No Regret Dynamics

Monday, October 16, 2023 1:52 PM

☆ External Regret Model.

Decision Maker / player / agent vs Adversary

A: set of action.

☆ $t=1, \dots, T$

1. Agent chooses $p^t \in \Delta(A)$.
2. Adversary (after looking at p^t) chooses $c^t : A \to [0,1]$
3. Agent draws $a^t \sim p^t$ (a^t w.p. $p^t(a^t)$)

incurs cost $c^t(a^t)$

Learns the entire c^t.

Learns $c^t(a^t)$ ≡ Bandit model.

Guarantees are similar upto some loss

Q: How bad $\sum_{t=1}^{T} c^t(a^t)$ (cost of the agent)

compared to "best possible" ?

$$\sum_{t=1}^{T} \text{"} \underbrace{\min_{a \in A} c^t(a)}$$

eg. $A = \{1, 2\}$

$t = 1, \ldots, T$

After looking at p^t adversary chooses ℓ^t as follows

$$\ell^t = (\overset{\ell^t(1)}{1}, \overset{\ell^t(2)}{0}) \quad \text{if } p^t(1) \geq p^t(2)$$

$$= (0, 1) \quad \text{o.w.}$$

$$\text{Expected cost of the agent} = \sum_{t=1}^{T} \boxed{\begin{array}{c}\max\{p^t(1), p^t(2)\}\\ \geq 1/2\end{array}}$$

$$\geq T/2$$

$$\text{"best possible"} = 0$$

★ Compare to the "best-action-in-hindsight".

$$\boxed{\operatorname*{argmin}_{a \in A} \sum_{t=1}^{T} \ell^t(a)}$$

★ Def Algo "$\mathcal{A}$" is no-regret if it's

$$\underline{\text{Time-avg-regret}} = \frac{1}{T}\left[E\left[\begin{array}{c}\text{cost of}\\ \text{the algo}\end{array}\right] - \left[\begin{array}{c}\text{cost of}\\ \text{Best-action-}\\ \text{in hindsight}\end{array}\right]\right]$$

$$\lim_{T \to \infty} \quad \downarrow$$

$$0$$

(Bad eg.) $A = \{1, 2\}$ $\quad c^t(1) \quad c^t(2)$

Oblivious Adversary : $(1, \quad 0)$ w.p. $1/2$

$(0, \quad 1)$ w.p. $1/2$

Expected cost-agent in round t $= \frac{p^t(1)}{2} + \frac{p^t(2)}{2}$

$= 1/2$

$\mathbb{E}[\text{cost-of-agent}] = T/2 \qquad \sum_{\substack{t=1\\ \shortparallel}}^{T} c^t(1) \qquad \sum_{\substack{t=1\\ \shortparallel}}^{T} c^t(2)$

$\mathbb{E}\begin{bmatrix}\text{cost of}\\ \text{best action}\end{bmatrix} = \min\left\{\begin{matrix}\#\text{ times}\\ (1,0)\end{matrix}, \begin{matrix}\#\text{ times}\\ (0,1)\end{matrix}\right\}$

$= T/2 - \Theta(\sqrt{T})$

$\uparrow$

variance.

T-A-Regret $= \frac{1}{\cancel{T}\sqrt{T}}\left[\cancel{T/2} - \cancel{T/2} + \Theta(\cancel{\sqrt{T}})\right]$

$= \Theta\left(\frac{1}{\sqrt{T}}\right) \longrightarrow 0$ as $T \to \infty$

$\downarrow$ extend

n-action $\qquad$ Regret $= \Theta\left(\sqrt{\frac{\ln n}{T}}\right)$

n-action   Regret $O\left(\sqrt{\frac{}{T}}\right)$

Thm: ∃ no-regret Algorithm with regret $\leq O\left(\sqrt{\frac{\ln n}{T}}\right)$

ε

(If want regret $= \varepsilon$ then $T = \frac{\ln n}{\varepsilon^2}$)

* Multiplicative-Weight-Update (MWU)

OR Hedge

OR expert advice

Intuition: Increase prob. of "good actions"

$\equiv$ aggressively punish bad actions.

* Player's MWU Algo:
  * $w^1(a) = 1$, $\forall a \in A$ (Init.)
  * For $t = 1, \ldots, T$

    ① p^t prop. to w^t : $\boxed{p^t(a) = \frac{w^t(a)}{\Gamma^t}}$, $\forall a \in A$

    where $\Gamma^t = \sum_{a \in A} w^t(a)$.

$$1 = \sum_{a \in A} w(a)$$

② After observing $c^t : A \to [0,1]$, update

$$\forall a \in A, \quad \boxed{w^{t+1}(a) = w^t(a)\,(1-\varepsilon)^{c^t(a)}}$$

$\varepsilon \to 0$: $w^{t+1}(a) \sim w^t(a)$ Explore

$\varepsilon \to 1$: $w^{(t+1)}(a) = B.R.$ Exploit.

$\varepsilon \in [0, \frac{1}{2}]$ pick later.

