August 13, 2020

Stochastic Linear Bandits

- **Stochastic contextual bandits:**
  - similar (mirror) of prev. lecture
  - key difference,

$$X_t = \underbrace{r(C_t, A_t)}_{\text{reward function}} + \eta_t$$

η_t → sub-Gaussian noise

$$r : \mathcal{C} \times [k] \to \mathbb{R}$$

  - w/o going into ~~sigma~~ σ-fields, idea is that

$$\mathbb{E}[X_t \mid H_{t-1}] = r(C_t, A_t)$$

  - if $r(\cdot, \cdot)$ was known, then,

$$A_t^* \in \underset{a \in [k]}{\operatorname{argmax}}\ r(C_t, a)$$

and so, the expected regret is

$$R_n = \mathbb{E}\left[\sum_{t=1}^{n} \max_a r(C_t, a) - \sum_{t=1}^{n} X_t\right]$$

Note: since no assumption is made on how rewards are chosen, it is possible to pick sub-optimal arms, one way to circumvent is to ensure that $A_1, \ldots A_{t-1}$ don't significantly alter $C_t, \ldots C_n$'s ...

- feature map-based

- assume learner has access to $\psi : C \times [k] \to \mathbb{R}^d$, and then, for some unknown $\theta^* \in \mathbb{R}^d$,

$$r(c,a) = \langle \theta_*, \psi(c,a) \rangle \quad \forall \ (c,a)$$

- ψ is referred to as feature map [o/p of NN before final layer]

- $|r(c,a) - r(c',a')| \leq \|\theta_*\| \, \|\psi(c,a) - \psi(c',a')\|_*$ [Hölder]

so, assumptions on $\theta_* \equiv$ "smoothness" of $r(\cdot,\cdot)$

- Stochastic linear Bandits

- the "action" is not as critical as the feature vector of an action, and thus,

- - let $A_t \subseteq \mathbb{R}^d$ be decision set, $A_t \in \mathcal{A}_t$

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

then,

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

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

"Decision set"

- if $A_t = \{e_1, \ldots, e_d\}$, then stochastic bandit
- if $A_t = \{\psi(\epsilon_t, i) \mid i \in [k]\}$, then contextual linear bandit.

• How to solve?

- Generalization of UCB: pick $C_t \subset \mathbb{R}^d$ that "contains θ_* w.h.p"
- assume, $\mathbb{E}\ \theta_* \in C_t$, then for any $a \in \mathbb{R}^d$, let

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

be an upper bound on mean pay off $\langle \theta^*, a \rangle$;
then select

$$A_t = \operatorname{argmax}_{a \in A_t} UCB_t(a)$$

- but how to construct C_t??
  - need an estimator for θ_* [analogous to μ_i for Stoch]
  - regularized least-squares estimator.

$$\hat{\theta}_t = \operatorname{arg\,min}_{\theta \in \mathbb{R}^d} \left(\sum_{s=1}^{t} (X_s - \langle \theta, A_s \rangle)^2 + \lambda \|\theta\|^2 \right)$$

$$\hat{\theta}_t = V_t^{-1} \sum_{s=1}^{t} A_s X_s \; ; \; V_0 = \lambda I \; ; \; V_t = V_0 + \sum_{s=1}^{t} A_s A_s^T$$

now, just pick

$$C_t \subseteq \mathcal{E}_t = \left\{ \theta \in \mathbb{R}^d \mid \|\theta - \hat{\theta}_{t+1}\|_{V_{t+1}}^2 \leq \beta_t \right\}$$

Remarks : A. $\|x\|_D^2 = x^T D x$, $D \succeq 0$

B $\beta_1 \geq 1$, , , . $\beta_1 \leq \beta_2 \cdots \leq \beta_n$

C: $\lambda_i(V_t)$ are "increasing" $\Rightarrow$ $\mathrm{vol}(\mathcal{E}_t)$ is shrinking

Main Result :

Assumption : We follow wing hold

- $1 \leq \beta_1 \leq \beta_2 \leq \cdots \leq \beta_n$
- $\max_{t \in [n]} \sup_{a,b \in A_t} \langle \theta_*, a-b \rangle \leq 1$
- $\|a\|_2 \leq L \quad \forall \; a \in \bigcup_{t=1}^{n} A_t$
- $\exists\, \delta \in (0,1)$ s.t w.p $1-\delta$, $\forall\, t \in [n]$ $\theta_* \in C_t$

Thm : w.p $1-\delta$, LinUCB satisfies

$$\hat{R}_n \leq \sqrt{8 n \beta_n \log\left(\frac{\det V_n}{\det V_0}\right)} \leq \sqrt{8 d n \beta_n \log\left(\frac{d\lambda + nL^2}{d\lambda}\right)}$$

$$R_n \leq C d \sqrt{n} \log(nL)$$

$$\sqrt{\beta_n} = \sqrt{\lambda}\, \|\theta_*\|_2 + \sqrt{2\log\left(\tfrac{1}{\delta}\right) + d \log\left(\frac{d\lambda + nL^2}{d\lambda}\right)}$$