بسم الله الرحمن الرحیم

یادگیری بندیت

جلسه ۱۱:

وحدت بندیت‌های خطی — فصل ۲۹

ترم بهار ۱۳۹۹ - ۱۴۰۰

مرور بندیت خطی

تصادفی و دشمنانه

بندیت خطی تصادفی

- انتخاب θ^*

t = ۱ تا n

- انتخاب A_t
- دریافت η_t و محاسبه

$$X_t = \langle \theta^*, A_t \rangle + \eta_t$$

- دریافت A_t
- انتخاب عمل A_t
- دریافت X_t

- انتخاب بهترین رقیب:
-

بندیت خطی تصادفی

- انتخاب θ^*

t=۱ تا n

- انتخاب $\mathcal{A}_t$
- دریافت η_t و محاسبه

$$X_t = \langle \theta^*, A_t \rangle + \eta_t$$

- دریافت $\mathcal{A}_t$
- انتخاب عمل A_t
- دریافت X_t

- انتخاب بهترین رقیب:

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

بندیت خطی تصادفی (نسخه ضرر)

- انتخاب θ^*

t = ۱ تا n

- انتخاب $\mathcal{A}_t$
- دریافت و محاسبه η_t

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

- دریافت $\mathcal{A}_t$
- انتخاب عمل A_t
- دریافت Y_t

- انتخاب بهترین رقیب:

$$\mathbb{E}\left[\sum_{t=1}^{n} Y_t - \min_{a_t \in \mathcal{A}_t} \langle \theta^*, a_t \rangle\right]$$

$$\mathbb{E}\left[\sum_{t=1}^{n} Y_t - \min_{a_t \in A_t} \langle \theta^*, a_t \rangle\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n} (1 - X_t) - \min_{a_t \in A_t} (1 - \langle \theta^*, a_t \rangle)\right]$$

$$= \mathbb{E}\left[-\sum_{t=1}^{n} X_t - \min_{a_t \in A_t} (- \langle \theta^*, a_t \rangle)\right]$$

$$= \mathbb{E}\left[\max_{a_t \in A_t} \langle \theta^*, a_t \rangle - \sum_{t=1}^{n} X_t\right]$$

بندیت خطی تصادفی (نسخه ضرر) ـ مجموعه عملیات ثابت

- انتخاب $\mathcal{A}$
- انتخاب θ^*

- دریافت $\mathcal{A}$

t= ۱ تا n

- انتخاب عمل A_t
- دریافت Y_t

- دریافت η_t و محاسبه

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

- انتخاب بهترین رقیب:

بندیت خطی تصادفی (نسخه ضرر) ـ مجموعه عملیات ثابت

- دریافت $\mathcal{A}$

- انتخاب $\mathcal{A}$
- انتخاب θ^*

$t = ۱$ تا n

- انتخاب عمل A_t
- دریافت Y_t

- دریافت η_t و محاسبه

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

- انتخاب بهترین رقیب:

$$R_n = \sum_{t=1}^{n} \mathbb{E}\left[\langle A_t, \theta \rangle\right] - n \inf_{a \in \mathcal{A}} \langle a, \theta \rangle,$$

بندیت خطی تصادفی ـ کران‌ها

برای الگوریتم UCB

$$R_n \leq Cd\sqrt{n}\log(nL),$$

بندیت خطی تصادفی ـ کران‌ها

برای الگوریتم UCB

$$R_n \leq Cd\sqrt{n}\log(nL),$$

در گوی واحد

$$R_n(\mathcal{A}, \theta) \geq d\sqrt{n}/(16\sqrt{3}).$$

بندیت خطی دشمنانه

- انتخاب $\mathcal{A}$
- انتخاب θ_tها

- دریافت $\mathcal{A}$

t = ۱ تا n

- انتخاب عمل A_t
- محاسبه $Y_t = \langle \theta_t, A_t \rangle$
- دریافت Y_t

- انتخاب بهترین رقیب:

بندیت خطی دشمنانه

- دریافت $\mathcal{A}$
- انتخاب $\mathcal{A}$
- انتخاب θ_tها

t = ۱ تا n

- انتخاب عمل A_t
- محاسبه $Y_t = \langle \theta_t, A_t \rangle$
- دریافت Y_t

- انتخاب بهترین رقیب:

$$R_n = \sum_{t=1}^{n} \mathbb{E}\left[\langle A_t, \theta_t \rangle\right] - n \inf_{a \in \mathcal{A}} \langle a, \bar{\theta}_n \rangle$$

بندیت خطی دشمنانه ـ کران‌ها

با الگوریتم EXP3 برای $k = |\mathcal{A}|$

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

بندیت خطی دشمنانه ــ کران‌ها

با الگوریتم EXP3 برای $k = |\mathcal{A}|$

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

با الگوریتم وزن‌های نمایی پیوسته برای $\mathcal{A}$ فشرده و محدب

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

بندیت خطی دشمنانه ـ کران‌ها

با الگوریتم EXP3 برای $k = |\mathcal{A}|$

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

با الگوریتم وزن‌های نمایی پیوسته برای $\mathcal{A}$ فشرده و محدب

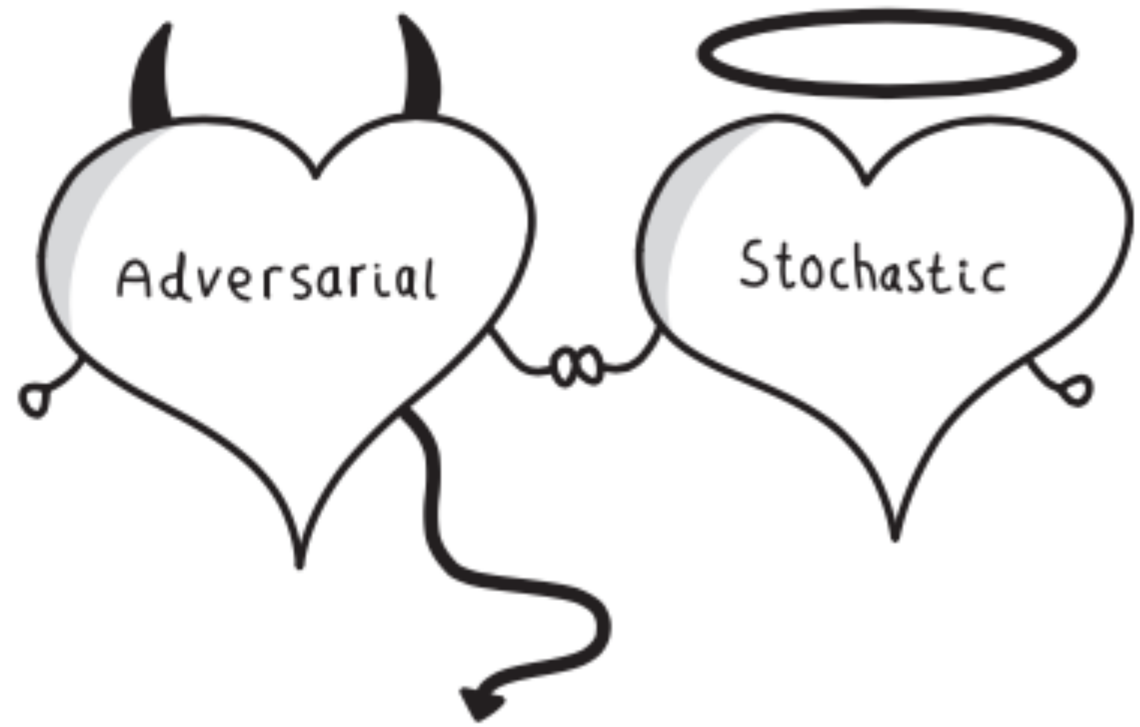
$$R_n \leq 2d\sqrt{3n(1 + \log_+(2n/d))}$$

