Algorithmic Chaining and the Role of Partial Feedback in Online Nonparametric Learning

Nicolò Cesa-Bianchi nicolo.cesa-bianchi@unimi.it
Pierre Gaillard pierre.gaillard@inria.fr
Claudio Gentile claudio.gentile@uninsubria.it
Sébastien Gerchinovitz sebastien.gerchinovitz@math.univ-toulouse.fr

Setting: online non parametric learning

At the beginning, the environment chooses Lipschitz loss functions $\ell_1, \dots, \ell_T : \mathcal{Y} \to [0,1]$.

For each round $t = 1, \dots, T$,

- Environment chooses $x_t \in \mathcal{X} = [0,1]^d$
- Learner observes x_t and predicts $\widehat{y}_t \in \mathcal{Y} = [0,1]^p$
- Learner suffers $\ell_t(\widehat{y}_t)$ and obtains feedback
  - the instantaneous loss $\ell_t(\widehat{y}_t)$ in the bandit setting
  - $\ell_t(x)$ for all $x \geq \widehat{y}_t$ in the one-sided feedback setting
  - the loss function ℓ_t in the full information setting.

Goal: minimize the regret

$$\mathrm{Reg}_T(\mathcal{F}) := \sum_{t=1}^{T} \ell_t(\widehat{y}_t) - \inf_{f \in \mathcal{F}} \sum_{t=1}^{T} \ell_t\big(f(x_t)\big)$$

over nonparametric function classes $\mathcal{F} \subseteq \mathcal{Y}^{\mathcal{X}}$.

Main contributions

We design efficient algorithms that achieve following bounds agains the class $\mathcal{F}$ of Lipschitz policies with $\mathcal{Y} = [0,1]$ (i.e., $p = 1$).

Feedback model	Loss functions	Upper bound
Bandit	Lipschitz	$T^{\frac{d+2}{d+3}}$
	Convex	$T^{\frac{d+1}{d+2}}$
One-sided full information	Semi-Lipschitz	$T^{\frac{d+1}{d+2}}$
	Lipschitz	$T^{\frac{d-1/3}{d+2/3}}$
Full information	Lipschitz	$T^{\frac{d-1}{d}}$

First explicit algorithm achieving the minimax rate for full information!

First $\mathcal{O}(\sqrt{T})$ regret bound on the sellers revenue in context-free second-price auctions.

Warmup: bandit feedback

Input: Ball radius $\epsilon > 0$, ϵ-covering $\mathcal{Y}_\epsilon$ of $\mathcal{Y}$ such that $|\mathcal{Y}_\epsilon| \lesssim \epsilon^{-p}$.
for $t = 1, 2, \dots$ **do**

1. Get context $x_t \in \mathcal{X}$;
2. If $x_t \notin$ any existing ball, create a new ball of radius ϵ centered on x_t, and allocate a fresh instance of Exp3;
3. Let the active Exp3 instance be the instance allocated to the existing ball whose center x_s is closest to x_t;
4. Draw an action $\widehat{y}_t$ using the active Exp3 instance;
5. Get $\ell_t(\widehat{y}_t)$ and use it to update the active Exp3 instance.

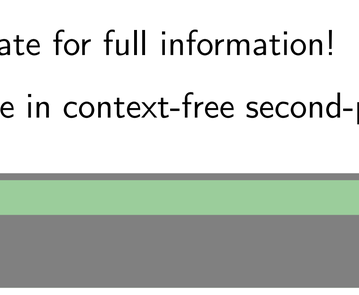
end

Lipschitz losses: if $\epsilon \approx T^{-\frac{1}{p+d+2}}$, then

$$\mathrm{Reg}_T(\mathcal{F}) = \tilde{\mathcal{O}}\big(T^{\frac{p+d+1}{p+d+2}}\big)\,.$$

Convex losses: it exists an algorithm such that

$$\mathrm{Reg}_T(\mathcal{F}) = \tilde{\mathcal{O}}\big(T^{\frac{d+1}{d+2}}\big)\,.$$



We recover the lower-bounds of context-free case when $d = 0$.

Full information

We provide an explicit algorithm that achieves a tight bound for nonparametric learning.

$$\mathrm{Reg}_T(\mathcal{F}) \lesssim \epsilon T + \int_\epsilon^1 \Big(\sqrt{T \ln \mathcal{N}_\infty(\mathcal{F}, x)} + \ln \mathcal{N}_\infty(\mathcal{F}, x)\Big)\,\mathrm{d}x\,. \qquad (1)$$

$$\lesssim T^{\frac{d-1}{d}} \qquad \text{for the class } \mathcal{F} \text{ of Lipschitz functions} \qquad (2)$$

The dimension p of the action space only appears as a multiplicative factor $p^{1/d}$.

Algorithm: the same as the one described to chain the bandits replacing local instances of Exp4 with instances of Hedge.

