Lecture #8: contextual/linear bandits

Setting 1 (contextual bandits)

For each round $t = 1, \dots, T$:

- agent observes context $c_t \in \mathcal{C}$ (arbitrarily chosen by nature)
- agent chooses action $a_t \in [K]$ — a_t is measurable w.r.t. $\mathcal{F}_{t-1} = \sigma(U_0, c_1, Y_1, U_1, \dots, Y_{t-1}, U_{t-1}, c_t)$
- agent observes and gets reward Y_t

  where $Y_t = r(a_t, c_t) + \eta_t$

  with $\eta_t \mid \mathcal{F}_{t-1}$ is 1 sub-Gaussian, 0 mean

  $$\begin{cases} \mathbb{E}[\eta_t \mid \mathcal{F}_{t-1}] = 0 \\ \forall \lambda \in \mathbb{R}, \ \mathbb{E}[e^{\lambda \eta_t} \mid \mathcal{F}_{t-1}] \leq \exp\left(\frac{\lambda^2}{2}\right) \end{cases}$$

$r : [K] \times \mathcal{C} \to \mathbb{R}$ is called the reward function ← object to estimate

(pseudo)-regret defined as: $R_T = \sum_{t=1}^{T} \max_{k \in [K]} r(k, c_t) - r(a_t, c_t)$

Without any assumption on r, independent bandit games for each context c.

- First possibility, r is "regular" (e.g. Lipschitz or Hölder) (see exercise sheet)

- A common assumption is that r is linear with respect to a known feature map $\Psi : [K] \times \mathcal{C} \to \mathbb{R}^d$ and a parameter $\theta^* \in \mathbb{R}^d$ (to estimate) such that

$$r(k, c) = \langle \theta^*, \Psi(k, c) \rangle \qquad \forall k, c.$$

This is equivalent to the following setting, with $\mathcal{A}_t = \{\psi(k, C_t) \mid k \in [K]\}$:

Setting 2 (linear bandits)

For each round $t = 1, \dots, T$:

- agent observes decision set $\mathcal{A}_t \subset \mathbb{R}^d$
- agent chooses action $a_t \in \mathcal{A}_t$ — a_t is measurable w.r.t. $\mathcal{F}_{t-1} = \sigma(U_0, C_1, Y_1, U_1, \dots, Y_{t-1}, U_{t-1}, C_t)$.
- agent observes and gets reward Y_t

  where $Y_t = \langle \theta^*, a_t \rangle + \eta_t$

  with $\eta_t \mid \mathcal{F}_{t-1}$ is 1 sub-Gaussian, 0 mean

Particular cases:

- $\mathcal{A}_t = \{e_1, \dots, e_d\} \longrightarrow$ classical multi-armed bandits with d arms and $\mu_k = \theta^*_k$.
- $\mathcal{A}_t \subset \{0,1\}^d \longrightarrow$ combinatorial bandits.

We want to build an adaptation of UCB for linear bandits, called

LinUCB.

The idea is to construct confidence sets $\mathcal{C}_t$ such that $\theta^* \in \mathcal{C}_t$ with high probability and pick at each round

$$a_t \in \operatorname*{argmax}_{a \in \mathcal{A}_t} \max_{\theta \in \mathcal{C}_t} \langle a, \theta \rangle$$

(with $\mathcal{C}_t$ as small as possible)

UCB score of arm a.

Before the confidence set, what is the estimate of θ^*? (ie "empirical mean")

Regularised least-squares estimator:

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

linear regression

$\lambda \geq 0$ is the penalty factor (or regularization parameter)

$\lambda > 0$ ensures uniqueness of the minimiser

We can indeed easily check that:

$$\hat{\theta}_t = V_t^{-1} \sum_{s=1}^{t} a_s Y_s \qquad \text{where } V_t = \lambda I_d + \sum_{s=1}^{t} a_s a_s^T$$

For any symmetric, positive definite matrix $M \in \mathbb{R}^{d \times d}$ and vector $u \in \mathbb{R}^d$, we denote

$$\|u\|_M^2 := \left(u^T M u\right)$$

Theorem (linear bandits concentration)