با الگوریتم پیروی از پیش‌روی منظم‌شده روی گوی واحد

$$R_n \leq 2\sqrt{3nd\log(n)}.$$

وحدت بندیت‌ها

تصادفی و دشمنانه



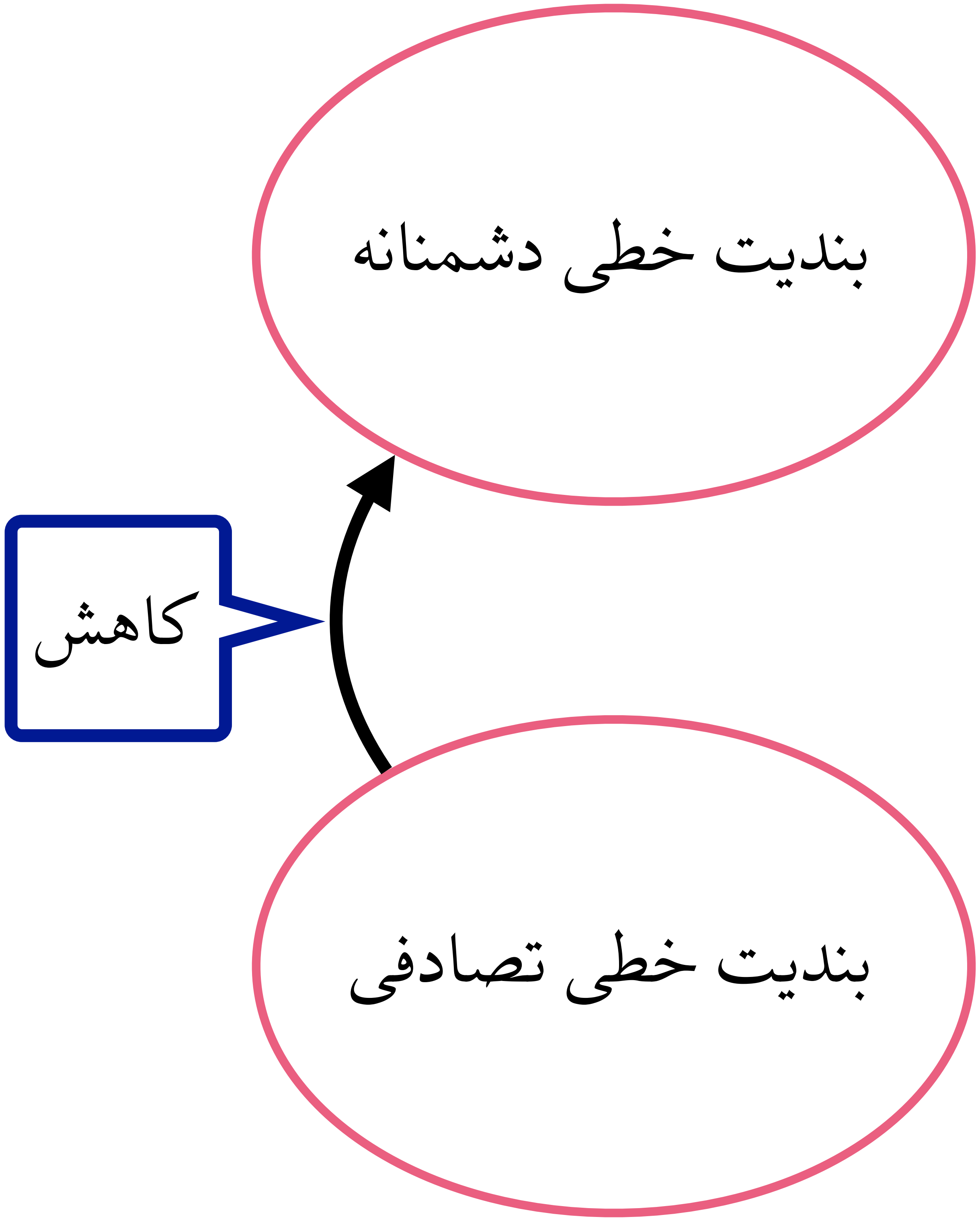
تصادفی

$$Y_t = \langle A_t, \theta \rangle + \eta_t, \qquad R_n = \sum_{t=1}^{n} \mathbb{E}\left[\langle A_t, \theta \rangle\right] - n \inf_{a \in \mathcal{A}} \langle a, \theta \rangle,$$

دشمنانه

$$Y_t = \langle A_t, \theta_t \rangle, \qquad R_n = \sum_{t=1}^{n} \mathbb{E}\left[\langle A_t, \theta_t \rangle\right] - n \inf_{a \in \mathcal{A}} \langle a, \bar{\theta}_n \rangle.$$

۱ـ کاهش تصادفی به دشمنانه



کاهش

- انتخاب $\mathcal{A}$
- انتخاب θ^*

t = ۱ تا n

- دریافت η_t و محاسبه

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

- دریافت $\mathcal{A}$
- انتخاب عمل A_t
- دریافت Y_t

???

دشمنانه

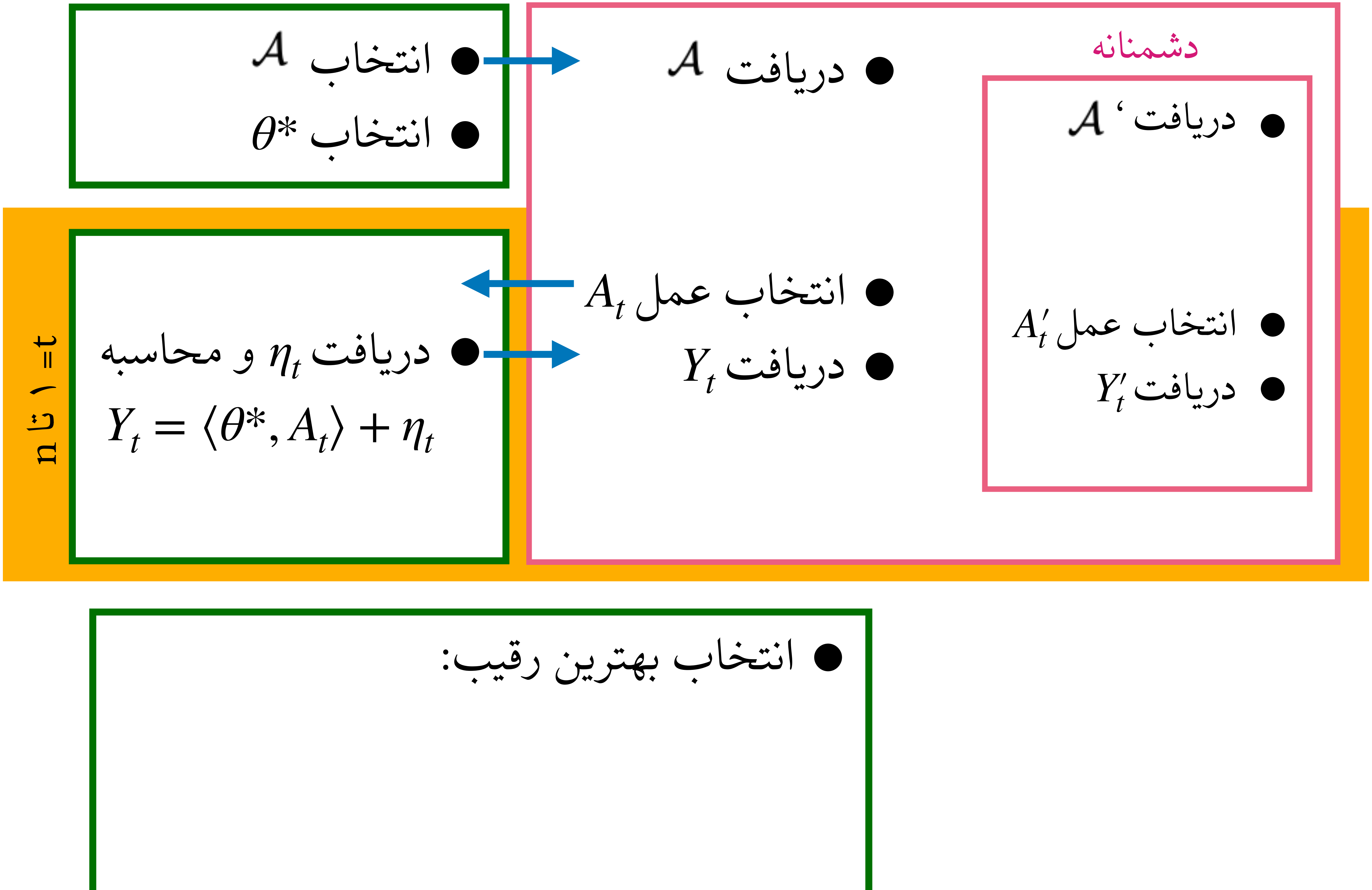
- دریافت $\mathcal{A}$ ‘
- انتخاب عمل A_t'
- دریافت Y_t'

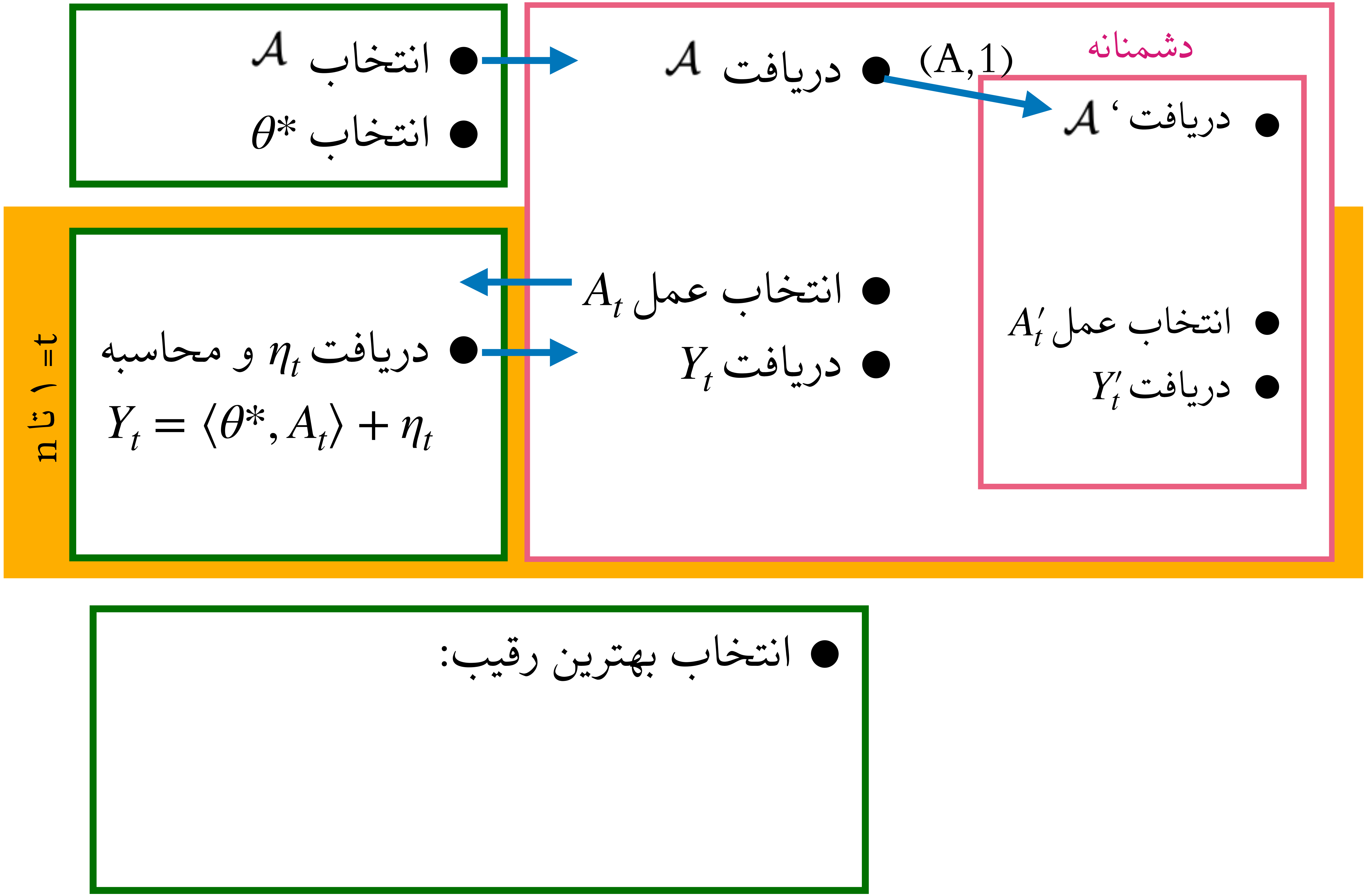
- انتخاب بهترین رقیب:

ایده: یک بعد جدید

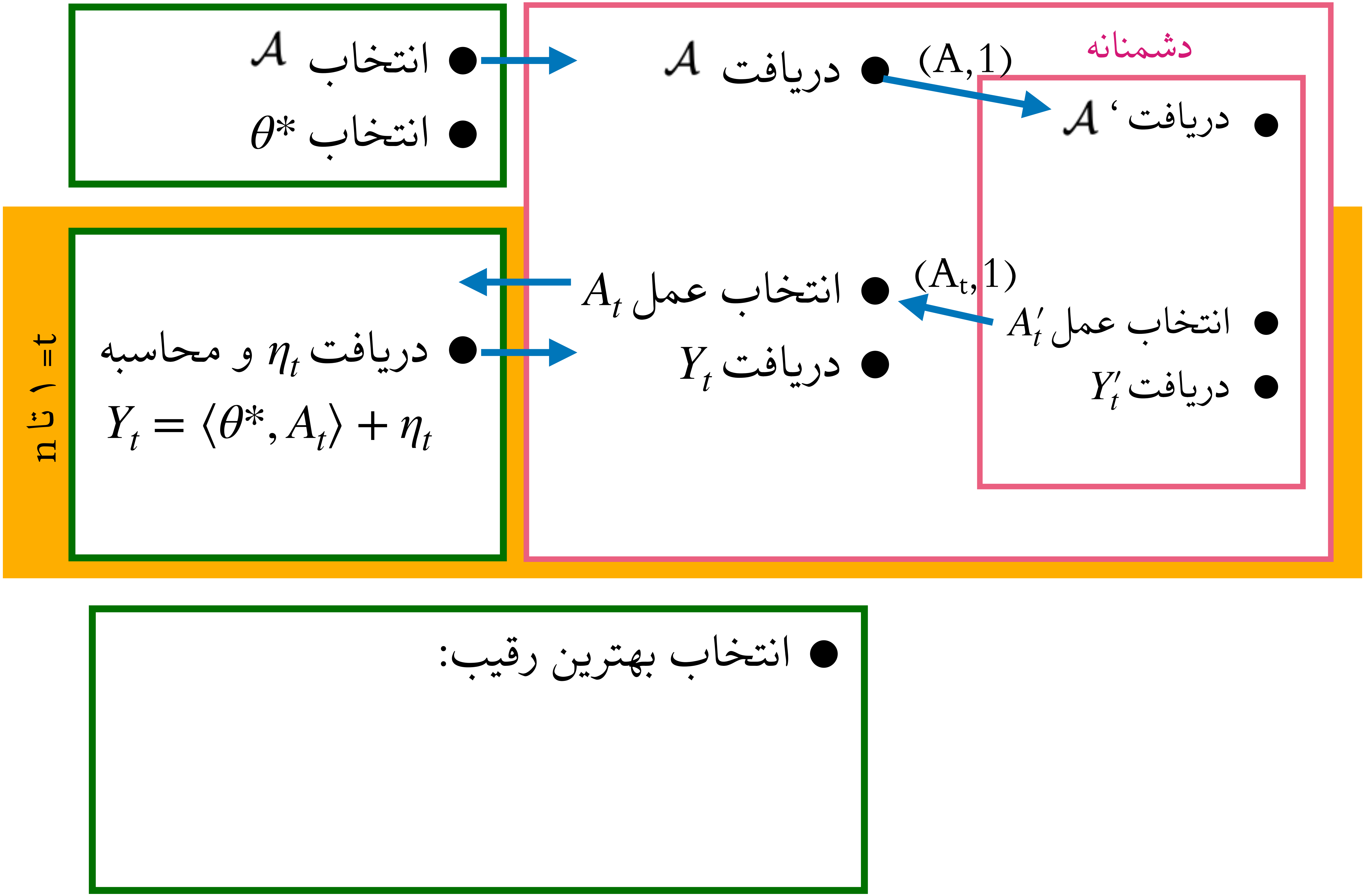
فرض: $|\langle a, \theta \rangle + \eta_t| \leq 1$

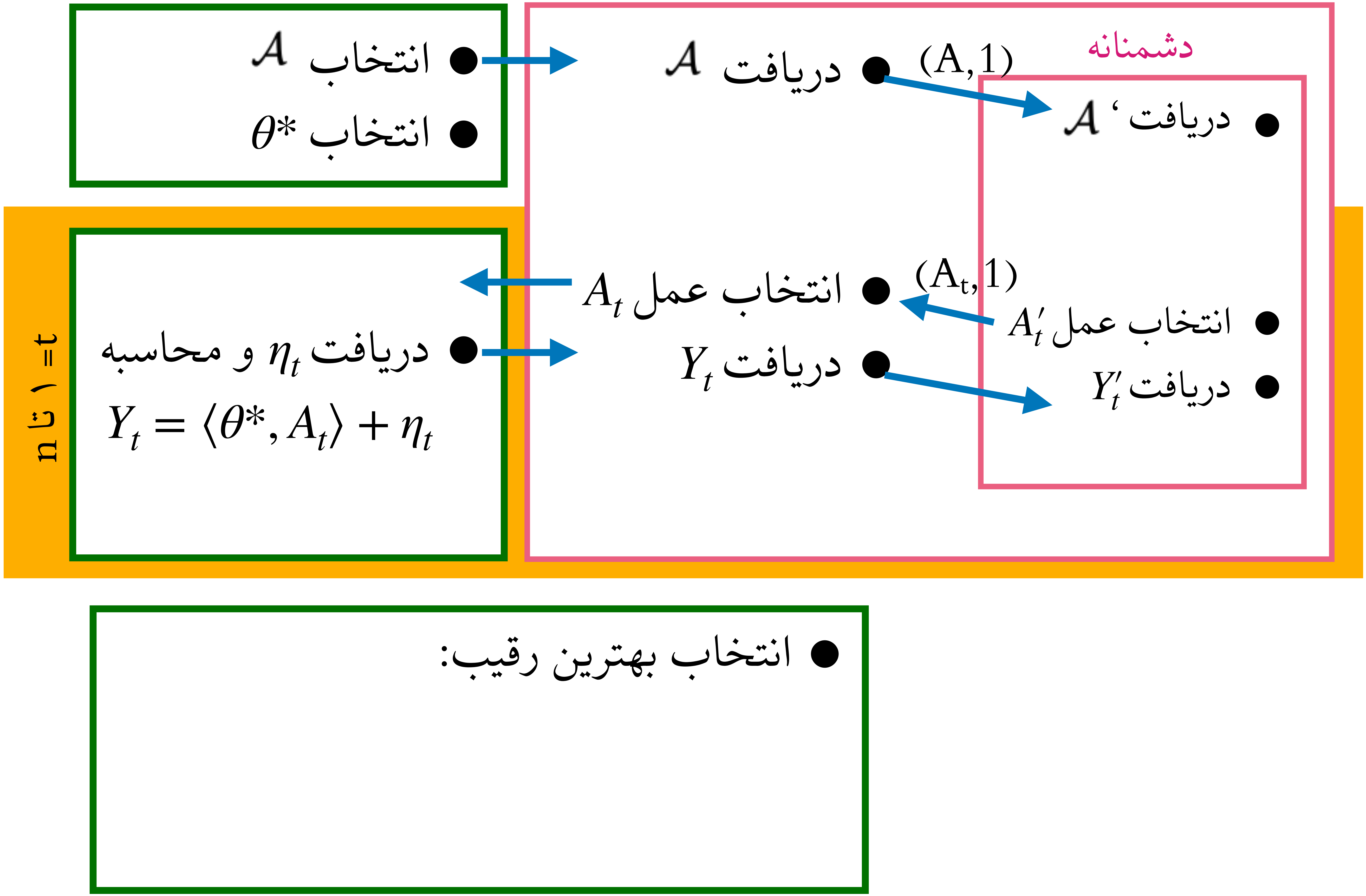
$$\mathcal{A}_{\text{aug}} = \{(a, 1) : a \in \mathcal{A}\} \subset \mathbb{R}^{d+1}$$


انتخاب $\mathcal{A}$
انتخاب θ^*
دریافت $\mathcal{A}$
دشمنانه
دریافت ‘ $\mathcal{A}$
انتخاب عمل A_t
دریافت Y_t
انتخاب عمل A'_t
دریافت Y'_t
دریافت η_t و محاسبه
$Y_t = \langle \theta^*, A_t \rangle + \eta_t$
t= ۱ تا n
انتخاب بهترین رقیب:

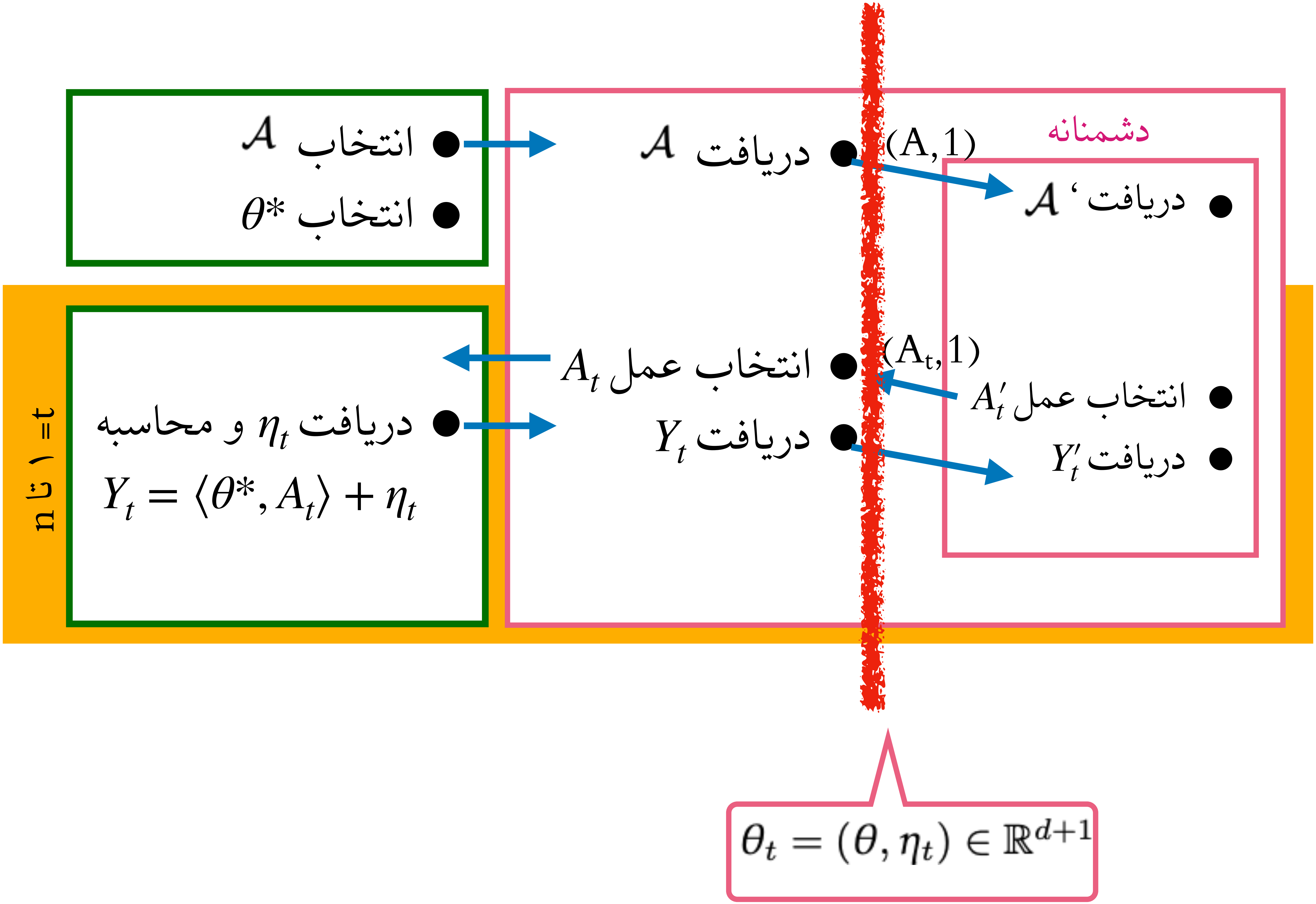


انتخاب $\mathcal{A}$
انتخاب θ^*
دشمنانه
(A,1)
دریافت $\mathcal{A}$
دریافت ‘$\mathcal{A}$
t= ۱ تا n
انتخاب عمل A_t
دریافت η_t و محاسبه
$Y_t = \langle \theta^*, A_t \rangle + \eta_t$
دریافت Y_t
انتخاب عمل A'_t
دریافت Y'_t
انتخاب بهترین رقیب:


انتخاب $\mathcal{A}$
انتخاب θ^*
دریافت $\mathcal{A}$
(A,1)
دشمنانه
دریافت ‘$\mathcal{A}$
t= ۱ تا n
انتخاب عمل A_t
$(A_t, 1)$
انتخاب عمل A'_t
دریافت η_t و محاسبه
$Y_t = \langle \theta^*, A_t \rangle + \eta_t$
دریافت Y_t
دریافت Y'_t
انتخاب بهترین رقیب:



انتخاب $\mathcal{A}$
انتخاب θ^*
دشمنانه
دریافت $\mathcal{A}$
(A,1)
دریافت $\mathcal{A}$ ‘
t= ۱ تا n
دریافت η_t و محاسبه
$Y_t = \langle \theta^*, A_t \rangle + \eta_t$
انتخاب عمل A_t
(A_t, 1)
دریافت Y_t
انتخاب عمل A_t'
دریافت Y_t'
انتخاب بهترین رقیب:



● انتخاب $\mathcal{A}$
● انتخاب θ^*
● دریافت $\mathcal{A}$
(A,1)
دشمنانه
● دریافت $\mathcal{A}$ ‘
● انتخاب عمل A_t
$(A_t, 1)$
● انتخاب عمل A'_t
● دریافت Y_t
● دریافت Y'_t
● دریافت η_t و محاسبه
$Y_t = \langle \theta^*, A_t \rangle + \eta_t$
t= ۱ تا n
$\theta_t = (\theta, \eta_t) \in \mathbb{R}^{d+1}$

الگوریتم دشمنانه:

$$R'_n = \mathbb{E}\left[\sum_{t=1}^{n}\langle A'_t, \theta_t\rangle - \inf_{a'\in\mathcal{A}_{\text{aug}}}\sum_{t=1}^{n}\langle a', \theta_t\rangle\right] \le B_n$$

$A'_t = (A_t, 1)$

$a' = (a, 1)$

$$\mathbb{E}\left[\sum_{t=1}^{n}\langle A_t, \theta\rangle - n\langle a_*, \theta\rangle\right] = \mathbb{E}\left[\sum_{t=1}^{n}\langle A'_t, \theta_t\rangle - n\langle a'_*, \bar{\theta}_n\rangle\right] \le R'_n \le B_n$$

- مشکل گوی واحد؟
- بین تصادفی و دشمنانه
- با ضرایب جدید: $R_n \leq Cd\sqrt{n \log(n)}$

۲ ـ بندیت خطی تصادفی با نوفه پارامتر

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

تصادفی و مستقل از
توزیع ν

$$\theta = \int x \nu(dx)$$

$$Y_t = \langle A_t, \theta_t \rangle = \langle A_t, \theta \rangle + \langle A_t, \theta_t - \theta \rangle.$$

$$Y_t = \langle A_t, \theta \rangle + \widetilde{\eta_t}$$

$$\tilde{\eta}_t = \langle A_t, \theta_t - \theta \rangle$$

$$\mathbb{E}_t \left[\widetilde{\eta_t} \right] = 0.$$

واریانس نوفه

$$\mathbb{V}_t[\tilde{\eta}_t] = \mathbb{E}_t[\langle A_t, \theta_t - \theta\rangle^2]$$

$$= A_t^\top \mathbb{E}_t[(\theta_t - \theta)(\theta_t - \theta)^\top] A_t$$

$$= A_t^\top \Sigma A_t ,$$

۳ــ بندیت خطی زمینه‌ای

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

$$\mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a_t^*, \theta \rangle\right] = \tilde{O}(d\sqrt{n}),$$

۳ـ بندیت خطی زمینه‌ای

$$\mathbb{E}\left[\sum_{t=1}^{n}\langle A_t - a_t^*, \theta\rangle\right] = \tilde{O}(d\sqrt{n}),$$

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

تعمیم به دشمنانه

۳ـ بندیت خطی زمینه‌ای

$$\mathrm{argmax}_{a \in \mathcal{A}_t} \langle a, \theta \rangle$$

$$\mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a_t^*, \theta \rangle\right] = \tilde{O}(d\sqrt{n}),$$

تعمیم به دشمنانه

$$R_n(\Theta) = \max_{\theta \in \Theta} \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a_t(\theta), y_t \rangle\right]$$

۳ــ بندیت خطی زمینه‌ای

$$\mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a_t^*, \theta \rangle\right] = \tilde{O}(d\sqrt{n}),$$

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

تعمیم به دشمنانه

$$R_n(\Theta) = \max_{\theta \in \Theta} \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a_t(\theta), y_t \rangle\right]$$

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

$$\mathcal{A}_t = \{a_1(t), \ldots, a_k(t)\}$$

$$R_n = \max_{i \in [k]} \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a_i(t), y_t \rangle\right] \leq 2\sqrt{3dn\log(k)}.$$

بسم الله الرحمن الرحیم

یادگیری بندیت

جلسه ۱۲:

پیروی از پیش‌روی منظم شده + کاهش آینه‌ای — فصل ۲۸

ترم بهار ۱۳۹۹ - ۱۴۰۰

صورت مسئله

بهینه‌سازی خطی برخط

$\mathcal{L} \subset \mathbb{R}^d \qquad \mathcal{A} \subset \mathbb{R}^d$

۱ــ انتخاب عمل $a_t \in A$

۲ــ دریافت $y_t \in L$

۳ــ هزینه: $\langle a_t, y_t \rangle$

تفاوت با بندیت

بهینه‌سازی خطی برخط

$\mathcal{L} \subset \mathbb{R}^d$ $\quad\quad$ $\mathcal{A} \subset \mathbb{R}^d$

