بسم الله الرحمن الرحیم

یادگیری برخط

جلسه هفدهم: بندیت خطی

بندیت خطی

- انتخاب بردارهای $y_t \in \mathbb{R}^d$
- در زمان t:
  - انتخاب یک $A_t \in \mathscr{A}$
  - زیان = $\langle A_t, y_t \rangle$

بندیت خطی

- انتخاب بردارهای $y_t \in \mathbb{R}^d$
- در زمان t:
  - انتخاب یک $A_t \in \mathscr{A}$
  - زیان = $\langle A_t, y_t \rangle$

بندیت دشمنانه

بندیت خطی

- انتخاب بردارهای $y_t \in \mathbb{R}^d$
- در زمان t:
  - انتخاب یک $A_t \in \mathscr{A}$
  - زیان = $\langle A_t, y_t \rangle$

بندیت دشمنانه

- ماتریس $y_t \in [0,1]^k$
- در مرحله t:
  - انتخاب عمل $A_t \in [k]$
  - دریافت نتیجه y_{t,A_t}

بندیت خطی

- انتخاب بردارهای $y_t \in \mathbb{R}^d$
- در زمان t:
  - انتخاب یک $A_t \in \mathscr{A}$
  - زیان = $\langle A_t, y_t \rangle$

بندیت دشمنانه

حالت خاص!

- ماتریس $y_t \in [0,1]^k$
- در مرحله t:
  - انتخاب عمل $A_t \in [k]$
  - دریافت نتیجه y_{t,A_t}

بندیت خطی

- انتخاب بردارهای $y_t \in \mathbb{R}^d$
- در زمان t:
  - انتخاب یک $A_t \in \mathcal{A}$
  - زیان = $\langle A_t, y_t \rangle$

فرض:

$$y_t \in \mathcal{L} = \{x \in \mathbb{R}^d : \sup_{a \in \mathcal{A}} |\langle a, x \rangle| \leq 1\}$$

بندیت دشمنانه

حالت خاص!

- ماتریس $y_t \in [0,1]^k$
- در مرحله t:
  - انتخاب عمل $A_t \in [k]$
  - دریافت نتیجه y_{t,A_t}

بندیت خطی

- انتخاب بردارهای $y_t \in \mathbb{R}^d$
- در زمان t:
  - انتخاب یک $A_t \in \mathscr{A}$
  - زیان = $\langle A_t, y_t \rangle$

کاهش گرادیان بندیتی؟
$R \leq O(n^{2/3} D^{3/2})$

فرض:

$$y_t \in \mathcal{L} = \left\{ x \in \mathbb{R}^d : \sup_{a \in \mathcal{A}} |\langle a, x \rangle| \leq 1 \right\}$$

الگوریتم

- در مرحله t:

  - توزیع P_t روی $\mathscr{A}$
  - انتخاب عمل $A_t \sim P_t$
  - دریافت نتیجه $\langle A_t, y_t \rangle$

الگوریتم

- در مرحله t:
  - توزیع P_t روی $\mathscr{A}$
  - انتخاب عمل $A_t \sim P_t$
  - دریافت نتیجه $\langle A_t, y_t \rangle$

نمایی

الگوریتم

الگوریتم

- ایده اول: احتمال متناسب با تابع نمایی

$$\tilde{P}_t(a) \propto \exp\left(-\eta \sum_{s=1}^{t-1} \hat{Y}_s(a)\right)$$

الگوریتم

- ایده اول: احتمال متناسب با تابع نمایی

$$\tilde{P}_t(a) \propto \exp\left(-\eta \sum_{s=1}^{t-1} \hat{Y}_s(a)\right)$$

- ایده دوم: + کمی گشت و گذار

$$P_t(a) = (1-\gamma)\tilde{P}_t(a) + \gamma\pi(a)$$

توزیع احتمال روی $\mathscr{A}$

الگوریتم

- در مرحله t:
  - توزیع P_t روی $\mathscr{A}$
  - انتخاب عمل $A_t \sim P_t$
  - دریافت نتیجه $Y_t := \langle A_t, y_t \rangle$

