# Algorithms for Incentive-Compatible and Incentive-Aware Learning

Chara Podimata (Harvard) podimata@g.harvard.edu, www.charapodimata.com

## ML Against Strategic Agents

ML increasingly used for decision-making that affects our lives: *loans, probation, interviews, …*

Heart of ML Paradigm (Especially offline ML)

Patterns in **past/training** data → accurate predictions for **future/test** data

### Incentives for Gaming

Past data describes people's private info →**can** be altered/gamed → achieve better outcomes

Strategic data sources for ML algorithms present both a challenge and an opportunity

Data received for training ML algos is not accurate

ML algorithms can adapt to the behavior of strategic agents.

Solution

Incentive-Compatibility

Incentive-Awareness

**Incentive-Compatibility (IC):** it is in the agents' best interest to report truthfully

## Incentive-Compatible ML

offline setting

[Chen, **Podimata**, Procaccia, Shah, EC18]

Family of IC algorithms for **linear regression**.

- Generalization of Resistant Line estimators (from statistics) in high dims (unknown prior to our work).
- Machinery: Ham Sandwich Theorem (Computational Geometry)

online setting

[Freeman, Pennock, **Podimata**, Vaughan, ICML20]

**No-regret** and IC online learning both for **full** and **partial** information. Optimal regret rate for full information.

- No tradeoff between accuracy + incentives in full info!
- Machinery: connection between online learning and wagering mechanisms

## Incentive-Aware ML

[Krishnamurthy, Lykouris, **Podimata**, Schapire, arXiv20]

Learning **pricing** strategies in the presence of some **irrational** agents.

- Regret deteriorates gracefully with # irrational agents.
- Connections with adversarial ML literature
- Machinery: online learning, convex analysis, behavioral economics

### Technique: *adaptive discretization*

[Chen, Liu, **Podimata**, NeurIPS20]

**Nearly optimal** learning rates for **strategic classification**

- Regret deteriorates gracefully with strategic power of agents.
- Machinery: online learning + bandits

[**Podimata**, Slivkins, arXiv20]

Learning **better** pricing strategies on **"nicer"** instances

- Results are way more general: better learning rates on 1-sided Lipschitz functions
- Machinery: online learning + bandits

[Feng, **Podimata**, Syrgkanis, EC18]

Strategic agents as learning agents.

- Exponentially better guarantees than standard algos
- Machinery: auctions + bandits

## Long-Term Agenda

- What are the limitations of IC and IA learning?
- Is robustness at odds with incentives?
- Connections to interpretability?