How do we design learning algorithms for multidimensional binary search that are robust to the presence of some irrational agents?

"Irrational" responses: inconsistent with $\theta^*$

→ Even few irrational responses, can have catastrophic effects in standard online learning algorithms.

Model

$\theta^* \in \mathbb{R}^d$: $\|\theta^*\| \leq 1$ : ground truth (unknown to the learner)

$\varepsilon > 0$: target accuracy

Repeated Interaction Protocol between Learner and Opponent

For round $t \in [T]$:  
1. Opponent chooses context $x_t \in \mathbb{R}^d$.
2. Learner observes $x_t$.
3. Opponent corrupts ($\theta_t = \text{arbitrary}$) or not ($\theta_t = \theta^*$).

→ happens at most $C$ times

Learner queries $\omega_t \in \mathbb{R}^d$ & observes feedback:

$y_t = \text{sign}(\langle \omega_t, \theta_t \rangle) \in \{-1, 1\}$.

Learner incurs loss: $\ell(\omega_t, \theta_t) = 1[|\omega_t - \langle \omega_t, \theta_t \rangle| > \varepsilon]$.

Learner’s Goals

1. Minimize: $R(T') = \sum_{t=1}^{T'} \ell(\omega_t, \theta_t)$

Same for standard multidimensional binary search, except $\theta_t = \theta^*$, $\forall t$

$R(T')$ degrading gracefully to $C$

(2.1) $R(T')$: agnostic $C$

Applications for learning in the presence of irrational agents:

Contextual dynamic pricing with linear buyers, Stackelberg Security Games, contextual dynamic pricing with Lipschitz buyers

Key Idea: Undesirability Levels

High level: undesirability expresses the likelihood of where $\theta^*$ lies

Undesirability: $u([0, \frac{1}{2}]) = 1$ but $u([\frac{1}{2}, 1]) \geq C + 1$

→ Generalize the undesirability levels idea for higher dimensions.

Challenges and Solutions

1. Eliminating regions with undesirability $\geq C + 1 \Rightarrow$ non-convexity

Solution: Always make hyperplane cuts

2. Existence of hyperplane cut with undesirability $\geq C + 1$.

Solution: After $2dC(d + 1) + 1$ contexts, there always exists a hyperplane with undesirability at least $C + 1$!

[Carathéodory’s Theorem] Every point $p$ in the convex hull of $P(C)$ can be written as the convex combination of at most $d + 1$ points in $P(C)$.

$u(p) \leq C(d + 1)$

→ For separating $P(C)$, suffices to have point: $u(p^*) = C(d + 1) + 1$.

[Landmarks.] Set of $2d$ points such that at each round one of them gets undesirability point. + pigeonhole


Solution: Building a dataset for perceptron with injected margin.

CorPV.K regret: $O \left(d^2 (d + 1) d \log d \right)$ and expected runtime: $O \left(\frac{d^2 C}{\varepsilon} \cdot \text{poly}(d \log d / \varepsilon, C) \right)$. 

4. Known $C$ and runtime exponential in it.

Solution: Extension of the multi-layering race technique of [Lykouris et al., ’18] for continuous spaces.

References


[Lykouris, Mirrokni, Paes Leme, ’18]. Stochastic Bandits Robust to Adversarial Corruptions. STOC’18.

Main Result

CorPV.A runs in quasipolynomial time, is agnostic in $C$ and achieves regret: $R(T) = O(d^3 (C + \log T) / T \log(d / \varepsilon))$.

Known $C$

Shaded region ($P(C)$): points with $u \leq C$ Patterned region: convex hull of $P(C)$