For any $\delta \in (0,1)$, $t \in \mathbb{N}$ and $\lambda > 0$:

$$\mathbb{P}\left(\|\hat{\theta}_t - \theta^*\|_{V_t} \geq \sqrt{\lambda}\,\|\theta^*\|_2 + \sqrt{2 \ln\left(\frac{1}{\delta}\right) + \ln\left(\frac{\det(V_t)}{\lambda^d}\right)} \right) \leq \delta$$

The proof relies on the following concentration lemma

Lemma

Let $S_t = \sum_{s=1}^{t} Y_s a_s$

For any $\lambda > 0, t \in \mathbb{N}$ and $\delta \in (0,1)$,

$$\mathbb{P}\left(\|S_t\|^2_{V_t^{-1}} \geq 2 \ln\left(\frac{1}{\delta}\right) + \ln\left(\frac{\det(V_t)}{\lambda^d}\right)\right) \leq \delta$$

Proof of the Theorem (based on lemma)

Note that $\hat{\theta}_t = V_t^{-1}\left(S_t + \sum_{s=1}^{t} a_s a_s^T \theta^*\right)$

$$= V_t^{-1} S_t + V_t^{-1}(V_t - \lambda I_d)\theta^*$$

So $\|\hat{\theta}_t - \theta^*\|_{V_t} = \| V_t^{-1} S_t - \lambda V_t^{-1} \theta^* \|_{V_t}$

$$\leq \|V_t^{-1} S_t\|_{V_t} + \lambda \|V_t^{-1}\theta^*\|_{V_t}$$

$$= \|S_t\|_{V_t^{-1}} + \lambda \underbrace{\|\theta^*\|_{V_t^{-1}}}_{\sqrt{\theta^{*T} V_t^{-1} \theta^*}}$$

$\sqrt{\theta^{*T} V_t^{-1} \theta^*} \leq \|V_t^{-1}\|_{op}^{1/2} \|\theta^*\|_2$

$\leq \lambda_{min}(V_t)^{-1/2} \|\theta^*\|_2$

$\leq \lambda^{-\frac{1}{2}} \|\theta^*\|_2$

$$\leq \|S_t\|_{V_t^{-1}} + \sqrt{\lambda}\, \|\theta^*\|_2.$$

$\square$

Proof of the lemma

For any $x \in \mathbb{R}^d$, define $\quad M_t(x) = \exp\left(\langle x, S_t \rangle - \frac{1}{2} \|x\|^2_{V_t - \lambda I_d} \right)$

1) We show by induction that $M_t(x)$ is a surpermartingale, so that

$$\mathbb{E}[M_t(x)] \leqslant M_0(x) = 1$$

$t \to t+1$

$$V_{t+1} = V_t + a_{t+1} a_{t+1}^T$$

$$M_{t+1}(x) = \exp\left(\langle x, S_{t+1} \rangle - \frac{1}{2} \left(x^T (V_{t+1} - \lambda I) x \right) \right)$$

$$= M_t(x) \cdot \exp\left(\langle x, a_{t+1} \rangle \eta_{t+1} - \frac{1}{2} \langle x, a_{t+1} \rangle^2 \right).$$

$$\mathbb{E}[M_{t+1}(x) | \mathcal{F}_t] \leqslant M_t(x)$$

$(\eta_{t+1} | \mathcal{F}_t$ is 1 sub-Gaussian$)$.

2) Let $\nu = \mathcal{N}(0, \lambda^{-1} I_d)$.

$$\bar{M}_t = \int M_t(x) \, d\nu(x)$$

is also a surpermartingale by Tonelli and

$$\bar{M}_t = \frac{1}{\sqrt{(2\pi)^d \lambda^{-d}}} \int_{\mathbb{R}^d} \exp\Big(\underbrace{\langle x, S_t \rangle - \frac{1}{2} \|x\|^2_{V_t - \lambda I} - \frac{1}{2} \|x\|^2_{\lambda I}}_{} \Big) dx$$

$$= x^T S_t - \frac{1}{2} x^T V_t x$$

$$= -\frac{1}{2} (x - V_t^{-1} S_t)^T V_t (x - V_t^{-1} S_t) + \frac{1}{2} S_t^T V_t^{-1} S_t$$