$$\tilde{P}_t(a) \propto \exp\left(-\eta \sum_{s=1}^{t-1} \hat{Y}_s(a)\right)$$

$$P_t(a) = (1-\gamma)\tilde{P}_t(a) + \gamma \pi(a)$$

الگوریتم

- در مرحله t:
  - توزیع P_t روی $\mathcal{A}$
  - انتخاب عمل $A_t \sim P_t$
  - دریافت نتیجه $Y_t := \langle A_t, y_t \rangle$

$$\tilde{P}_t(a) \propto \exp\left(-\eta \sum_{s=1}^{t-1} \hat{Y}_s(a)\right)$$

$$P_t(a) = (1-\gamma)\tilde{P}_t(a) + \gamma \pi(a)$$

تخمین y_t

$$Y_t = \langle A_t, y_t \rangle$$

دریافت از محیط

$$Y_t = \langle A_t, y_t \rangle$$

دریافت از محیط

$$\hat{Y}_t = R_t \dot{A}_t Y_t$$

فرض

$$Y_t = \langle A_t, y_t \rangle$$

دریافت از محیط

$$\hat{Y}_t = R_t \dot{A}_t Y_t$$

فرض

$$\mathbb{E}_t[\hat{Y}_t]$$

$$Y_t = \langle A_t, y_t \rangle$$

دریافت از محیط

$$\hat{Y}_t = R_t A_t Y_t$$

فرض

$$\mathbb{E}_t[\hat{Y}_t] = R_t \mathbb{E}_t[A_t A_t^\top] y_t$$

$$Y_t = \langle A_t, y_t \rangle$$

دریافت از محیط

$$\hat{Y}_t = R_t A_t Y_t$$

فرض

$$\mathbb{E}_t[\hat{Y}_t] = R_t \mathbb{E}_t[A_t A_t^\top] y_t = R_t \left(\sum_{a \in \mathcal{A}} P_t(a) a a^\top \right) y_t$$

$$Y_t = \langle A_t, y_t \rangle$$

دریافت از محیط

$$\hat{Y}_t = R_t A_t Y_t$$

فرض

$$\mathbb{E}_t[\hat{Y}_t] = R_t \mathbb{E}_t[A_t A_t^\top] y_t = R_t \underbrace{\left(\sum_{a \in \mathcal{A}} P_t(a) a a^\top \right)}_{Q_t} y_t$$

$$Y_t = \langle A_t, y_t \rangle$$

دریافت از محیط

$$\hat{Y}_t = R_t A_t Y_t$$

فرض

$$\mathbb{E}_t[\hat{Y}_t] = R_t \mathbb{E}_t[A_t A_t^\top] y_t = R_t \underbrace{\left(\sum_{a \in \mathcal{A}} P_t(a) a a^\top \right)}_{Q_t} y_t$$

اگر بگذاریم

$$R_t = Q_t^{-1}$$

$$Y_t = \langle A_t, y_t \rangle$$

دریافت از محیط

$$\hat{Y}_t = R_t A_t Y_t$$

فرض

$$\mathbb{E}_t[\hat{Y}_t] = R_t \mathbb{E}_t[A_t A_t^\top] y_t = R_t \underbrace{\left(\sum_{a \in \mathcal{A}} P_t(a) a a^\top \right)}_{Q_t} y_t$$

اگر بگذاریم

$$R_t = Q_t^{-1}$$

وارون‌پذیر

$$Y_t = \langle A_t, y_t \rangle$$

دریافت از محیط

$$\hat{Y}_t = R_t A_t Y_t$$

فرض

$$\mathbb{E}_t[\hat{Y}_t] = R_t \mathbb{E}_t[A_t A_t^\top] y_t = R_t \underbrace{\left(\sum_{a \in \mathcal{A}} P_t(a) a a^\top \right)}_{Q_t} y_t$$

