CB continued

Last time:

Explaining method called SquareCB

"Reduce"

Inverse Propensity Gap Weighting

Contextual Bandits

Online regression (we have studied)

ϵ-Greedy Strategy for CB.

$\epsilon \in [0,1]$

For $t = 1$ to T:

- Nature gives a context x_t.
- Use "oracle" to forecast rewards for all arms.

$$\hat{f}_t : [K] \to [0,1] \quad \text{estimate (based on } x_t \text{ \& past)}$$

$$f^*(x_t, a_m) = \mathbb{E}[r_t \mid x_t, a_t] \quad \underline{\text{UNKNOWN}}$$

$\pi_t \sim p_t$

- With probability ϵ, pick π_t or Unif$[K]$
- $1-\epsilon$, pick $\pi_t = \underset{\pi}{\operatorname{argmax}}\, \hat{f}_t(\pi)$.

- Observe r_t of our action.

Thm: given an oracle achieving

$$\sum_{t=1}^{T} \mathbb{E}_{\pi_t}\left[\left(\hat{f}_t(x_t, \pi_t) - f^*(x_t, \pi_t)\right)^2\right] \le \text{Reg}_{SQ}$$

(typically $o(T)$)

Then ε-Greedy achieves (for some ε)

$$\mathbb{E}[\text{Reg}_{CB}] \le K^{1/3} T^{2/3} \text{Reg}_{SQ}^{1/3}$$

$$\sum_{t=1}^{T}\left[f^*(x_t, \pi^*_t) - f^*(x_t, \pi_t)\right]$$

where $\pi^*_t = \operatorname{argmax}_{\pi} f^*(x_t, \pi)$

Pf:

$$\mathbb{E}\sum_{t=1}^{T}\left[f^*(x_t,\pi^*_t)-f^*(x_t,\pi_t)\right]$$

$$=\mathbb{E}\sum_{t=1}^{T}\Big[\underbrace{\left(\hat{f}(x_t,\pi_t^*)-\hat{f}_t(x_t,\pi_t)\right)}_{\text{I}}+\underbrace{\left(f^*(x_t,\pi_t^*)-\hat{f}_t(x_t,\pi_t^*)\right)}_{\text{II}}+\underbrace{\left(\hat{f}_t(x_t,\pi_t)-f^*(x_t,\pi_t)\right)}_{\text{III}}\Big]$$

Observe: $\text{I}=\sum\left(\hat{f}_t(\pi_t^*)-\hat{f}_t(\pi_t)\right)$

$\leq \varepsilon T\cdot 1=\varepsilon T$

$|\langle u, v \rangle| \leq |u|_2 |v|_2$ C.S.

$$\mathbb{E}\left[\sum_{t=1}^{T} \left(\hat{f}_t(\pi_t) - f^*(\pi_t) \right)\right]$$

$$\leq \sqrt{T}\left[\sum_{t=1}^{T} \mathbb{E}\left[\left(\hat{f}_t(\pi_t) - f^*(\pi_t) \right)^2\right]\right]$$

$$\leq \sqrt{T}\,\sqrt{\text{RegSQ}}$$

(II)

$$\sum_{t=1}^{T} \left(f^*(x_t, \pi_t^*) - \hat{f}_t(\pi_t^*) \right)$$

$$= \sum_{t=1}^{T} \frac{\sqrt{p_t(\pi_t^*)}}{\sqrt{p_t(\pi_t^*)}} \left(f^*(x_t, \pi_t^*) - \hat{f}_t(\pi_t^*) \right)$$

$p_t(\pi^*) \geq \frac{\varepsilon}{K}$

$$\leq \underbrace{\sqrt{\sum_{t=1}^{T} \frac{1}{p_t(\pi_t^*)}}}_{\leq \sqrt{\frac{KT}{\varepsilon}}} \underbrace{\sqrt{\sum_{t=1}^{T} p_t(\pi_t^*) \left(f^*(x_t, \pi_t^*) - \hat{f}_t(\pi_t^*) \right)^2}}_{\leq \sqrt{\text{Reg}_{\text{SQ}}}}$$

$$\text{I} + \text{II} + \text{III} \leq \varepsilon T + \sqrt{\frac{KT}{\varepsilon}}\sqrt{R_{eg_{sq}}} \left(+ \sqrt{T}\sqrt{R_{eg_{sq}}} \right)$$

Let $\varepsilon = \frac{K^{1/3} R_{eg_{sq}}^{1/3}}{T^{1/3}}$

$$\hookrightarrow \leq 3 K^{1/3} R_{eg_{sq}}^{1/3} T^{2/3}$$

Alg SquareCB:

$\gamma > 0$ (parameter to be optimized later)

For $t = 1$ to T:

- Nature reveals x_t
- Oracle forecast is $\hat{f}_t : [K] \to [0,1]$
- Select $\lambda_t \in [1, K]$ s.t.

$$\sum_{\pi \in [K]} p_t(\pi) = 1$$

$$p_t(\pi) = \frac{1}{\lambda_t + \gamma\left(\hat{f}_t(\hat{\pi}_t) - \hat{f}_t(\pi)\right)}$$

$\hat{\pi}_t = \arg\max_{\pi} \hat{F}(\pi)$

- Play $\pi_t \sim p_t$, get r_t.

Thm As before, assume oracle

$$\mathbb{E}\left(\hat{f}_t(\pi_t) - f^*(x_t, \pi_t)\right)^2 \le Reg_{SQ}.$$

For some $\delta > 0$,

$$\mathbb{E}\, Reg_{CB} \le \sqrt{KT\, Reg_{SQ}}$$

pf? Use

$$|ab| \le \frac{a^2}{2} + \frac{b^2}{2} \quad \longleftrightarrow \quad (a-b)^2 \ge 0$$

(AM-GM Inequality)

in a clever way.

Lemma: $\hat{f} : [K] \to [0,1]$

$f^* : [K] \to [0,1]$

$\lambda \in [1, A]$

$\hat{\pi} = \underset{\pi}{\operatorname{argmax}}\ \hat{f}(\pi)$

$\pi^* = \underset{\pi}{\operatorname{argmax}}\ f^*(\pi)$

$$p(\pi) \simeq \frac{1}{\lambda + 2\gamma\left(\hat{f}(\hat{\pi}) - \hat{f}(\pi)\right)}$$

$\pi \sim p$

Then:

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

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

Pf Thm:

By Lemma

$$\mathbb{E}\,\mathrm{Reg}_{CB} = \mathbb{E}\sum_{t=1}^{T}\left(f^*(x_t,\pi^*_t) - f^*(x_t,\pi_t)\right)$$

$$\leq \sum_{t=1}^{T}\frac{K}{\gamma} + \gamma\sum_{t=1}^{T}\mathbb{E}\left(f^*(x_t,\pi^*_t) - f^*(x_t,\pi_t)\right)^2$$

$$= \frac{KT}{\gamma} + \gamma\,\mathrm{Reg}_{SQ} \qquad \text{let } \gamma = \sqrt{\frac{KT}{\mathrm{Reg}_{SQ}}}$$

$$\leq 2\sqrt{KT\,\mathrm{Reg}_{SQ}} \qquad \square$$

Pf of Lemma:

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

$$= \mathop{\mathbb{E}}_{p}\Big[\underbrace{\left(f^*(\pi^*) - \hat{f}(\pi^*)\right)}_{\text{I}}$$

$$+ \underbrace{\left(\hat{f}(\pi^*) - \hat{f}(\hat{\pi})\right)}_{\text{II}}$$

$$+ \underbrace{\left(\hat{f}(\hat{\pi}) - \hat{f}(\pi)\right)}_{\text{III}}$$

$$+ \underbrace{\left(\hat{f}(\pi) - f^*(\pi)\right)}_{\text{IV}}\Big]$$

(III) & (IV) easy to bound. Interesting part is (I) & (II).

(I)

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

$$= \frac{\gamma p(\pi^*)}{\gamma p(\pi^*)} \left(f^*(\pi^*) - \hat{f}(\pi^*) \right)$$

$$\leq \frac{\gamma p(\pi^*)}{2} \left(f^*(\pi^*) - \hat{f}(\pi^*) \right)^2 + \frac{1}{2\gamma p(\pi^*)}$$

$$\leq \frac{\gamma \mathbb{E}_p\left[\left(f^*(\pi) - \hat{f}(\pi) \right)^2 \right]}{2} + \frac{1}{2\gamma p(\pi^*)}$$

$$\text{I} + \text{II} \leq \frac{\gamma}{2} \mathbb{E}_{\pi \sim p}\left[\left(f^*(\pi) - \hat{f}(\pi) \right)^2 \right] + \frac{1}{2\gamma p(\pi^*)} + \left(\hat{f}(\pi^*) - \hat{f}(\hat{\pi}) \right)$$

Recall: $\frac{1}{2\gamma p(\pi^*)} = \frac{\lambda + 2\gamma\left(\hat{f}(\hat{\pi}) - \hat{f}(\pi^*)\right)}{2\gamma}$

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