Pf : $\Gamma^t = \sum_{a \in A} w^t(a)$

Goal: Relate $\mathbb{E}[\text{cost of algo}] = \sum_{t=1}^{T} \underbrace{\sum_{a \in A} p^t(a) c^t(a)}_{\text{expected cost in round } t.}$

& $\mathbb{E}\left[\text{cost of best action}\right] = \underbrace{\underset{a \in A}{\overset{a^* \,=}{\operatorname{argmin}}} \sum_{t=1}^{T} c^t(a)}_{OPT}$

via Γ^t.

$$\Gamma^T = \sum_{a \in A} w^T(a) \geq w^T(a^*)$$

$$= w^1(a^*) \prod_{t=1}^{T} (1-\varepsilon)^{c^t(a^*)}$$

$$\Gamma' = \sum_{a \in A} w\ (a) = w(\ )$$

$$= w^{1}(a^*) \prod_{t=1}^{T} (1-\varepsilon)^{c^t(a^*)}$$

$$= 1 \cdot (1-\varepsilon)^{\sum_{t=1}^{T} c^t(a^*)}$$

$$= (1-\varepsilon)^{\text{OPT.}} \longrightarrow ①$$

- Expected cost of agent in time t.

$$\nu^t = \sum_{a \in A} p^t(a)\, c^t(a) = \sum_{a \in A} \frac{w^t(a)}{\Gamma^t} c^t(a) \rightarrow ②$$

$$\Gamma^{t+1} = \sum_{a \in A} w^{t+1}(a)$$

$$= \sum_{a \in A} w^t(a)\ (1-\varepsilon)^{c^t(a)} \qquad \left(\begin{array}{l} \varepsilon \in [0, 1/2] \\ c^t \in [0,1] \end{array}\right)$$

$$\underset{\text{(verify)}}{\leq} \sum_{a \in A} w^t(a)\ (1 - \varepsilon c^t(a))$$

$$= \Gamma^t \left(\boxed{\frac{\sum_{a \in A} w^t(a)}{\Gamma^t}} - \varepsilon \underline{\sum_{a \in A} \frac{w^t(a)}{\Gamma^t} c^t(a)} \right)$$

$$= \underline{\Gamma^t} (1 - \varepsilon \nu^t)$$

$$\Gamma^{t-1} (\ \ \varepsilon \nu^{t-1}) (1 - \varepsilon \nu^t)$$

$$\leq \underbrace{\Gamma^{t-1}\left(1-\varepsilon\nu^{t-1}\right)} \left(1-\varepsilon\nu\right.$$

$$\Gamma^{t+1} \leq \Gamma^{1} \prod_{k=1}^{t} \left(1-\varepsilon \underset{\sim}{\nu}^{k}\right)$$

↳ expected cost of agt in time k.

$$(1-\varepsilon)^{OPT} \underset{\because ①}{\leq} \Gamma^{T} \leq \underset{\underset{n}{\|}}{\Gamma^{1}} \prod_{t=1}^{T} \left(1-\varepsilon\nu^{t}\right) \leftarrow$$

Apply ln

$$OPT \ln(1-\varepsilon) \leq \ln n + \sum_{t=1}^{T} \ln\left(1-\varepsilon\nu^{t}\right)$$

$$\boxed{\begin{aligned} &\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots \\ &-x-x^2 \leq \ln(1-x) \leq -x \end{aligned}}$$

$$\frac{(OPT)\left(+\varepsilon+\varepsilon^2\right)}{\varepsilon} \geq \frac{-\ln n}{\varepsilon} + \sum_{t=1}^{T} \frac{\left(+\varepsilon\nu^{t}\right)}{\varepsilon}$$

$$\Rightarrow \mathbb{E}\left[\text{cost of agt using MWU}\right] = \sum_{t=1}^{T} \nu^{t} - OPT \leq \frac{\ln n}{\varepsilon} + \varepsilon\, OPT$$

$$\Rightarrow \mathbb{E}\left[\text{cost of alg running MWU}\right] = \cancel{t=1} - OPT \qquad \varepsilon.$$

$$\varepsilon = \sqrt{\frac{\ln n}{T}}$$

$$\text{avg Regret} = \frac{1}{T}\left[\sum_{t=1}^{T} v^t - OPT\right] \le \frac{1}{T}\left[\frac{\overset{\sqrt{\ln n}}{\cancel{\ln n}}}{\frac{\cancel{\sqrt{\ln n}}}{\sqrt{T}}} + \sqrt{\frac{\ln n}{T}}\cdot \overset{T}{\cancel{OPT}}\ \frac{T}{\sqrt{T}}\right]$$

$$\le \frac{\cancel{1}}{\cancel{T}\ \sqrt{T}}\left[\sqrt{\ln n}\cdot\cancel{\sqrt{T}} + \sqrt{\ln n}\ \cancel{\sqrt{T}}\right]$$

$$= 2\sqrt{\frac{\ln n}{T}}$$

N players:

$i \in N$ $\quad S_i$: action set.

✲ If each player plays as per MWU.

$$\text{after } \underbrace{T = \frac{\ln n}{\varepsilon^2}}\ \text{rounds} \rightarrow \overset{T.A-}{\text{regret}} \le \varepsilon$$

the "avg play" is ε-CCE.

That is,

suppose player i plays $p_i^t \in \Delta(S_i)$

$$\sigma^t = \prod_{i\in N} p_i^t \qquad \sigma^t(\underset{(s_1,\ldots,s_m)}{\underset{\shortparallel}{s}}) = \prod_{i\in N} p_i^t(s_i) \quad \forall s \in \bigtimes_{i\in N} S_i$$

$$\sigma = \frac{1}{T}\sum_{t=1}^{T} \sigma^t \quad \text{is} \quad \underline{\varepsilon\text{-CCE}}$$

$\forall i, \ \forall s_i' \in S_i$

$$\mathbb{E}_{s\sim\sigma}\left[c_i(s)\right] \underset{(exe)}{\le} \mathbb{E}_{s\sim\sigma}\left[c_i(s_i', s_{-i})\right] + \varepsilon$$

$$\Big\Downarrow \qquad\qquad\qquad\qquad \Big\Downarrow$$

$$\mathbb{E}\left[\begin{matrix}\text{avg}\\ \text{cost of the}\\ \text{agent}\end{matrix}\right] \qquad \mathbb{E}\left[\begin{matrix}\text{avg}\\ \text{cost of the}\\ \text{best-action-in}\\ \text{hindsight}\end{matrix}\right]$$