$$= -\frac{1}{2}\|x - V_t^{-1}S_t\|_{V_t}^2 + \frac{1}{2}\|S_t\|_{V_t^{-1}}^2$$

$$\bar{M}_t = \exp\left(\frac{1}{2}\|S_t\|_{V_t^{-1}}^2\right) \cdot \left(\frac{\lambda}{2\pi}\right)^{d/2} \int_{\mathbb{R}^d} \underbrace{\exp\left(-\frac{1}{2}\|x - V_t^{-1}S_t\|_{V_t}^2\right)}_{\text{up to rescaling, pdf of } \mathcal{N}(V_t^{-1}S_t, V_t)} dx$$

$$= \exp\left(\frac{1}{2}\|S_t\|_{V_t^{-1}}^2\right) \frac{\lambda^{d/2}}{\sqrt{\det(V_t)}}$$

$$\|S_t\|_{V_t^{-1}}^2 = 2\ln(\bar{M}_t) - \ln\left(\frac{\lambda^d}{\det(V_t)}\right)$$

3)

$$\mathbb{P}\left(\|S_t\|_{V_t^{-1}}^2 \geq 2\ln\left(\frac{1}{\delta}\right) + \ln\left(\frac{\det(V_t)}{\lambda^d}\right)\right) = \mathbb{P}\left(\ln(\bar{M}_t) \geq \ln\left(\frac{1}{\delta}\right)\right)$$

$$= \mathbb{P}\left(\bar{M}_t \geq \frac{1}{\delta}\right) \leq \mathbb{E}[\bar{M}_t]\,\delta \leq \delta. \qquad \square$$

Algo LinUCB: suppose we know m with $\|\theta^*\|_2 \leq m$

For each $t \in \mathbb{N}$

Play $a_t \in \underset{a \in A_t}{\arg\max} \; \underset{\theta \in \mathcal{C}_{t-1}}{\max} \langle \theta, a_t \rangle$

can be computed efficiently for our specific form of $\mathcal{C}_t$ and nice A_t.

with $\hat{\theta}_t = \underset{\theta \in \mathbb{R}^d}{\arg\min} \sum_{s=1}^{t} \left(Y_s - \langle \theta, a_s \rangle \right)^2 + \lambda \|\theta\|_2^2$ $\qquad V_t := \lambda I + \sum_{s=1}^{t} a_s a_s^T$

and $\mathcal{C}_t = \left\{ \theta \in \mathbb{R}^d \;\middle|\; \|\hat{\theta}_t - \theta\|_{V_t} \leqslant \sqrt{\lambda}\, m + \sqrt{4 \ln(t) + \ln\left(\frac{\det(V_t)}{\lambda^d}\right)} \right\}$

Theorem:

If $\|\theta^*\|_2 \leqslant m$ and for any t, $\sup_{a \in \mathcal{A}_t} \|a\|_2 \leqslant L$, then the regret of LinUCB satisfies for any $\lambda > 0$:

$$\mathbb{E}[R_T] \leqslant c \sqrt{T m^2 \lambda + \ln(T) T + d T \ln\left(1 + \frac{T L^2}{d \lambda}\right)} \sqrt{d \ln\left(1 + \frac{T L^2}{d \lambda}\right)} + c_2 m L$$

for univ constants c, c_2

Corollary:

Taking $\lambda = 1$ and considering the main factor in T we have:

$$\mathbb{E}[R_T] = O\left(d \sqrt{T} \ln T \right)$$

Comments:

- distribution free bound
- if $\mathcal{A}_t$ is finite, and the same for every t, we can get a $\log(T)$ instance dependent bound.

