Adversarial Online Learning with noise

Alon Resler Yishay Mansour

Tel Aviv University

Jun 13, 2019

Adversarial bandits

- A T rounds game between a learner and an adversary
- Set of K actions $A = \{1, \ldots, K\}$
- On round t:
  - The adversary selects a loss vector $\vec{\ell_t} \in \{0,1\}^K$ where $\ell_{i,t}$ is the loss associated with action i at round t
  - The learner chooses an action I_t (usually random)
  - The learner incurs a loss $\ell_{I_t,t}$
  - Finally, the learner observes a feedback

Feedback Types and Regret

- **Full information feedback**: the learner observes $\vec{\ell_t}$
- **Bandit feedback**: the learner observes $\ell_{I_t,t}$
- The learner goal is to minimize the expected regret:

$$Regret(T) = \mathrm{E}\left[\sum_{t=1}^{T} \ell_{I_t,t}\right] - \min_{i \in A} \sum_{t=1}^{T} \ell_{i,t}$$

- We say that the algorithm has vanishing regret if $Regret(T) = o(T)$

Our work

- We study online learning settings in which the feedback is corrupted by random noise
- We consider binary losses xored with the noise, which is a Bernoulli random variable
- We consider both settings: *bandit* feedback and *full information* feedback

Results Summary

Feedback type \ Noise model	Constant noise	Variable noise (Uniform)
Full information (known noise)	$\Theta(\frac{1}{\epsilon}\sqrt{T \ln K})$	$\Theta(T^{2/3} \ln^{1/3} K)$
Full Information (unknown noise)	$\Theta(\frac{1}{\epsilon}\sqrt{T \ln K})$	$\Theta(T)$
Bandit (known noise)	$\tilde{\Theta}(\frac{1}{\epsilon}\sqrt{TK})$	$\tilde{\Theta}(T^{2/3}K^{1/3})$
Bandit (unknown noise)	$\tilde{\Theta}(\frac{1}{\epsilon}\sqrt{TK})$	$\Theta(T)$

Poster @ Pacific Ballroom #156