Yuval Filmus, Steve Hanneke, Idan Mehalel and Shay Moran

Bandit-Feedback Online Multiclass Classification: Variants and Tradeoffs

NeurIPS, December 2024

Yuval Filmus
Technion

Steve Hanneke
Purdue University

Shay Moran
Technion, Google

Online classification

- A repeated game between Lrn and Adv.
- Given: *domain* $\mathcal{X}$, *label set* $\mathcal{Y}$, *concept class* $\mathcal{H}$ of $\mathcal{X} \to \mathcal{Y}$ functions.
- Adv chooses a target function $h^\star \in \mathcal{H}$.
- For $t = 1, \ldots, T$:
  - Adv chooses an instance $x_t \in \mathcal{X}$ and sends it to Lrn.
  - Lrn predicts $\hat{y}_t \in \mathcal{Y}$ (possibly at random, distribution is non-private).
  - Adv provides feedback:
    - *Full-information*: $h^\star(x_t)$, or
    - *Bandit*: $1[\hat{y}_t = h^\star(x_t)]$.
- Target min/max value: the *mistake bound* $M = \mathbb{E}\left[\sum_{t=1}^{T} 1[\hat{y}_t \neq h^\star(x_t)]\right]$.
  - Interested in the *minimax optimal* $M := \mathtt{M}^\star(\mathcal{H}, \text{randomness?}, \text{feedback})$.

Full information

- Full information:
  - [Daniely, Sabato, Ben-David, Shalev-Shwartz ’15]:
    - $\mathtt{M}^\star(\mathscr{H}, \text{rand}, \text{full}) \approx \mathtt{M}^\star(\mathscr{H}, \text{det}, \text{full}) \approx \mathsf{Ldim}(\mathscr{H})$.

Bandit feedback - deterministic learners

- What happens with bandit-feedback?
  - [Auer and Long ’99]:
    - $\mathtt{M}^\star(\mathscr{H}, \text{det}, \text{bandit}) = O\left(\text{Ldim}(\mathscr{H})\,|\,\mathscr{Y}\,|\log|\,\mathscr{Y}\,|\right)$.
  - [Long ’20, Genesson '21]:
    - $\exists \mathscr{H}: \mathtt{M}^\star(\mathscr{H}, \text{det}, \text{bandit}) = \Omega\left(\text{Ldim}(\mathscr{H})\,|\,\mathscr{Y}\,|\log|\,\mathscr{Y}\,|\right)$.

Bandit feedback - randomized learners

- What happens with bandit-feedback, when randomness is allowed?
  - [easy]:
    - $\exists \mathcal{H}: \mathtt{M}^\star(\mathcal{H}, \text{rand}, \text{bandit}) = \Omega\left(\mathsf{Ldim}(\mathcal{H})\,|\,\mathcal{Y}\,|\right)$.
  - Previous SOTA upper bound of order $\mathsf{Ldim}(\mathcal{H})\,|\,\mathcal{Y}\,|\log|\,\mathcal{Y}\,|$ is **deterministic**.
  - Question: Can a randomized algorithm shave the $\log|\,\mathcal{Y}\,|$ factor?
  - Answer: Yes! **[this work]**:
    - $\mathtt{M}^\star(\mathcal{H}, \text{rand}, \text{bandit}) = O\left(\mathsf{Ldim}(\mathcal{H})\,|\,\mathcal{Y}\,|\right)$.

The price of determinism

- Trivially: $\mathtt{M}^\star(\mathscr{H}, \text{det}, \text{bandit}) \geq \mathtt{M}^\star(\mathscr{H}, \text{rand}, \text{bandit})$.
- By how much?
  - [easy]:
    - $\mathtt{M}^\star(\mathscr{H}, \text{det}, \text{bandit}) = O\left(|\mathscr{Y}| \log |\mathscr{Y}| \cdot \mathtt{M}^\star\left(\mathscr{H}, \text{rand}, \text{bandit}\right)\right)$.
  - Question: Is this tight for some classes?
  - Answer: Yes (at least up to a log factor) **[this work]**:
    - $\exists \mathscr{H}\colon \mathtt{M}^\star(\mathscr{H}, \text{det}, \text{bandit}) = \Omega\left(|\mathscr{Y}| \cdot \mathtt{M}^\star\left(\mathscr{H}, \text{rand}, \text{bandit}\right)\right)$.

Main technical contribution

- Tight bound for online classification with expert advice.
- Theorem **[this work]**: Suppose that n experts publish sequential predictions for a target sequence $Y = y_1, \ldots, y_T \in \mathcal{Y}$. Then, there exists a randomized algorithm sequentially predicting Y, that receives bandit feedback and has the mistake bound

$$\mathtt{M}^{\star}(n, \mathsf{OPT}) = O\left(|\mathcal{Y}| \left[\log_{|\mathcal{Y}|} n + \mathsf{OPT} \right] \right).$$

- OPT is the #mistakes made by the best expert.

Thank you!

- **Future work:**
  - Multilabel setting (Multiple correct labels)
  - Similar questions for other feedback models