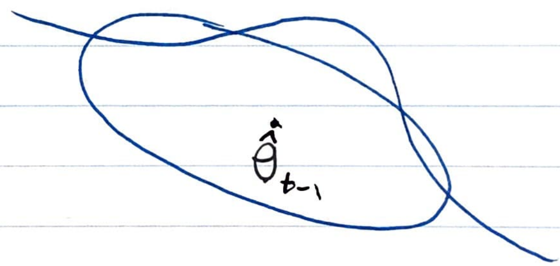
Is ther any other approach than aptimism?

Thompson Sampling. [Thompson Sampling for Contextual Bandit with Linear Pay off.
Linear Thompson Sampling Revisited.

For multi-armed bandit:



For linear bandit



I draw a θ_{TS} then

$$\|\theta - \hat{\theta}_{t-1}\|^2_{V_{t-1}} \leqslant \beta_{t-1}(\delta)$$

$$A_t = \arg\max \langle a, \theta_{TS} \rangle$$

Saprse linear bandit: we assume θ^* is an spare vector means $\|\theta^*\|_0 \leqslant S$.

Low dimensional bandit: Arms are close to a linear subspace with smal dimension

---

Adversarial linear bandit: $\theta_1 \cdots \theta_n$, $X_t = \langle A_t, \theta_t \rangle$