Proof:

Let us bound the instantaneous regret first.

$$r_t = \langle \theta^*, A_t^* - a_t \rangle \qquad \text{when } A_t^* \in \operatorname{argmax}_{a \in \mathcal{A}_t} \langle \theta^*, a \rangle .$$

Define the good event
$$\mathcal{E}_t = \left\{ \theta^* \in \mathcal{C}_{t-1} \right\}$$

Thanks to our concentration theorem, $\mathbb{P}(\neg \mathcal{E}_t) \leqslant \frac{1}{(t-1)^2}$

So
$$\mathbb{E}[r_t] \leqslant mL\, \mathbb{P}(\neg \mathcal{E}_t) + \mathbb{E}[r_t \mathbb{1}_{\mathcal{E}_t}]$$
$$\leqslant \frac{mL}{(t-1)^2} + \mathbb{E}[r_t \mathbb{1}_{\mathcal{E}_t}] .$$

if $\mathcal{E}_t$, $\theta^* \in \mathcal{C}_{t-1}$ so:

$$\langle \theta^*, A_t^* \rangle \leqslant \max_{\theta \in \mathcal{C}_{t-1}} \langle \theta, A_t^* \rangle$$

$$\leq \max_{\theta \in \mathcal{C}_{r-1}} \langle \theta, a_r \rangle$$ by defn of a_r.

$$= \langle \tilde{\theta}_r, a_r \rangle \text{ for some } \tilde{\theta}_r \in \mathcal{C}_{r-1}.$$

Cauchy Schwarz gives:

$$r_r = \langle \theta^*, A_r^* - a_r \rangle \leq \langle \tilde{\theta}_r - \theta^*, a_r \rangle \leq \|\tilde{\theta}_r - \theta^*\|_{V_{r-1}} \|a_r\|_{V_{r-1}^{-1}}$$

$$\leq \|a_r\|_{V_{r-1}^{-1}} \left(\|\tilde{\theta}_r - \hat{\theta}_{r-1}\|_{V_{r-1}} + \|\theta^* - \hat{\theta}_{r-1}\|_{V_{r-1}} \right)$$

$$\leq 2 \|a_r\|_{V_{r-1}^{-1}} \cdot \underbrace{\left(\sqrt{\lambda}\, m + \sqrt{4 \ln(r_{-1}) + \ln\left(\frac{\det(V_{r-1})}{\lambda^d}\right)} \right)}_{\text{define } \alpha_r = \max(\,\cdot\,,\ mL)}$$

also by assumption, $r_r \leq 2mL$, so

$$r_r \leq 2\alpha_r \left(1 \wedge \|a_r\|_{V_{r-1}^{-1}} \right) \quad (\text{if } \mathcal{E}_r \text{ holds}).$$

overall:

$$R_T \leq \sum_{r=2}^{T} \mathbb{E}\left[r_r \mathbb{1}_{\mathcal{E}_r} \right] + mL \left(\frac{1}{(r-1)^2} \wedge 1 \right) \quad + mL$$

$$\leqslant 2 \sum_{t=2}^{T} \alpha_t \left(1 \wedge \|a_t\|_{V_{t-1}^{-1}} \right) + mL\left(1+\frac{\pi^2}{6}\right)$$

$$\leqslant 2\sqrt{\sum_{t=1}^{T} \alpha_t^2} \sqrt{\sum_{t=1}^{T} \left(1 \wedge \|a_t\|^2_{V_{t-1}^{-1}}\right)} + mL\left(1+\frac{\pi^2}{6}\right)$$

$$\leqslant 2\sqrt{2\sum_{t=1}^{T} \lambda m^2 + 4\ln(t) + \ln\left(\frac{\det(V_t)}{\lambda^d}\right)} \sqrt{\sum_{t=1}^{T} \left(1 \wedge \|a_t\|^2_{V_{t-1}^{-1}}\right)} + mL\left(2+\frac{\pi^2}{6}\right)$$

Bound on: $\ln\left(\frac{\det(V_n)}{\lambda^d}\right)$ $\qquad \prod_{i=1}^{d} \lambda_i \leqslant \left(\frac{\sum_{i=1}^{d} \lambda_i}{d}\right)^d$ (arithmetic vs geometric mean)

$$\frac{\det(V_t)}{\lambda^d} = \det\left(\frac{V_t}{\lambda}\right) \leqslant \left(\frac{\operatorname{tr}\left(\frac{V_t}{\lambda}\right)}{d}\right)^d = \left(\frac{\operatorname{tr}(V_t)}{\lambda d}\right)^d$$

$$\operatorname{tr}(V_t) = \operatorname{tr}\left(\lambda I + \sum_{s=1}^{t} a_s a_s^T\right) = \lambda d + \sum_{s=1}^{t} \underbrace{\operatorname{tr}(a_s a_s^T)}_{= \|a_s\|_2^2}$$

$$\leqslant \lambda d + t L^2.$$

$$\boxed{\text{so } \ln\left(\frac{\det(V_t)}{\lambda^d}\right) \leqslant d \ln\left(1+\frac{tL^2}{\lambda d}\right)}$$