اگر بگذاریم

$$R_t = Q_t^{-1}$$

وارون‌پذیر

$$Q(\pi) = \sum_{a \in \mathcal{A}} \pi(a) a a^\top$$

$$Y_t = \langle A_t, y_t \rangle$$

دریافت از محیط

$$\hat{Y}_t = R_t A_t Y_t$$

فرض

$$\mathbb{E}_t[\hat{Y}_t] = R_t \mathbb{E}_t[A_t A_t^\top] y_t = R_t \underbrace{\left(\sum_{a \in \mathcal{A}} P_t(a) a a^\top \right)}_{Q_t} y_t$$

اگر بگذاریم

$$R_t = Q_t^{-1}$$

وارون‌پذیر

$$Q(\pi) = \sum_{a \in \mathcal{A}} \pi(a) a a^\top$$

$$\mathbb{E}_t[\hat{Y}_t] = y_t$$

$$P_t(a) = (1-\gamma)\tilde{P}_t(a) + \gamma\pi(a)$$

$$\underbrace{R_t\left(\sum_{a\in\mathcal{A}} P_t(a)aa^\top\right)y_t}_{Q_t}$$

وارون‌پذیر

$$Q(\pi) = \sum_{a\in\mathcal{A}} \pi(a)aa^\top$$

وارون‌پذیر =>

Q_t وارون‌پذیر

الگوریتم EXP3 برای بندیت خطی

Input Finite action set $\mathcal{A} \subset \mathbb{R}^d$, learning rate η, exploration distribution π, exploration parameter γ

for $t = 1, 2, \ldots, n$ **do**

  Compute sampling distribution:

$$P_t(a) = \gamma \pi(a) + (1-\gamma) \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \hat{Y}_s(a)\right)}{\sum_{a' \in \mathcal{A}} \exp\left(-\eta \sum_{s=1}^{t-1} \hat{Y}_s(a')\right)}.$$