تفاوت با بندیت

۱ـ انتخاب عمل $a_t \in A$

۲ـ دریافت $y_t \in L$

۳ـ هزینه: $\langle a_t, y_t \rangle$

$$R_n(a) = \sum_{t=1}^{n} \langle a_t - a, y_t \rangle$$

$$R_n = \max_{a \in \mathcal{A}} R_n(a)$$

الگوریتم‌های پیشنهادی

روش‌های کاهش

$$A = \mathbb{R}^d$$

روش‌های کاهش

$$a_{t+1} = y_t$$

$$A = \mathbb{R}^d$$

روش‌های کاهش

$$A = \mathbb{R}^d$$

$$a_{t+1} = y_t$$

$$a_{t+1} = a_t - \eta y_t$$

روش‌های کاهش

$$a_{t+1} = y_t$$

$A = \mathbb{R}^d$

$$a_{t+1} = a_t - \eta y_t$$

A محدب

روش‌های کاهش

$A = \mathbb{R}^d$

$$a_{t+1} = y_t$$

$$a_{t+1} = a_t - \eta y_t$$

A محدب

$$a_{t+1} = \arg\min_{a \in A} \langle a, y_t \rangle$$

$$a_{t+1} = \Pi(a_t - \eta y_t)$$

روش‌های کاهش

$A = \mathbb{R}^d$

$$a_{t+1} = y_t$$

$$a_{t+1} = a_t - \eta y_t$$

A محدب

$$a_{t+1} = \arg\min_{a \in A} \langle a, y_t \rangle$$

$$a_{t+1} = \Pi(a_t - \eta y_t)$$

$$a_{t+1} = \arg\min_{a \in A} \eta \langle a, y_t \rangle + D_F(a, a_t)$$

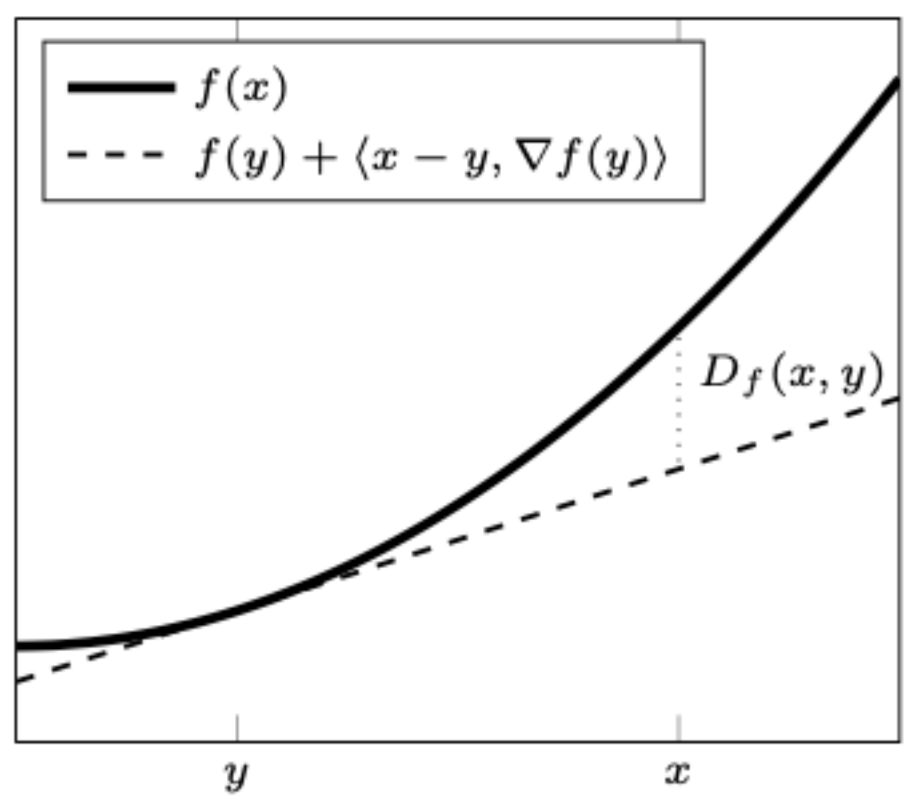
کاهش آینه‌ای

لژاندر: مانند بشقاب ته‌گود

$$a_1 = \mathrm{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

$$D_f(x, y) = f(x) - f(y) - \langle x - y, \nabla f(y) \rangle$$



روش‌های پیروی از پیش‌رو

$$a_{t+1} = \arg\min_{a \in A} \sum_{s} \langle a, y_s \rangle$$

روش‌های پیروی از پیش‌رو

$$a_{t+1} = \arg\min_{a \in A} \sum_s \langle a, y_s \rangle$$

$$a_{t+1} = \arg\min_{a \in A} \eta \sum_s \langle a, y_s \rangle + F(a)$$

پیروی از پیش‌روی منظم شده

$$a_1 = \mathrm{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

تابع لژاندر

(a) C is non-empty;

(b) f is differentiable and strictly convex on C; and

(c) $\lim_{n\to\infty} \|\nabla f(x_n)\|_2 = \infty$ for any sequence $(x_n)_n$ with $x_n \in C$ for all n and $\lim_{n\to\infty} x_n = x$ and some $x \in \partial C$.

چند تابع خوب لژاندر

$f(x) = \frac{1}{2}\|x\|_2^2$ $\mathrm{dom}(f) = \mathbb{R}^d$

چند تابع خوب لژاندر

$f(x) = \frac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y) \rangle$$

چند تابع خوب لژاندر

$f(x) = \frac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$

$$D_f(x, y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle$$

$$= \frac{1}{2}||x||_2^2 - \frac{1}{2}||y||_2^2 - \langle x - y, y\rangle$$

چند تابع خوب لژاندر

$f(x) = \frac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle$$

$$= \frac{1}{2}||x||_2^2 - \frac{1}{2}||y||_2^2 - \langle x - y, y\rangle$$

$$= \frac{1}{2}(||x||_2^2 + ||y||_2^2 - 2\langle x, y\rangle)$$

چند تابع خوب لژاندر

$f(x) = \frac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y) \rangle$$

$$= \frac{1}{2}||x||_2^2 - \frac{1}{2}||y||_2^2 - \langle x - y, y \rangle$$

$$= \frac{1}{2}(||x||_2^2 + ||y||_2^2 - 2\langle x, y \rangle) \quad = \frac{1}{2}(||x - y||_2^2)$$

چند تابع خوب لژاندر

$f(x) = \frac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle$$

$$= \frac{1}{2}||x||_2^2 - \frac{1}{2}||y||_2^2 - \langle x - y, y\rangle$$

$$= \frac{1}{2}(||x||_2^2 + ||y||_2^2 - 2\langle x, y\rangle) \quad = \frac{1}{2}(||x-y||_2^2)$$

$f(x) \;=\; \sum_i x_i \log(x_i) \;-\; x_i \qquad \mathrm{dom}(f) \;=\; [0,\infty)^d$

چند تابع خوب لژاندر

$f(x) = \frac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$

$$D_f(x, y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle$$

$$= \frac{1}{2}||x||_2^2 - \frac{1}{2}||y||_2^2 - \langle x - y, y\rangle$$

$$= \frac{1}{2}(||x||_2^2 + ||y||_2^2 - 2\langle x, y\rangle) \quad = \frac{1}{2}(||x - y||_2^2)$$

$f(x) \;=\; \sum_i x_i \log(x_i) \;-\; x_i \qquad \mathrm{dom}(f) \;=\; [0, \infty)^d$

$$\nabla f(x) \;=\; \log(x)$$

چند تابع خوب لژاندر

$$f(x) = \tfrac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y) \rangle$$

$$= \frac{1}{2}||x||_2^2 - \frac{1}{2}||y||_2^2 - \langle x - y, y \rangle$$

$$= \frac{1}{2}(||x||_2^2 + ||y||_2^2 - 2\langle x, y \rangle) \quad = \frac{1}{2}(||x - y||_2^2)$$

$$f(x) \;=\; \textstyle\sum_i x_i \log(x_i) \;-\; x_i \qquad \mathrm{dom}(f) \;=\; [0,\infty)^d$$

$$\nabla f(x) \;=\; \log(x)$$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y) \rangle$$

چند تابع خوب لژاندر

$$f(x) = \tfrac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle$$

$$= \frac{1}{2}||x||_2^2 - \frac{1}{2}||y||_2^2 - \langle x - y, y\rangle$$

$$= \frac{1}{2}(||x||_2^2 + ||y||_2^2 - 2\langle x, y\rangle) \quad = \frac{1}{2}(||x - y||_2^2)$$

$$f(x) = \sum_i x_i \log(x_i) - x_i \qquad \mathrm{dom}(f) = [0,\infty)^d$$

$$\nabla f(x) = \log(x)$$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle$$

$$= \langle x, \log(x)\rangle - \langle x, 1\rangle - \langle y, \log(y)\rangle + \langle y, 1\rangle - \langle x, \log(y)\rangle + \langle y, \log(y)\rangle$$

چند تابع خوب لژاندر

$$f(x) = \tfrac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle$$

$$= \frac{1}{2}||x||_2^2 - \frac{1}{2}||y||_2^2 - \langle x - y, y\rangle$$

$$= \frac{1}{2}(||x||_2^2 + ||y||_2^2 - 2\langle x, y\rangle) \quad = \frac{1}{2}(||x - y||_2^2)$$

$$f(x) \;=\; \textstyle\sum_i x_i \log(x_i) \;-\; x_i \qquad \mathrm{dom}(f) \;=\; [0,\infty)^d$$

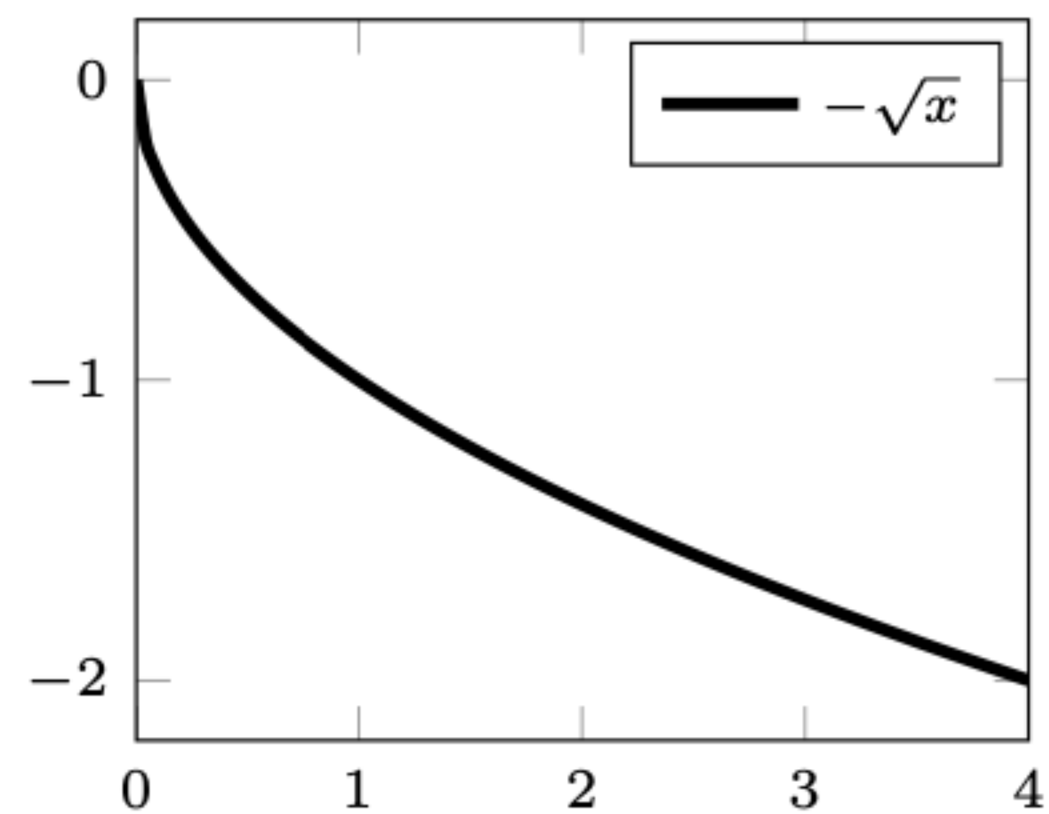
$$\nabla f(x) \;=\; \log(x)$$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle$$

$$= \langle x, \log(x)\rangle - \langle x, 1\rangle - \langle y, \log(y)\rangle + \langle y, 1\rangle - \langle x, \log(y)\rangle + \langle y, \log(y)\rangle$$

$$= \sum x_i \log(x_i / y_i)$$

$$f(x) = -2\sum_{i=1}^{d} \sqrt{x_i} \quad \mathrm{dom}(f) = [0, \infty)^d$$



شباهت دو الگوریتم

برابری دو مدل برای $\mathcal{D} \subseteq \mathcal{A}$.

