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

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

- 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

<table>
<thead>
<tr>
<th>Feedback type \ Noise model</th>
<th>Constant noise</th>
<th>Variable noise (Uniform)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full information (known noise)</td>
<td>$\Theta\left(\frac{1}{\epsilon} \sqrt{T \ln K}\right)$</td>
<td>$\Theta\left( T^{2/3} \ln^{1/3} K \right)$</td>
</tr>
<tr>
<td>Full Information (unknown noise)</td>
<td>$\Theta\left(\frac{1}{\epsilon} \sqrt{T \ln K}\right)$</td>
<td>$\Theta\left( T \right)$</td>
</tr>
<tr>
<td>Bandit (known noise)</td>
<td>$\tilde{\Theta}\left(\frac{1}{\epsilon} \sqrt{TK}\right)$</td>
<td>$\tilde{\Theta}\left( T^{2/3} K^{1/3} \right)$</td>
</tr>
<tr>
<td>Bandit (unknown noise)</td>
<td>$\tilde{\Theta}\left(\frac{1}{\epsilon} \sqrt{TK}\right)$</td>
<td>$\Theta\left( T \right)$</td>
</tr>
</tbody>
</table>

Poster @ Pacific Ballroom #156