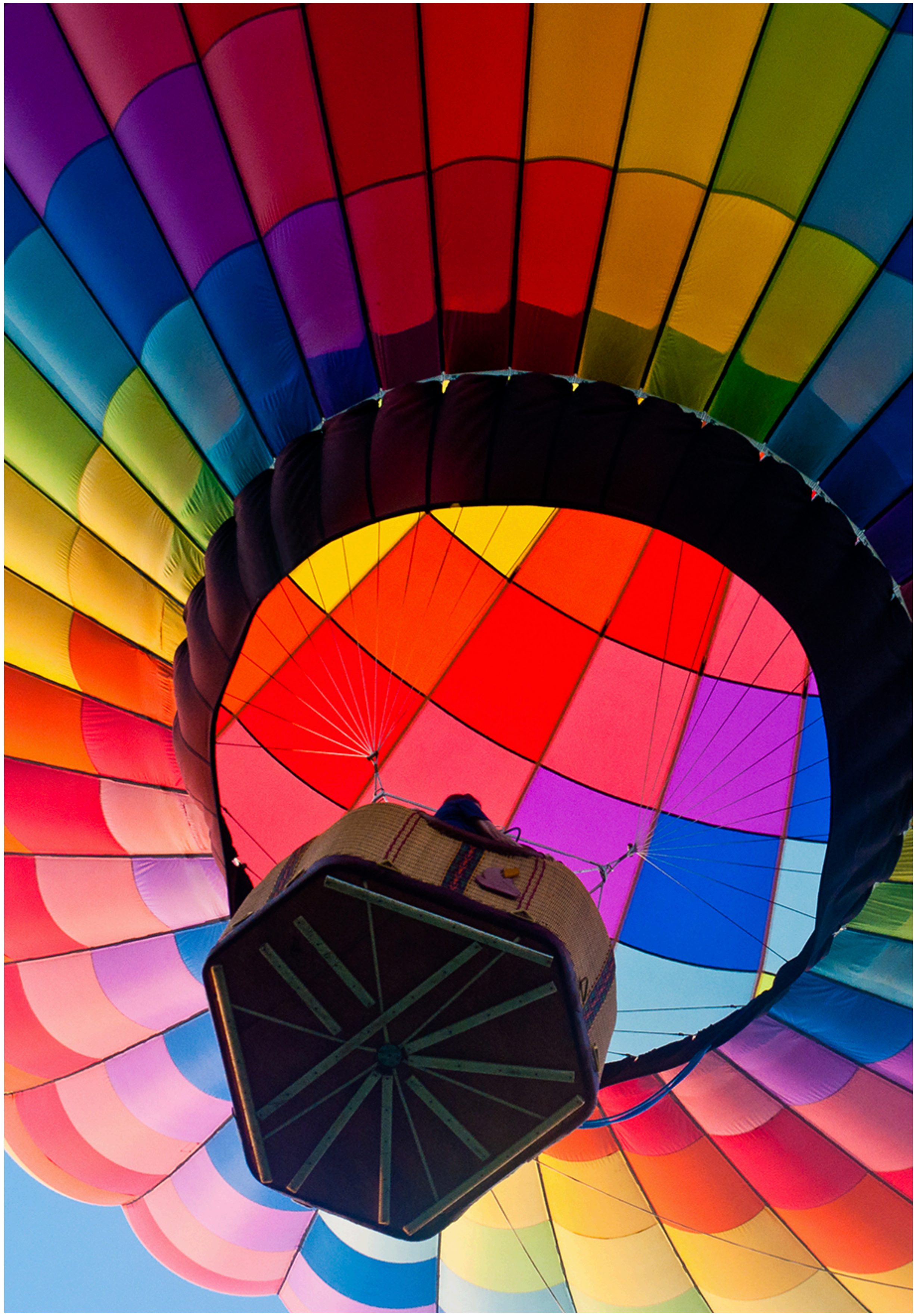
  Sample action $A_t \sim P_t$

  Observe loss $Y_t = \langle A_t, y_t \rangle$ and compute loss estimates:

$$\hat{Y}_t = Q_t^{-1} A_t Y_t \qquad \text{and} \qquad \hat{Y}_t(a) = \langle a, \hat{Y}_t \rangle.$$

end for

تحلیل الگوریتم



الگوریتم EXP3 برای بندیت خطی

$$P_t(a) = (1-\gamma)\tilde{P}_t(a) + \gamma\pi(a)$$

$$R_n \le \frac{\log k}{\eta} + \boxed{2\gamma n} + \eta \sum_{t=1}^{n} \mathbb{E}\left[\sum_{a\in\mathcal{A}} P_t(a)\hat{Y}_t^2(a)\right]$$

الگوریتم EXP3 برای بندیت خطی

$$P_t(a) = (1-\gamma)\tilde{P}_t(a) + \gamma\pi(a)$$

$$R_n \leq \frac{\log k}{\eta} + \boxed{2\gamma n} + \eta \sum_{t=1}^{n} \mathbb{E}\left[\sum_{a \in \mathcal{A}} P_t(a)\hat{Y}_t^2(a)\right]$$

کرانی برای این قسمت

$$\sum_{t=1}^{n} \mathbb{E}\left[\sum_{a \in \mathcal{A}} P_t(a) \hat{Y}_t^2(a)\right]$$

$$\sum_{t=1}^{n} \mathbb{E}\left[\sum_{a \in \mathcal{A}} P_t(a) \hat{Y}_t^2(a)\right]$$

$$\sum_{t=1}^{n} \mathbb{E}\left[\sum_{a \in \mathcal{A}} P_t(a) \hat{Y}_t^2(a)\right]$$