Bound on $\sum_{t=1}^{T} \left(1 \wedge \|a_t\|^2_{V_{t-1}^{-1}}\right)$

$u \wedge 1 \leqslant 2\ln(1+u)$

so $\displaystyle \sum_{t=1}^{T} \left(1 \wedge \|a_t\|^2_{V_{t-1}^{-1}}\right) \leqslant \underbrace{2\sum_{t=1}^{T} \ln\left(1+\|a_t\|^2_{V_{t-1}^{-1}}\right)}_{=\ln\left(\det\left(\frac{V_T}{V_0}\right)\right)}$

Indeed, $V_r = V_{r-1} + a_r a_r^T = V_{r-1}^{1/2}\left(I + V_{r-1}^{-1/2} a_r a_r^T V_{r-1}^{-1/2}\right) V_{r-1}^{1/2}$

so $\det(V_r) = \det(V_{r-1}) \cdot \det\Big(I + \underbrace{V_{r-1}^{-1/2} a_r a_r^T V_{r-1}^{-1/2}}_{yy^T \text{ is a rank one matrix.}}\Big)$

$1 + yy^T$ has eigenvalues: $\Big(\underbrace{1+\|y\|^2}_{\text{eigenvector } y.}, 1, \dots, 1\Big)$

$\det(V_r) = \det(V_{r-1}) \cdot \left(1 + \|V_{r-1}^{-1/2} a_r\|_2^2\right)$

$= \det(V_{r-1})\left(1 + \|a_r\|^2_{V_{r-1}^{-1}}\right)$

So by induction $\ln(\det(V_r)) = \ln(\det(V_0)) + \sum_{s=1}^{r} \ln\left(1 + \|a_r\|^2_{V_{r-1}^{-1}}\right)$

so $\sum_{r=1}^{T} \left(1 + \|a_r\|^2_{V_{r-1}^{-1}}\right) \leq 2\ln\left(\frac{\det(V_T)}{\underbrace{\det(V_0)}_{\lambda^d}}\right)$

$\leq 2d \ln\left(1 + \frac{TL^2}{\lambda d}\right)$ Thanks to previous bound.

In conclusion, gathering everything we get.

$R_T \leq 4\sqrt{2Tm^2\lambda + 4\ln(T)T + d^T \ln\left(1 + \frac{TL^2}{d\lambda}\right)}\sqrt{d\ln\left(1 + \frac{TL^2}{d\lambda}\right)} + mL\left(2 + \frac{\pi^2}{6}\right)$

$\square$

LinUCB has regret $R_T = O\left(d\sqrt{T} \ln T \right)$.

Can we do better?

Theorem (minimax lower bound, linear bandits)

Let $\mathcal{A}_t = [-1, 1]^d$ and $\Theta = \left\{ -\frac{1}{\sqrt{T}}, \frac{1}{\sqrt{T}} \right\}^d$. Then for any algorithm, there exists $\theta^* \in \Theta$ s.t.:

$$\mathbb{E}\left[R_T(\theta^*)\right] \geq \frac{e^{-2}}{8} d\sqrt{T}$$

. Here $m = \sqrt{\frac{d}{T}}$ and $L = \sqrt{d}$. So for $T \geq d$, LinUCB is optimal, up to a $\ln T$ term.

Proof relies on the following lemma, which is the equivalent of the fundamental inequality for linear bandits.

