CS 396: Online Markets

Lecture 5: Online Learning (Cont)

Last Time:

- online learning
- follow the leader
- exponential weights

Today:

- online learning (cont)
- warmup up: be the leader
- perturbed follow the leader

Exercise: Be the Leader

Setup:

- **Alg:** be the leader
  - let $V_j^i = \sum_{r=1}^i v_j^i$
  - in round i choose: $j^i = \text{argmax}_j V_j^i$
- **Input:**

	1	2	3	4	5
Action 1	1/2	0	1	0	1
Action 2	0	1	0	1	0

Question:

- what is OPT (in hindsight) payoff?
- what is payoff of BTL on this input?
- in general, which is bigger OPT or BTL?

Online Learning

"make decisions over time, learn to do well"

Model:

- k actions
- n rounds
- action j's payoff in round i: $v_j^i \in [0, h]$
- in round i:
  (a) choose an action j^i
  (b) learn payoffs $v_1^i, \ldots, v_k^i$
  (c) obtain payoff $v_{j^i}^i$.
- payoff ALG $= \sum_{i=1}^n v_{j^i}^i$

Goal: profit close to best action in hindsight

Def: the **best in hindsight** payoff is

$$\text{OPT} = \max_j \sum_{i=1}^n v_j^i$$

Def: the **regret** of the algorithm is

$\text{Regret}_n = {}^1\!/\!_n[\text{OPT} - \text{ALG}]$

Be the Leader

Alg 0: Be the Leader (BTL)

- let $V_j^i = \sum_{r=1}^i v_j^i$
- in round i choose: $> j^i = \text{argmax}_j V_j^i$

Example: $k = 2$ actions

	1	2	3	4
Action 1	.4	.3	0	1
Action 2	.2	.1	1	0
BTL	.4	.7	1.7	2.7
OPT	.4	.7	1.3	1.7

PICTURE

Thm: BTL $\geq$ OPT

Proof:

- let OPT_i = best-in-hindsight after i rounds.
- let $\text{opt}_i = \text{OPT}_i - \text{OPT}_{i-1}$ (change in OPT_i)
- claim: $\text{btl}_i \geq \text{opt}_i$
  - opt_i = change in leaders' payoffs over round i
  - btl_i = full payoff received by that leader in round i
- $\Rightarrow$ BTL $\geq$ OPT.

Follow the Perturbed Leader

Alg 2: Follow the Perturbed Leader (FTPL)

- learning rate ϵ
- hallucinate: $v_j^0 = h\times$ "num tails of ϵ-bias coin flipped in a row"
- let $V_j^i = \sum_{r=1}^{i} v_j^i$
- in round i choose: $j^i = \operatorname{argmax}_j v_j^0 + V_j^{i-1}$

Example: $k = 2$ actions

	0	1	2	3	4	5	6	...
Action 1	**2**	1/2	0	1	0	1	0	...
Action 2	**3**	0	1	0	1	0	1	...

- OPT $\approx n/2$
- FTPL $\approx n/2$
- "no regret"

Thm: for payoffs in $[0, h]$,

$$\mathbf{E}[\text{FTPL}] \geq (1-\epsilon)\,\text{OPT} - \frac{h}{\epsilon}\ln k.$$

Cor: in n rounds and payoffs in $[0, h]$, tune learning rate ϵ so

$$\mathbf{E}[\text{Regret(FTPL)}] \leq 2h\sqrt{\frac{\ln k}{n}}$$

Proof: same as for EW.

Exercise: Learning Rate

Setup:

- $n = 200$ rounds.
- $k = 10$ actions.
- follow-the-perturbed-leader (FTPL) algorithm
- learning rate $\epsilon = 0.1$

Question: You find out you are going to run for for 400 days? Should you increase or decrease your learning rate?

Q: Why does FTPL work?

A:

1. stability: FTPL $\approx$ BTPL
2. small perturbation: $BTPL \gtrsim$ OPT

Lemma 1: (Stabity) FTPL $\geq (1-\epsilon)$ BTPL

Lemma 2: (Small Perturbation) BTPL $\geq$ OPT $-O(\frac{h}{\epsilon}\log k)$

Proof of Thm:

- combine: FTPL $\geq (1-\epsilon)$ OPT $-O(\frac{h}{\epsilon}\log k)$

Proof of Lemma 2: (intuition)

- BTPL $\geq$ BTL $-\mathbf{E}[\max_j v_j^0]$
- $\mathbf{E}[\max_j v_j^i] = O(\frac{h}{\epsilon}\ln k)$
  - flip coins in rounds.
  - about $(1-\epsilon)$ fraction of actions remaining in each round
  - no actions remain after $\log_{1/(1-\epsilon)} k \approx \frac{1}{\epsilon}\log k$ rounds
- formal proof:
  - compare max of geometric r.v.s to max of exponential r.v.s
  - calculus

Proof of Lemma 1:

- coupling argument
- start with raw scores
- add perturbation as:
  - pick action with lowest total score
  - flip coin:
    * heads: discard
    * tails: add h to score.
  - repeat until one action j^* left
- flip j^*'s coin:
  - tails: (w.p. $1-\epsilon$)
    * best action score > h + second-best score
    * FTPL and FTPL pick j^*
  - heads:
    * FTPL ≥ 0