$\hat{Y}_t^2(a)$

$$\sum_{t=1}^{n} \mathbb{E}\left[\sum_{a \in \mathcal{A}} P_t(a) \hat{Y}_t^2(a)\right]$$

$$\hat{Y}_t^2(a) = (a^\top Q_t^{-1} A_t Y_t)^2$$

$$\sum_{t=1}^{n} \mathbb{E}\left[\sum_{a\in\mathcal{A}} P_t(a)\hat{Y}_t^2(a)\right]$$

$$\hat{Y}_t^2(a) = (a^\top Q_t^{-1} A_t Y_t)^2 = Y_t^2 A_t^\top Q_t^{-1} a a^\top Q_t^{-1} A_t$$

$$\sum_{t=1}^{n} \mathbb{E}\left[\sum_{a \in \mathcal{A}} P_t(a) \hat{Y}_t^2(a)\right]$$

$$\hat{Y}_t^2(a) = (a^\top Q_t^{-1} A_t Y_t)^2 = Y_t^2 A_t^\top Q_t^{-1} a a^\top Q_t^{-1} A_t$$

$$\sum_{t=1}^{n} \mathbb{E}\left[\sum_{a\in\mathcal{A}} P_t(a)\hat{Y}_t^2(a)\right]$$

$$\hat{Y}_t^2(a) = (a^\top Q_t^{-1} A_t Y_t)^2 = Y_t^2 A_t^\top Q_t^{-1} a a^\top Q_t^{-1} A_t$$

$\sum_a P_t(a)\hat{Y}_t^2(a)$

$$\sum_{t=1}^{n} \mathbb{E}\left[\sum_{a\in\mathcal{A}} P_t(a)\hat{Y}_t^2(a)\right]$$

$$\hat{Y}_t^2(a) = (a^\top Q_t^{-1} A_t Y_t)^2 = Y_t^2 A_t^\top Q_t^{-1} a a^\top Q_t^{-1} A_t$$

$$\sum_a P_t(a)\hat{Y}_t^2(a) = Y_t^2 A_t^\top Q_t^{-1} A_t$$

$$\sum_{t=1}^{n} \mathbb{E}\left[\sum_{a\in\mathcal{A}} P_t(a)\hat{Y}_t^2(a)\right]$$

$$\hat{Y}_t^2(a) = (a^\top Q_t^{-1} A_t Y_t)^2 = Y_t^2 A_t^\top Q_t^{-1} a a^\top Q_t^{-1} A_t$$

$$\sum_a P_t(a)\hat{Y}_t^2(a) = Y_t^2 A_t^\top Q_t^{-1} A_t \leq A_t^\top Q_t^{-1} A_t$$

$$\sum_{t=1}^{n} \mathbb{E}\left[\sum_{a\in\mathcal{A}} P_t(a)\hat{Y}_t^2(a)\right]$$

$$\hat{Y}_t^2(a) = (a^\top Q_t^{-1} A_t Y_t)^2 = Y_t^2 A_t^\top Q_t^{-1} a a^\top Q_t^{-1} A_t$$

$$\sum_a P_t(a)\hat{Y}_t^2(a) = Y_t^2 A_t^\top Q_t^{-1} A_t \leq A_t^\top Q_t^{-1} A_t = \text{trace}(A_t A_t^\top Q_t^{-1})$$

$$\sum_{t=1}^{n} \mathbb{E}\left[\sum_{a \in \mathcal{A}} P_t(a) \hat{Y}_t^2(a)\right]$$

$$\hat{Y}_t^2(a) = (a^\top Q_t^{-1} A_t Y_t)^2 = Y_t^2 A_t^\top Q_t^{-1} a a^\top Q_t^{-1} A_t$$

$$\sum_a P_t(a) \hat{Y}_t^2(a) = Y_t^2 A_t^\top Q_t^{-1} A_t \leq A_t^\top Q_t^{-1} A_t = \operatorname{trace}(A_t A_t^\top Q_t^{-1})$$

$$\sum_{t=1}^{n} \mathbb{E}\left[\sum_{a\in\mathcal{A}} P_t(a)\hat{Y}_t^2(a)\right]$$

$$\hat{Y}_t^2(a) = (a^\top Q_t^{-1} A_t Y_t)^2 = Y_t^2 A_t^\top Q_t^{-1} a a^\top Q_t^{-1} A_t$$