$$\operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

$$=$$

$$\arg\min_{a \in A} \eta \sum_s \langle a, y_s \rangle + F(a)$$

برابری دو مدل برای $\mathcal{D} \subseteq \mathcal{A}$.

$$\operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$\Phi_t(a) = \eta\langle a, y_t\rangle + D_F(a, a_t) =$$

$$\arg\min_{a\in A} \eta \sum_s \langle a, y_s\rangle + F(a)$$

برابری دو مدل برای $\mathcal{D} \subseteq \mathcal{A}$.

$$\text{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$\Phi_t(a) = \eta\langle a, y_t\rangle + D_F(a, a_t) = \eta\langle a, y_t\rangle + F(a) - F(a_t) - \langle a - a_t, \nabla F(a_t)\rangle$$

$$\arg\min_{a\in A} \eta \sum_s \langle a, y_s\rangle + F(a)$$

برابری دو مدل برای $\mathcal{D} \subseteq \mathcal{A}$.

$$\operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$\Phi_t(a) = \eta\langle a, y_t\rangle + D_F(a, a_t) = \eta\langle a, y_t\rangle + F(a) - F(a_t) - \langle a - a_t, \nabla F(a_t)\rangle$$

$$\eta y_t = \nabla F(a_t) - \nabla F(a_{t+1})$$

$$\arg\min_{a\in A} \eta \sum_s \langle a, y_s\rangle + F(a)$$

برابری دو مدل برای $\mathcal{D} \subseteq \mathcal{A}$.

$$\operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$\Phi_t(a) = \eta\langle a, y_t\rangle + D_F(a, a_t) = \eta\langle a, y_t\rangle + F(a) - F(a_t) - \langle a - a_t, \nabla F(a_t)\rangle$$

$$\nabla\Phi_t(a_{t+1}) \;=\; 0 \qquad \eta y_t = \nabla F(a_t) - \nabla F(a_{t+1})$$

$$\arg\min_{a\in A} \eta \sum_s \langle a, y_s\rangle + F(a)$$

برابری دو مدل برای $\mathcal{D} \subseteq \mathcal{A}$.

$$\operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$\Phi_t(a) = \eta\langle a, y_t\rangle + D_F(a, a_t) = \eta\langle a, y_t\rangle + F(a) - F(a_t) - \langle a - a_t, \nabla F(a_t)\rangle$$

$$\nabla\Phi_t(a_{t+1}) \;=\; 0 \qquad \eta y_t = \nabla F(a_t) - \nabla F(a_{t+1})$$

$$= -\eta\sum_{s=1}^{t} y_s$$

$$\arg\min_{a\in A} \eta \sum_{s} \langle a, y_s\rangle + F(a)$$

برابری دو مدل برای $\mathcal{D} \subseteq \mathcal{A}$.

$$\operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\Phi_t(a) = \eta\langle a, y_t\rangle + D_F(a, a_t) = \eta\langle a, y_t\rangle + F(a) - F(a_t) - \langle a - a_t, \nabla F(a_t)\rangle$$

$$\nabla\Phi_t(a_{t+1}) = 0 \qquad \eta y_t = \nabla F(a_t) - \nabla F(a_{t+1})$$

$$= \nabla F(a_1) - \eta \sum_{s=1}^{t} y_s = -\eta \sum_{s=1}^{t} y_s;$$

$$\arg\min_{a \in A} \eta \sum_s \langle a, y_s \rangle + F(a)$$

برابری دو مدل برای $\mathcal{D} \subseteq \mathcal{A}$.

$$\operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$\Phi_t(a) = \eta\langle a, y_t\rangle + D_F(a, a_t) = \eta\langle a, y_t\rangle + F(a) - F(a_t) - \langle a - a_t, \nabla F(a_t)\rangle$$

$$\nabla\Phi_t(a_{t+1}) = 0 \qquad \eta y_t = \nabla F(a_t) - \nabla F(a_{t+1})$$

$$\nabla F(a_{t+1}) = -\eta y_t + \nabla F(a_t) = \nabla F(a_1) - \eta\sum_{s=1}^{t} y_s = -\eta\sum_{s=1}^{t} y_s;$$

$$\arg\min_{a\in A} \eta\sum_{s} \langle a, y_s\rangle + F(a)$$

برابری دو مدل برای $\mathcal{D} \subseteq \mathcal{A}$.

$$\text{argmin}_{a\in\mathcal{A}}\left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$\nabla F(a_{t+1}) \quad = -\eta \sum_{s=1}^{t} y_s;$$

$$\arg\min_{a\in A} \eta \sum_{s} \langle a, y_s\rangle + F(a)$$

برابری دو مدل برای $\mathcal{D} \subseteq \mathcal{A}$.

$$\mathrm{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$\nabla F(a_{t+1}) \quad = -\eta \sum_{s=1}^{t} y_s ;$$

$$\arg\min_{a\in A} \eta \sum_{s} \langle a, y_s\rangle + F(a)$$

$$\eta \sum_{s=1}^{t} \langle a, y_s\rangle + F(a)$$

برابری دو مدل برای $\mathcal{D} \subseteq \mathcal{A}.$

$$\operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$\nabla F(a_{t+1}) \quad = -\eta \sum_{s=1}^{t} y_s;$$

$$\arg\min_{a\in A} \eta \sum_{s} \langle a, y_s\rangle + F(a)$$

$$\Phi_t'(a) \ = \eta \textstyle\sum_{s=1}^{t}\langle a, y_s\rangle + F(a)$$

برابری دو مدل برای $\mathcal{D} \subseteq \mathcal{A}$.

$$\mathrm{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$\nabla F(a_{t+1}) \quad = -\eta \sum_{s=1}^{t} y_s;$$

$$\arg\min_{a\in A} \eta \sum_s \langle a, y_s\rangle + F(a)$$

$$\Phi_t'(a) \ = \eta \textstyle\sum_{s=1}^{t}\langle a, y_s\rangle + F(a)$$

$$\nabla F(a_{t+1}) = -\eta \sum_{s=1}^{t} y_s.$$

برابری دو مدل برای $\mathcal{D} \subseteq \mathcal{A}$.

$$\operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$\nabla F(a_{t+1}) \quad = -\eta \sum_{s=1}^{t} y_s;$$

$$\arg\min_{a\in A} \eta \sum_{s} \langle a, y_s\rangle + F(a)$$

$$\Phi_t'(a) \;= \eta \textstyle\sum_{s=1}^{t}\langle a, y_s\rangle + F(a)$$

$$\nabla\Phi_t'(a_{t+1}) = 0 \qquad\qquad \nabla F(a_{t+1}) = -\eta\sum_{s=1}^{t} y_s.$$

مثال: کاهش آینه‌ای

$F(a) = \frac{1}{2}\|a\|_2^2$ $\quad\quad$ $\nabla F(a) = a$ $\quad\quad$ $D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathbb{R}^d} \eta \langle a, y_t \rangle + \frac{1}{2}\|a - a_t\|_2^2$$

مثال: کاهش آینه‌ای

$$F(a) = \tfrac{1}{2}\|a\|_2^2 \qquad \nabla F(a) = a \qquad D(a, a_t) = \tfrac{1}{2}\|a - a_t\|_2^2$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathbb{R}^d} \eta\langle a, y_t\rangle + \frac{1}{2}\|a - a_t\|_2^2$$

$$\nabla F(a_{t+1}) = -\eta \sum_{s=1}^{t} y_s$$

مثال: کاهش آینه‌ای

$F(a) = \frac{1}{2}\|a\|_2^2 \qquad \nabla F(a) = a \qquad D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$

$$a_{t+1} = \text{argmin}_{a \in \mathbb{R}^d}\, \eta\langle a, y_t\rangle + \frac{1}{2}\|a - a_t\|_2^2$$

$$\nabla F(a_{t+1}) = -\eta \sum_{s=1}^{t} y_s$$

کاهش گرادیان برخط

مثال: کاهش آینه‌ای

$F(a) = \frac{1}{2}\|a\|_2^2$ $\nabla F(a) = a$ $D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathbb{R}^d} \eta\langle a, y_t\rangle + \frac{1}{2}\|a - a_t\|_2^2$$

$$\nabla F(a_{t+1}) = -\eta \sum_{s=1}^{t} y_s$$

$$a_{t+1} = a_t - \eta y_t$$

کاهش گرادیان برخط

مثال: کاهش آینه‌ای

$$F(a) = \tfrac{1}{2}\|a\|_2^2 \qquad \nabla F(a) = a \qquad D(a, a_t) = \tfrac{1}{2}\|a - a_t\|_2^2$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathbb{R}^d}\, \eta \langle a, y_t \rangle + \frac{1}{2}\|a - a_t\|_2^2$$

$$\nabla F(a_{t+1}) = -\eta \sum_{s=1}^{t} y_s$$

$$a_{t+1} = a_t - \eta y_t$$

کاهش گرادیان برخط

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}}\, \eta \langle a, y_t \rangle + \frac{1}{2}\|a - a_t\|_2^2$$

