Contextual Bandits & OCB

- ε-Greedy
- SquareCB (inverse prop. gap weighting)

Proved given **Lemma** $\forall \gamma > 0$

$\forall \hat{f}, f^*: [K] \to [0,1]$

$$\pi \sim p(\pi) = \frac{1}{\lambda + \gamma(\hat{f}(\hat{\pi}) - \hat{f}(\pi))}$$

$$\mathbb{E}_{\pi\sim p}\left[f^*(\pi^*) - f^*(\pi)\right]$$

$$\leq \frac{K}{\gamma} + \gamma\,\mathbb{E}\left[\left(\hat{f}(\pi) - f^*(\pi)\right)^2\right]$$

Pf: $\mathbb{E}_p\left[f^*(\pi^*) - f^*(\pi)\right]$

$$= \mathbb{E}_p\Big[f^*(\pi^*) - \hat{f}(\pi^*)$$

$$+ \hat{f}(\pi^*) - \hat{f}(\hat{\pi})$$

$$+ \cdots\Big]$$

$$\le \frac{K}{\gamma} + \gamma\, \mathbb{E}_p\left[\left(\hat{f}(\pi) - f^*(\pi)\right)^2\right]$$

$$\mathbb{E}\,\text{Regret} \le \frac{KT}{\gamma} + \gamma \underbrace{\sum_{t=1}^{T} \mathbb{E}_p\left[\left(\hat{f}_t(\pi) - f^*(\pi)\right)^2\right]}_{\text{RegretSQ}}$$

(over T rounds)

$$\gamma = \sqrt{\frac{KT}{\text{RegretSQ}}}$$

$$= 2\sqrt{KT\,\text{RegretSQ}}$$

Today: LinUCB

High-level: 1. Use linear regression ($\vec{\theta}_t$) to estimate the reward of each arm

2. Use uncertainty estimates & $\vec{\theta}_t$ to get UCB on each arm

3. "Optimistic" pick arm w/ best UCB.

Algorithm: $\theta^* \in \mathbb{R}^d$, $\beta > 0$ (Assume $\phi(\cdot,\cdot)$ known, $\|\theta^*\|_2 \leq 1$)

$\mathbb{E}[r_t \mid x_t, \pi_t] = \langle \theta^*, \phi(x_t, \pi_t) \rangle$

For $t = 0$ to $T-1$.

- $$\hat{\theta}_t = \underset{\|\theta\|_2 \leq 1}{\operatorname{argmin}} \sum_{s=1}^{t} \big(r_s - \langle \theta, \underbrace{\phi(x_s, \pi_s)}_{\phi_s} \rangle \big)^2$$

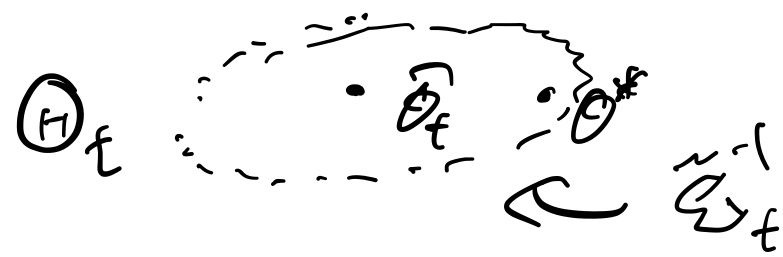
- Define

$$\tilde{\Sigma}_t = \sum_{s=1}^{t} \phi_s \phi_s^T + I$$

(so $\langle u, \tilde{\Sigma}_t u \rangle = \sum_{s=1}^{t} \langle \phi_s, u \rangle^2 + \|u\|_2^2$)

"how much we know about direction u"

- Observe context x_{t+1}.



- Select

$$\pi_{t+1} \in \underset{\pi \in [K]}{\operatorname{argmax}} \left[\max_{\substack{\theta \text{ s.t.} \\ \langle \theta - \hat{\theta}_t, \tilde{\Sigma}_t (\theta - \hat{\theta}_t) \rangle \le (16\beta + 4)^2}} \langle \theta, \phi(x_{t+1}, \pi) \rangle \right]$$

"Uncertainty set" $= \Theta_t$

- Play π_{t+1}, get reward r_{t+1}.

Possible issues:

- Θ_t too small, doesn't include θ^*
- Θ_t too big.

Thm: $\mathbb{E}\,\text{Regret} = \mathbb{E}\sum_{t=1}^{T}\left(f^*(x_t,\pi_t^*) - f^*(x_t,\pi_t)\right)$

$$\lesssim d\sqrt{T}\log T$$

where $\beta = \Theta(d\log T)$.

Pf Has two parts.

Part I: Show (with high probability) θ^* is always in the uncertainty set.

$$\pi^*_{t+1} = \underset{\pi \in [K]}{\arg\max} \langle \theta^*, \phi(x_{t+1}, \pi) \rangle$$

Pf:

$$\mathbb{E}\,\mathrm{Regret} = \sum_{t=0}^{T-1} \Big(\langle \theta^*, \phi(x_{t+1}, \pi^*_{t+1}) \rangle - \langle \theta^*, \phi(x_{t+1}, \pi_{t+1}) \rangle \Big)$$

Step 1 (HW) $\theta^* \in \Theta_t$

$$\leq \sum_{t=0}^{T-1} \Big(\max_{\theta \in \Theta_t} \langle \theta, \phi(x_{t+1}, \pi^*_{t+1}) \rangle - \langle \theta^*, \phi(x_{t+1}, \pi_{t+1}) \rangle \Big)$$

$$\leq \sum_{t=0}^{T-1} \Big(\max_{\theta \in \Theta_t} \langle \theta, \phi(x_{t+1}, \pi_{t+1}) \rangle - \langle \theta^*, \phi(x_{t+1}, \pi_{t+1}) \rangle \Big)$$

$$\max \langle \theta - \theta^*, \phi(x_{t+1}, \pi_{t+1}) \rangle$$

$$\max_{\theta} \langle \theta - \theta^*, \phi(x_t, \pi_t) \rangle$$

$$= \max_{\theta} \left\langle \tilde{\Sigma}_t^{1/2} (\theta - \theta^*), \tilde{\Sigma}_t^{-1/2} \phi(x_t, \pi_t) \right\rangle$$

<u>Lemma</u> (Elliptical Potential Lemma)

Suppose $\phi_1 \ldots \phi_T \in \mathbb{R}^d$ $\|\phi_i\| \leq 1$

$$\tilde{\Sigma}_t = \sum_{s=1}^{t} \phi_s \phi_s^T + I$$

Then

$$\sum_{t=1}^{T} \underbrace{\phi_t \tilde{\Sigma}_{t-1}^{-1} \phi_t}_{\text{"how surprising } \phi_t \text{ is"}} \leq 2d \log T.$$