$$\sum_a P_t(a)\hat{Y}_t^2(a) = Y_t^2 A_t^\top Q_t^{-1} A_t \leq A_t^\top Q_t^{-1} A_t = \operatorname{trace}(A_t A_t^\top Q_t^{-1})$$

$$\mathbb{E}_t\left[\sum_{a\in\mathcal{A}} P_t(a)\hat{Y}_t^2(a)\right] \leq \operatorname{trace}\left(\sum_{a\in\mathcal{A}} P_t(a) a a^\top Q_t^{-1}\right)$$

$$\sum_{t=1}^{n} \mathbb{E}\left[\sum_{a\in\mathcal{A}} P_t(a)\hat{Y}_t^2(a)\right]$$

$$\hat{Y}_t^2(a) = (a^\top Q_t^{-1} A_t Y_t)^2 = Y_t^2 A_t^\top Q_t^{-1} a a^\top Q_t^{-1} A_t$$

$$\textstyle\sum_a P_t(a)\hat{Y}_t^2(a) = Y_t^2 A_t^\top Q_t^{-1} A_t \leq A_t^\top Q_t^{-1} A_t = \operatorname{trace}(A_t A_t^\top Q_t^{-1})$$

$$\mathbb{E}_t\left[\sum_{a\in\mathcal{A}} P_t(a)\hat{Y}_t^2(a)\right] \leq \operatorname{trace}\left(\sum_{a\in\mathcal{A}} P_t(a) a a^\top Q_t^{-1}\right) = d.$$

$$\sum_{t=1}^{n} \mathbb{E}\left[\sum_{a\in\mathcal{A}} P_t(a)\hat{Y}_t^2(a)\right] \le nd$$

$$\hat{Y}_t^2(a) = (a^\top Q_t^{-1} A_t Y_t)^2 = Y_t^2 A_t^\top Q_t^{-1} a a^\top Q_t^{-1} A_t$$

$$\textstyle\sum_a P_t(a)\hat{Y}_t^2(a) = Y_t^2 A_t^\top Q_t^{-1} A_t \le A_t^\top Q_t^{-1} A_t = \mathrm{trace}(A_t A_t^\top Q_t^{-1})$$

$$\mathbb{E}_t\left[\sum_{a\in\mathcal{A}} P_t(a)\hat{Y}_t^2(a)\right] \le \mathrm{trace}\left(\sum_{a\in\mathcal{A}} P_t(a) a a^\top Q_t^{-1}\right) = d.$$

$\leq nd$

$$R_n \leq \frac{\log k}{\eta} + 2\gamma n + \eta \sum_{t=1}^{n} \mathbb{E}\left[\sum_{a \in \mathcal{A}} P_t(a) \hat{Y}_t^2(a)\right]$$

$\leq nd$

$$R_n \leq \frac{\log k}{\eta} + 2\gamma n + \eta \sum_{t=1}^{n} \mathbb{E}\left[\sum_{a \in \mathcal{A}} P_t(a) \hat{Y}_t^2(a)\right]$$

$g(\pi) = \max_{a \in \mathcal{A}} \|a\|^2_{Q^{-1}(\pi)}$,

$$\gamma = \eta g(\pi) \quad \Rightarrow \quad |\eta \hat{Y}_t(a)| \leq 1$$

$\leq nd$

$$R_n \leq \frac{\log k}{\eta} + 2\gamma n + \eta \sum_{t=1}^{n} \mathbb{E}\left[\sum_{a\in\mathcal{A}} P_t(a)\hat{Y}_t^2(a)\right]$$

$g(\pi) = \max_{a\in\mathcal{A}} \|a\|^2_{Q^{-1}(\pi)}$,

$$\gamma = \eta g(\pi) \Rightarrow |\eta \hat{Y}_t(a)| \leq 1$$

$$R_n \leq \frac{\log k}{\eta} + \eta n(2g(\pi) + d)$$

$\leq nd$

$$R_n \leq \frac{\log k}{\eta} + 2\gamma n + \eta \sum_{t=1}^{n} \mathbb{E}\left[\sum_{a \in \mathcal{A}} P_t(a) \hat{Y}_t^2(a)\right]$$