مثال: کاهش آینه‌ای

$$F(a) = \tfrac{1}{2}\|a\|_2^2 \qquad \nabla F(a) = a \qquad D(a, a_t) = \tfrac{1}{2}\|a - a_t\|_2^2$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathbb{R}^d} \eta \langle a, y_t \rangle + \frac{1}{2}\|a - a_t\|_2^2$$

$$\nabla F(a_{t+1}) = -\eta \sum_{s=1}^{t} y_s$$

$$a_{t+1} = a_t - \eta y_t$$

کاهش گرادیان برخط

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \eta \langle a, y_t \rangle + \frac{1}{2}\|a - a_t\|_2^2 \;= \Pi(a_t - \eta y_t)$$

مثال: پیش‌رو

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

مثال: پیش‌رو

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$a_{t+1,i} = \frac{\exp\left(-\eta \sum_{s=1}^{t} y_{si}\right)}{\sum_{j=1}^{d} \exp\left(-\eta \sum_{s=1}^{t} y_{sj}\right)}.$$

محاسبه کاهش آینه‌ای

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

محاسبه کاهش آینه‌ای

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

محاسبه کاهش آینه‌ای

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\tilde{a}_{t+1} = \operatorname{argmin}_{a \in \mathcal{D}} \eta \langle a, y_t \rangle + D_F(a, a_t)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} D_F(a, \tilde{a}_{t+1}).$$

محاسبه کاهش آینه‌ای

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

$$\tilde{a}_{t+1} = \operatorname{argmin}_{a \in \mathcal{D}} \eta \langle a, y_t \rangle + D_F(a, a_t)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} D_F(a, \tilde{a}_{t+1}).$$

$$\nabla(\ldots) =$$

محاسبه کاهش آینه‌ای

$$a_{t+1} = \operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$\tilde{a}_{t+1} = \operatorname{argmin}_{a\in\mathcal{D}} \eta\langle a, y_t\rangle + D_F(a, a_t)$$

$$a_{t+1} = \operatorname{argmin}_{a\in\mathcal{A}} D_F(a, \tilde{a}_{t+1}).$$

$$D_f(x, y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle$$

$$\nabla(\ldots) = \eta y_t + \nabla F(\tilde{a}_{t+1}) - \nabla F(a_t) = 0.$$

محاسبه کاهش آینه‌ای

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\tilde{a}_{t+1} = \operatorname{argmin}_{a \in \mathcal{D}} \eta \langle a, y_t \rangle + D_F(a, a_t)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} D_F(a, \tilde{a}_{t+1}).$$

$$D_f(x, y) = f(x) - f(y) - \langle x - y, \nabla f(y) \rangle$$

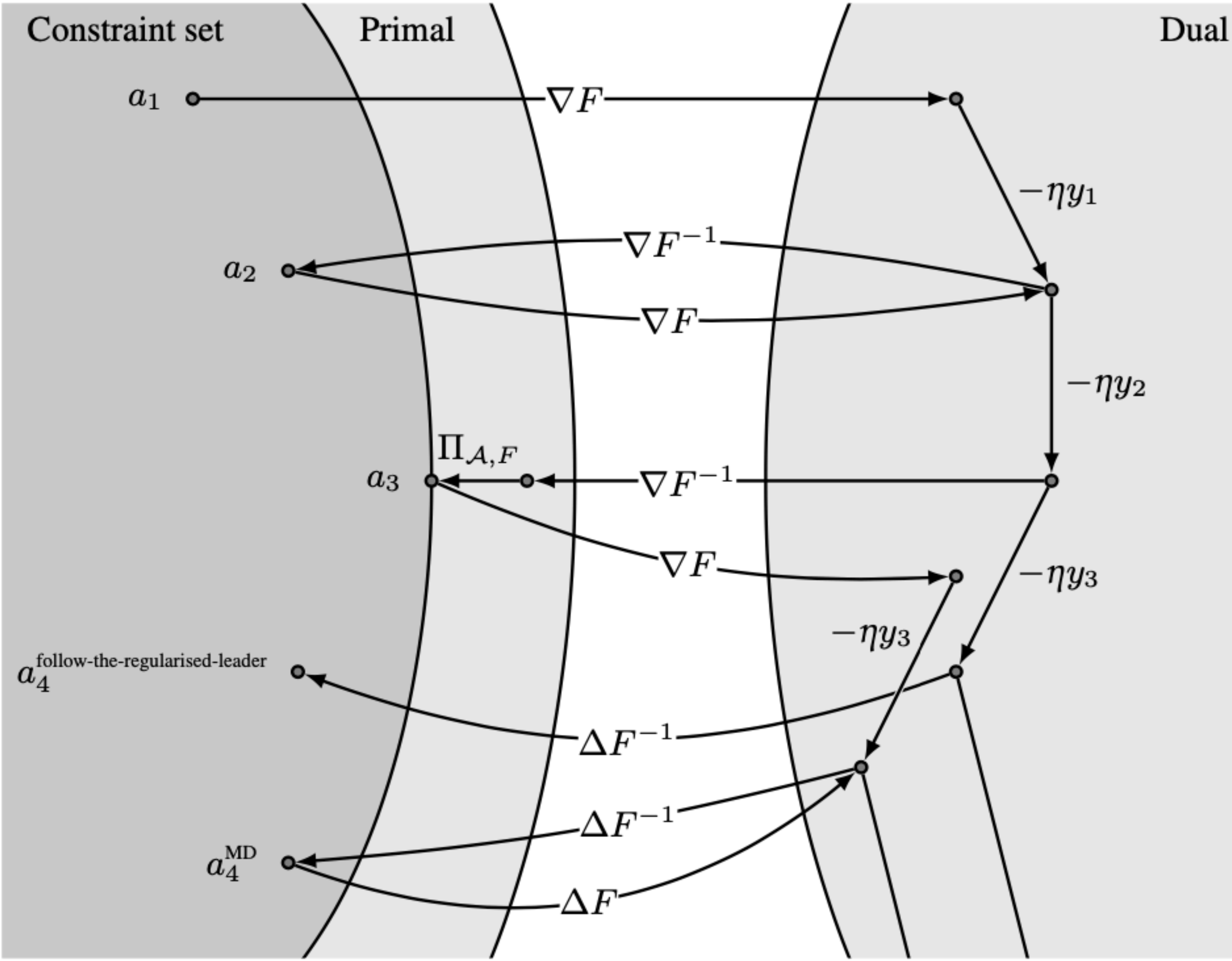
$$\nabla(\ldots) = \eta y_t + \nabla F(\tilde{a}_{t+1}) - \nabla F(a_t) = 0.$$

$$\tilde{a}_{t+1} = (\nabla F)^{-1}(\nabla F(a_t) - \eta y_t)$$

محاسبه پیروی از پیش‌روی منظم شده

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

$$a_{t+1} = \Pi_{\mathcal{A},F} \left(\nabla F^{-1} \left(-\eta \sum_{s=1}^{t} y_s \right) \right)$$



Constraint set
Primal
Dual
a_1
∇F
$-\eta y_1$
∇F^{-1}
a_2
∇F
$-\eta y_2$
$\Pi_{\mathcal{A},F}$
a_3
∇F^{-1}
∇F
$-\eta y_3$
$-\eta y_3$
$a_4^{\text{follow-the-regularised-leader}}$
ΔF^{-1}
ΔF^{-1}
a_4^{MD}
ΔF

تحلیل پشیمانی

کاهش آینه‌ای

پیروی از پیش‌روی منظم شده

تحلیل پشیمانی کاهش آینه‌ای

قضیه:

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$R_n(a) \leq \frac{1}{\eta} \left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1}) \right)$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \geq 0$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \geq 0$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\langle a_t - a, y_t\rangle = \langle a_t - a_{t+1}, y_t\rangle + \langle a_{t+1} - a, y_t\rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t)\rangle \geq 0$$

$$a_{t+1} = \text{argmin}_{a \in \mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$\langle a_{t+1} - a, y_t\rangle \leq \frac{1}{\eta}\langle a - a_{t+1}, \nabla F(a_{t+1}) - \nabla F(a_t)\rangle$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \geq 0$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\begin{aligned} \langle a_{t+1} - a, y_t \rangle &\leq \frac{1}{\eta} \langle a - a_{t+1}, \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \\ &= \frac{1}{\eta} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right) \end{aligned}$$

$$\begin{array}{ccc} F(a) & -F(a) & -F(a_{t+1}) \\ -F(a_t) & +F(a_{t+1}) & +F(a_t) \\ -\langle a - a_t, \nabla F(a_t) \rangle & +\langle a - a_{t+1}, \nabla F(a_{t+1}) \rangle & +\langle a_{t+1} - a_t, \nabla F(a_t) \rangle \end{array}$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \geq 0$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\begin{aligned} \langle a_{t+1} - a, y_t \rangle &\leq \frac{1}{\eta} \langle a - a_{t+1}, \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \\ &= \frac{1}{\eta} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right) \end{aligned}$$

$$\begin{array}{ccc} \cancel{F(a)} & -\cancel{F(a)} & -F(a_{t+1}) \\ -F(a_t) & +F(a_{t+1}) & +F(a_t) \\ -\langle a - a_t, \nabla F(a_t) \rangle & +\langle a - a_{t+1}, \nabla F(a_{t+1}) \rangle & +\langle a_{t+1} - a_t, \nabla F(a_t) \rangle \end{array}$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \geq 0$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\begin{aligned} \langle a_{t+1} - a, y_t \rangle &\leq \frac{1}{\eta} \langle a - a_{t+1}, \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \\ &= \frac{1}{\eta} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right) \end{aligned}$$

$$\begin{array}{ccc} \cancel{F}(a) & -\cancel{F}(a) & -\cancel{F}(a_{t+1}) \\ -F(a_t) & +\cancel{F}(a_{t+1}) & +F(a_t) \\ -\langle a - a_t, \nabla F(a_t) \rangle & +\langle a - a_{t+1}, \nabla F(a_{t+1}) \rangle & +\langle a_{t+1} - a_t, \nabla F(a_t) \rangle \end{array}$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \geq 0$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\begin{aligned} \langle a_{t+1} - a, y_t \rangle &\leq \frac{1}{\eta} \langle a - a_{t+1}, \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \\ &= \frac{1}{\eta} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right) \end{aligned}$$