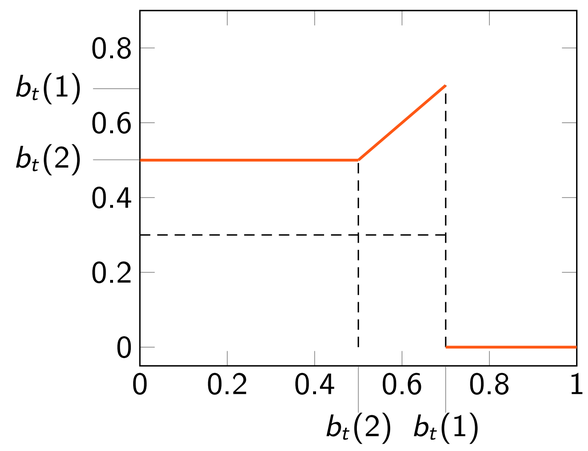
One-sided feedback – example

Exemple: nonparametric second-price auctions

At each round $t = 1, \dots, T$

- The seller (learner) observes x_t and computes a reserve price $\widehat{y}_t$
- Simultaneously, a set of buyers propose bids that we order
  $$b_t(1) \geq b_t(2) \geq \dots$$
- The seller observes the highest bid $b_t(1)$ and his revenue $g_t(\widehat{y}_t)$
  - if $\widehat{y}_t \leq b_t(2)$, the item is sold at the price $b_t(2)$
  - if $b_t(2) \leq \widehat{y}_t \leq b_t(1)$, the item is sold at the price $\widehat{y}_t$
  - if $b_t(1) \leq \widehat{y}_t$, the item is not sold

Knowing $g_t(\widehat{y}_t)$ and $b_t(1)$ allows to compute $g_t(y)$ for all $y \geq \widehat{y}_t$



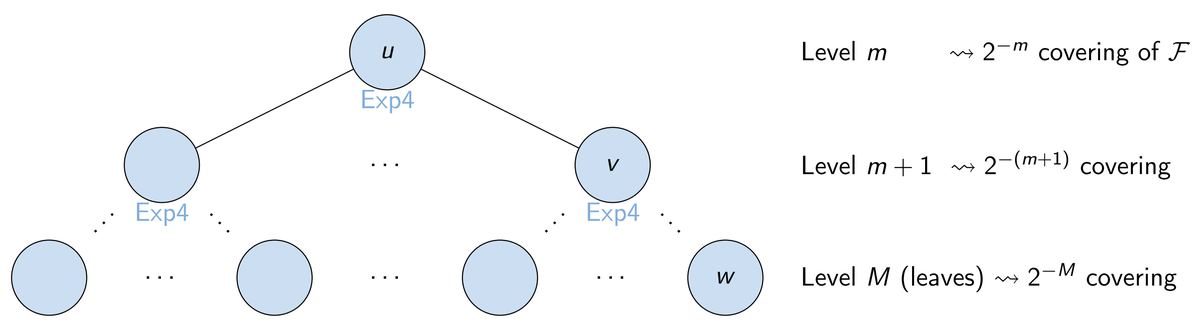
Substituting Exp3 with a variant that takes advantage of the additional feedback to reduce the variance of the loss estimates, in ContextualExp3 yields

$$\mathrm{Reg}_T(\mathcal{F}) = \tilde{\mathcal{O}}\big(T^{\frac{d+1}{d+2}}\big)\,.$$

Chaining the bandits

Ideas of the algorithm: Hierarchical covering of $\mathcal{F}$ = tree whose nodes are functions in $\mathcal{F}$

- The nodes at each depth m define a (2^{-m})-covering of $\mathcal{F}$
- Any function $f^* \in \mathcal{F}$ is represented by a unique path (or chain) in the tree
- The algorithm runs an instance of Exp4 on each node of the tree. The instance A_f at node f uses the predictions of the instances running on the nodes that are children of f as expert advice.



Key elements of the proof:

- Small local ranges: the losses associated with neighboring nodes are close
- Local version of Exp4 that scales with the loss range: possible because of richer feedback
- Sum over a path: the regret against f^* with path $f_0 \to f_1 \to \dots \to f_M \to f^*$ can be written as

$$\sum_{t=1}^{T} \Big(\mathbb{E}\big[\ell_t(A_0(x_t))\big] - \ell_t(f^*(x_t))\Big) \leq \sum_{m=0}^{M-1} \mathbb{E}\Big[\underbrace{\sum_{t=1}^{T} \Big(\ell_t(A_m(x_t)) - \ell_t(A_{m+1}(x_t))\Big)}_{\text{Regret of Exp4 at level } m \,\approx\, 2^{-m}\sqrt{\frac{T \log N_{m+1}}{\gamma}}}\Big] + \frac{T}{2^M}$$

Result: for well-chosen parameters γ and M, if the losses are Lipschitz

$$\mathrm{Reg}_T(\mathcal{F}) \lesssim \frac{T}{2^m} + \sum_{m=0}^{M-1} \sqrt{\frac{T \ln N_{m+1}}{\gamma}} \lesssim \gamma T + \int_\gamma^1 \sqrt{\frac{T}{\gamma} \ln \mathcal{N}(\mathcal{F}, \epsilon)}\mathrm{d}\epsilon$$

$$\lesssim T^{\frac{d}{d+1}} \qquad \text{for the class } \mathcal{F} \text{ of Lipschitz functions}$$

Improvements: efficient algorithm with better rates using penalized loss estimates $\to \tilde{\mathcal{O}}(\sqrt{T})$ regret for non-parametric bandit with one-sided feedback with $d = 1$.