$g(\pi) = \max_{a \in \mathcal{A}} \|a\|^2_{Q^{-1}(\pi)}$,

$$\gamma = \eta g(\pi)$$

$|\eta \hat{Y}_t(a)| \leq 1$

$$R_n \leq \frac{\log k}{\eta} + \eta n (2g(\pi) + d)$$

$$= 2\sqrt{(2g(\pi) + d) n \log(k)}$$

$\eta = \sqrt{\frac{\log(k)}{(2g(\pi)+d)n}}$

$\leq nd$

$$R_n \leq \frac{\log k}{\eta} + 2\gamma n + \eta \sum_{t=1}^{n} \mathbb{E}\left[\sum_{a \in \mathcal{A}} P_t(a) \hat{Y}_t^2(a)\right]$$

$g(\pi) = \max_{a \in \mathcal{A}} \|a\|^2_{Q^{-1}(\pi)}$,

$$\gamma = \eta g(\pi)$$

$|\eta \hat{Y}_t(a)| \leq 1$

$$R_n \leq \frac{\log k}{\eta} + \eta n(2g(\pi) + d)$$

$$= 2\sqrt{(2g(\pi) + d)n \log(k)}$$

$\eta = \sqrt{\frac{\log(k)}{(2g(\pi)+d)n}}$

$$\leq 2\sqrt{3dn \log(k)}.$$

$g(\pi) \leq d.$

THEOREM 27.1. *Assume that $\mathcal{A}$ is non-empty and let $k = |\mathcal{A}|$. For any exploration distribution π, for some parameters η and γ, for all $(y_t)_t$ with $y_t \in \mathcal{L}$, the regret of Algorithm 15 satisfies*

$$R_n \leq 2\sqrt{(2g(\pi) + d)n \log(k)}, \tag{27.1}$$

where $g(\pi) = \max_{a \in \mathcal{A}} \|a\|^2_{Q^{-1}(\pi)}$. Furthermore, there exists an exploration distribution π and parameters η and γ such that $g(\pi) \leq d$, and hence $R_n \leq 2\sqrt{3dn \log(k)}$.

پیوسته

پیوسته؟

THEOREM 27.1. *Assume that $\mathcal{A}$ is non-empty and let $k = |\mathcal{A}|$. For any exploration distribution π, for some parameters η and γ, for all $(y_t)_t$ with $y_t \in \mathcal{L}$, the regret of Algorithm 15 satisfies*

$$R_n \leq 2\sqrt{(2g(\pi) + d)n \log(k)}, \tag{27.1}$$

where $g(\pi) = \max_{a \in \mathcal{A}} \|a\|^2_{Q^{-1}(\pi)}$. Furthermore, there exists an exploration distribution π and parameters η and γ such that $g(\pi) \leq d$, and hence $R_n \leq 2\sqrt{3dn \log(k)}$.

روش اول:

- تقریب $\mathscr{A}$ با $C \subseteq \mathbb{R}^d$

$$\sup_{a \in \mathcal{A}} \min_{b \in \mathcal{C}} \sup_{y \in \mathcal{L}} |\langle a - b, y \rangle| \leq 1/n.$$

روش اول:

- تقریب $C \subseteq \mathbb{R}^d$ با $\mathscr{A}$

$$\sup_{a\in\mathcal{A}} \min_{b\in\mathcal{C}} \sup_{y\in\mathcal{L}} |\langle a-b, y\rangle| \leq 1/n.$$

همیشه با $(6dn)^d$ تا بردار می‌توان

روش اول:

- تقریب $\mathscr{A}$ با $C \subseteq \mathbb{R}^d$

$$\sup_{a\in\mathcal{A}} \min_{b\in\mathcal{C}} \sup_{y\in\mathcal{L}} |\langle a-b, y\rangle| \leq 1/n.$$