$$\begin{array}{ccc} \cancel{F(a)} & -\cancel{F(a)} & -\cancel{F(a_{t+1})} \\ -\cancel{F(a_t)} & +\cancel{F(a_{t+1})} & +\cancel{F(a_t)} \\ -\langle a - a_t, \nabla F(a_t) \rangle & +\langle a - a_{t+1}, \nabla F(a_{t+1}) \rangle & +\langle a_{t+1} - a_t, \nabla F(a_t) \rangle \end{array}$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \geq 0$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\begin{aligned} \langle a_{t+1} - a, y_t \rangle &\leq \frac{1}{\eta} \langle a - a_{t+1}, \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \\ &= \frac{1}{\eta} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right) \end{aligned}$$

$$\begin{array}{ccc} \cancel{F(a)} & -\cancel{F(a)} & -\cancel{F(a_{t+1})} \\ -\cancel{F(a_t)} & +\cancel{F(a_{t+1})} & +\cancel{F(a_t)} \\ -\langle a - \cancel{a_t}, \nabla F(a_t) \rangle & +\langle a - a_{t+1}, \nabla F(a_{t+1}) \rangle & +\langle a_{t+1} - \cancel{a_t}, \nabla F(a_t) \rangle \end{array}$$

$$
\begin{aligned}
R_n &= \sum_{t=1}^{n} \langle a_t - a, y_t \rangle \\
&\leq \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \frac{1}{\eta} \sum_{t=1}^{n} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right)
\end{aligned}
$$

$$
\begin{aligned}
R_n &= \sum_{t=1}^{n} \langle a_t - a, y_t \rangle \\
&\le \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \frac{1}{\eta} \sum_{t=1}^{n} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right) \\
&= \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \frac{1}{\eta} \left(D(a, a_1) - D(a, a_{n+1}) - \sum_{t=1}^{n} D(a_{t+1}, a_t) \right)
\end{aligned}
$$

$$
\begin{aligned}
R_n &= \sum_{t=1}^{n} \langle a_t - a, y_t \rangle \\
&\leq \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \frac{1}{\eta} \sum_{t=1}^{n} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right) \\
&= \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \frac{1}{\eta} \left(D(a, a_1) - D(a, a_{n+1}) - \sum_{t=1}^{n} D(a_{t+1}, a_t) \right) \\
&\leq \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \frac{F(a) - F(a_1)}{\eta} - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t)
\end{aligned}
$$

$$\eta y_t = \nabla F(a_t) - \nabla F(a_{t+1})$$

$$\langle a_t - a_{t+1}, y_t \rangle = \frac{1}{\eta} \langle a_t - a_{t+1}, \nabla F(a_t) - \nabla F(\tilde{a}_{t+1}) \rangle$$

$$\eta y_t = \nabla F(a_t) - \nabla F(a_{t+1})$$

$$\langle a_t - a_{t+1}, y_t \rangle = \frac{1}{\eta} \langle a_t - a_{t+1}, \nabla F(a_t) - \nabla F(\tilde{a}_{t+1}) \rangle$$

$$= \frac{1}{\eta} \left(D(a_{t+1}, a_t) + D(a_t, \tilde{a}_{t+1}) - D(a_{t+1}, \tilde{a}_{t+1}) \right)$$

$$\eta y_t = \nabla F(a_t) - \nabla F(a_{t+1})$$

$$\begin{aligned}
\langle a_t - a_{t+1}, y_t \rangle &= \frac{1}{\eta} \langle a_t - a_{t+1}, \nabla F(a_t) - \nabla F(\tilde{a}_{t+1}) \rangle \\
&= \frac{1}{\eta} \left(D(a_{t+1}, a_t) + D(a_t, \tilde{a}_{t+1}) - D(a_{t+1}, \tilde{a}_{t+1}) \right) \\
&\leq \frac{1}{\eta} \left(D(a_{t+1}, a_t) + D(a_t, \tilde{a}_{t+1}) \right)
\end{aligned}$$

تحلیل پشیمانی پیروی از پیش‌روی منظم شده

قضیه:

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

مثال

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $F(a) = \frac{1}{2}\|a\|_2^2$ $\eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

مثال

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $\quad F(a) = \frac{1}{2}\|a\|_2^2 \qquad \eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

مثال

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $\quad F(a) = \frac{1}{2}\|a\|_2^2 \qquad \eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2$$

مثال

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $F(a) = \frac{1}{2}\|a\|_2^2$ $\eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

مثال

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $\quad F(a) = \frac{1}{2}\|a\|_2^2 \qquad \eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

$$R_n \leq \frac{\operatorname{diam}_F(\mathcal{A})}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\|y_t\|_2^2$$

مثال

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $F(a) = \frac{1}{2}\|a\|_2^2$ $\eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

$$R_n \leq \frac{\operatorname{diam}_F(\mathcal{A})}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\|y_t\|_2^2 \leq \frac{1}{2\eta} + \frac{\eta n}{2}$$

مثال

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $F(a) = \frac{1}{2}\|a\|_2^2 \qquad \eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

$$R_n \leq \frac{\operatorname{diam}_F(\mathcal{A})}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\|y_t\|_2^2 \leq \frac{1}{2\eta} + \frac{\eta n}{2} = \sqrt{n}.$$

مثال کاهش آینه‌ای روی سادک

$$R_n \leq \sqrt{2n\log(d)}$$

نسخه بندیتی

بهینه‌سازی خطی برخط

$\mathcal{L} \subset \mathbb{R}^d \quad \mathcal{A} \subset \mathbb{R}^d$

تفاوت با یادگیری برخط
$y_t \in L$

۱ـ انتخاب عمل $A_t \in A$
۲ـ دریافت $\langle A_t, y_t \rangle$

$$R_n(a) = \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a, y_t \rangle\right]$$

$$R_n = \max_{a \in \mathcal{A}} R_n(a)$$

Input Legendre potential F, action set $\mathcal{A}$ and learning rate $\eta > 0$

Choose $\bar{A}_1 = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} F(a)$

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

  Choose measure P_t on $\mathcal{A}$ with mean $\bar{A}_t$

  Sample action A_t from P_t and observe $\langle A_t, y_t \rangle$

  Compute estimate $\hat{Y}_t$ of the loss vector y_t

  Update:

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \langle a, \hat{Y}_t \rangle + D_F(a, \bar{A}_t) \qquad \text{(Mirror descent)}$$

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \sum_{s=1}^{t} \langle a, \hat{Y}_s \rangle + F(a) \qquad \text{(follow-the-regularised-leader)}$$

end for

تحلیل پشیمانی دو الگوریتم بندیت

$$R_n(a) \le \mathbb{E}\left[\frac{F(a) - F(\bar{A}_1)}{\eta} + \sum_{t=1}^{n} \langle \bar{A}_t - \bar{A}_{t+1}, \hat{Y}_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(\bar{A}_{t+1}, \bar{A}_t)\right]$$

$$R_n \le \frac{\mathrm{diam}_F(\mathcal{A})}{\eta} + \frac{1}{\eta} \sum_{t=1}^{n} \mathbb{E}\left[D(\bar{A}_t, \tilde{A}_{t+1})\right]$$

$$\tilde{A}_{t+1} = \mathrm{argmin}_{a \in \mathrm{dom}(F)}\, \eta \langle a, \hat{Y}_t \rangle + D_F(a, \bar{A}_t)$$

$$\mathbb{E}\left[\langle A_t, y_t\rangle\right] = \mathbb{E}\left[\langle \bar{A}_t, y_t\rangle\right] = \mathbb{E}\left[\mathbb{E}\left[\langle \bar{A}_t, y_t\rangle \mid \bar{A}_t\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\langle \bar{A}_t, \hat{Y}_t\rangle \,\middle|\, \bar{A}_t\right]\right]$$

$$R_n(a) = \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t, y_t\rangle - \langle a, y_t\rangle\right] = \mathbb{E}\left[\sum_{t=1}^{n} \langle \bar{A}_t - a, \hat{Y}_t\rangle\right]$$

مثال: بندیت خطی روی گوی واحد

$U_t \quad \{\pm e_1, \ldots, \pm e_d\}$

$E_t \in \{0,1\} \quad \mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$

مثال: بندیت خطی روی گوی واحد

$U_t \quad \{\pm e_1, \dots, \pm e_d\}$

$E_t \in \{0,1\} \quad \mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$

$$A_t = E_t U_t + \frac{(1 - E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

مثال: بندیت خطی روی گوی واحد

$$U_t \quad \{\pm e_1, \ldots, \pm e_d\}$$

$$E_t \in \{0,1\} \quad \mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$$

$$\hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t \rangle}{1 - \|\bar{A}_t\|}$$

$$A_t = E_t U_t + \frac{(1 - E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

مثال: بندیت خطی روی گوی واحد

$$U_t \quad \{\pm e_1, \ldots, \pm e_d\}$$

$$E_t \in \{0,1\} \quad \mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$$

$$A_t = E_t U_t + \frac{(1-E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

$$\hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t \rangle}{1 - \|\bar{A}_t\|}$$

$$F(a) = -\log\left(1 - \|a\|\right) - \|a\|$$

مثال: بندیت خطی روی گوی واحد

$$E_t \in \{0,1\} \quad \mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$$

$$U_t \quad \{\pm e_1, \ldots, \pm e_d\}$$

$$A_t = E_t U_t + \frac{(1-E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

$$\hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t \rangle}{1 - \|\bar{A}_t\|}$$

$$F(a) = -\log\left(1 - \|a\|\right) - \|a\|$$

$$\tilde{\mathcal{A}} = \{x \in \mathbb{R}^d : \|x\|_2 \le r\}$$

مثال: بندیت خطی روی گوی واحد

$$E_t \in \{0,1\} \quad \mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\| \qquad\qquad U_t \quad \{\pm e_1, \ldots, \pm e_d\}$$

$$A_t = E_t U_t + \frac{(1-E_t)\bar{A}_t}{\|\bar{A}_t\|} \qquad\qquad \hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t \rangle}{1 - \|\bar{A}_t\|}$$

$$F(a) = -\log(1 - \|a\|) - \|a\|$$

$$\tilde{\mathcal{A}} = \{x \in \mathbb{R}^d : \|x\|_2 \leq r\}$$

آنگاه $R_n \leq 2\sqrt{3nd\log(n)}.$

پایان