Lemma (fundamental inequality, linear bandits)

For all linear bandit instances, θ and θ' in $\mathbb{R}^d$

for all strategies and events A that are $\sigma(H_T)$-measurable,

law of H_T under θ (and π)

$$\mathbb{E}_\theta\left[\sum_{t=1}^{T} KL\left(\mathcal{N}(\langle a_t, \theta \rangle, 1), \mathcal{N}(\langle a_t, \theta' \rangle, 1)\right)\right] = KL\left(\mathbb{P}_\theta^{H_T}, \mathbb{P}_{\theta'}^{H_T} \right)$$

and $\frac{1}{2} e^{-KL(\mathbb{P}_\theta^{H_T}, \mathbb{P}_{\theta'}^{H_T})} \leq \mathbb{P}_\theta^{H_T}(A) + \mathbb{P}_{\theta'}^{H_T}(A^c)$ (Bretagnolle-Huber inequality)

Proof of the Theorem

Consider any $\theta \in \Theta$ and $i \in [d]$. Define θ' as $\theta'_j = \begin{cases} \theta_j & \text{if } i \neq j \\ -\theta_j & \text{if } i = j \end{cases}$

1) By the above equality:

$$KL\left(\mathbb{P}_\theta^{H_T}, \mathbb{P}_{\theta'}^{H_T}\right) = \mathbb{E}_\theta\left[\sum_{t=1}^{T} KL\left(\mathcal{N}(\langle a_t, \theta\rangle, 1), \mathcal{N}(\langle a_t, \theta'\rangle, 1)\right)\right]$$

KL between Gaussian of same variance

$$= \frac{1}{2} \sum_{t=1}^{T} \mathbb{E}\left[\langle a_t, \theta - \theta'\rangle^2\right]$$

$$\leq \frac{1}{2} \sum_{t=1}^{T} \frac{4}{T} = 2$$

2) Moreover, define the event $A(\theta, i) = \left\{ \sum_{t=1}^{T} \mathbb{1}\{\operatorname{sgn}(a_{it}) \neq \operatorname{sgn}(\theta_{it})\} \geq \frac{T}{2} \right\}$.

Then Bretagnolle-Huber inequality states:

$$\underbrace{\mathbb{P}_\theta(A(\theta, i)) + \mathbb{P}_{\theta'}(A(\theta, i)^c)}_{\parallel} \geq \frac{1}{2} e^{-KL(\mathbb{P}_\theta^{H_T}, \mathbb{P}_{\theta'}^{H_T})} \geq \frac{1}{2} e^{-2}$$

$$\mathbb{P}_\theta(A(\theta, i)) + \mathbb{P}_{\theta'}(A(\theta', i))$$

So that, when summing over all possible θ: (Calveraging hammer argument)

$$\sum_{\theta \in \Theta} \mathbb{P}_\theta(A(\theta, i)) + \mathbb{P}_{\theta'}(A(\theta', i)) = 2 \sum_{\theta \in \Theta} \mathbb{P}_\theta(A(\theta), i) \geq \frac{|\Theta|}{2} e^{-2}$$

Now summing over all i:

$$\frac{1}{|\Theta|} \sum_{\theta \in \Theta} \sum_{i \in [d]} \mathbb{P}_\theta(A(\theta), i) \geq \frac{d}{4} e^{-2}.$$

In particular, there exists $\theta \in \Theta$ s.t. $\sum_{i \in [d]} \mathbb{P}_\theta(A(\theta), i) \geq \frac{d}{4} e^{-2}$.

Consider such a Θ in the following.

Then we can lower bound the regret on Θ:

$$R_T(\Theta) = \sum_{t=1}^{T} \max_{a \in A_t} \langle a - a_t, \theta \rangle$$

$$= \sum_{t=1}^{T} \sum_{i=1}^{d} (\mathrm{sgn}(\Theta_i) - a_{it}) \cdot \Theta_i$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{d} |\mathrm{sgn}(\Theta_i) - a_{it}|.$$

$$\geq \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{d} \mathbb{1}\{\mathrm{sgn}(\Theta_i) \neq \mathrm{sgn}(a_{it})\}$$

so that $\mathbb{E}[R_T(\Theta)] \geq \frac{1}{\sqrt{T}} \sum_{i=1}^{d} \frac{T}{2} \mathbb{P}(A(\Theta, i))$

$$\geq \frac{\sqrt{T}}{8} d e^{-2} \qquad \square$$