همیشه با $(6dn)^d$ تا بردار می‌توان

$$R_n \leq 2\sqrt{3dn\log(k)}.$$

$$R_n = O(d\sqrt{n\log(nd)}).$$

روش اول:

- تقریب $\mathscr{A}$ با $C \subseteq \mathbb{R}^d$

$$\sup_{a \in \mathcal{A}} \min_{b \in \mathcal{C}} \sup_{y \in \mathcal{L}} |\langle a - b, y \rangle| \leq 1/n.$$

همیشه با $(6dn)^d$ تا بردار می‌توان

$$R_n \leq 2\sqrt{3dn\log(k)}.$$

مشکل:
زمان اجرا

$$R_n = O(d\sqrt{n\log(nd)}).$$

روش دوم، EXP3 پیوسته

یک توزیع روی $\mathscr{A}$

$$P_t = (1-\gamma)\tilde{P}_t + \gamma\pi,$$

$$\tilde{P}_t(B) = \frac{\int_B \exp\left(-\eta \sum_{s=1}^{t-1} \hat{Y}_s(a)\right) da}{\int_{\mathcal{A}} \exp\left(-\eta \sum_{s=1}^{t-1} \hat{Y}_s(a)\right) da}.$$

THEOREM 27.3. *Assume that $\mathcal{A}$ is compact, convex and has volume* $\mathrm{vol}(\mathcal{A}) = \int_{\mathcal{A}} da > 0$. *Then an appropriately tuned instantiation of the continuous exponential weights algorithm with Kiefer–Wolfowitz exploration has regret bounded by*

$$R_n \leq 2d\sqrt{3n(1 + \log_+(2n/d))}.$$

الگوریتم

- در مرحله t:
  - توزیع P_t روی $\mathscr{A}$
  - انتخاب عمل $A_t \sim P_t$
  - دریافت نتیجه $Y_t := \langle A_t, y_t \rangle$

یک توزیع روی $\mathscr{A}$

$$P_t = (1-\gamma)\tilde{P}_t + \gamma\pi,$$

$$\tilde{P}_t(B) = \frac{\int_B \exp\left(-\eta \sum_{s=1}^{t-1} \hat{Y}_s(a)\right) da}{\int_{\mathcal{A}} \exp\left(-\eta \sum_{s=1}^{t-1} \hat{Y}_s(a)\right) da}.$$

تخمین y_t

$$\underbrace{\left(\sum_{a\in\mathcal{A}} P_t(a) a a^\top\right)}_{Q_t}$$

$$\hat{Y}_t = Q_t^{-1} A_t Y_t$$

نمونه گرفتن از Pt؟

THEOREM 27.2. *Let* $p(a) \propto \mathbb{I}_{\mathcal{A}}(a)\exp(-f(a))$ *be a density with respect to the Lebesgue measure on* $\mathcal{A}$ *such that* $f : \mathcal{A} \to \mathbb{R}$ *is a convex function. Then there exists a polynomial-time algorithm for sampling from* p, *provided one can compute the following efficiently:*

1 *(First-order information):* $\nabla f(a)$ *where* $a \in \mathcal{A}$.

2 *(Euclidean projections):* $\operatorname{argmin}_{x\in\mathcal{A}} \|x - y\|_2$ *where* $y \in \mathbb{R}^d$.

روش سوم، BGD:

THEOREM 27.3. *Assume that $\mathcal{A}$ is compact, convex and has volume* $\operatorname{vol}(\mathcal{A}) = \int_{\mathcal{A}} da > 0$. *Then an appropriately tuned instantiation of the continuous exponential weights algorithm with Kiefer–Wolfowitz exploration has regret bounded by*

$$R_n \leq 2d\sqrt{3n(1 + \log_+(2n/d))}.$$

کاهش گرادیان بندیتی؟

$R \leq O(n^{2/3} D^{3/2})$

توزیع Kiefer–Wolfowitz

$$g(\pi) \doteq \max_{v \in \mathcal{A}} v^\top Q^{-1}(\pi) v \,.$$

پاپان