Incentive Compatible Regression Learning

Ofer Dekel[1] **Felix Fischer**[2] Ariel D. Procaccia[1]

[1]School of Computer Science and Engineering
The Hebrew University of Jerusalem

[2]Institut für Informatik
Ludwig-Maximilians-Universität München

19th ACM-SIAM Symposium on Discrete Algorithms

Motivation

- Goal: *learn* from information held by *strategic agents*
- Agents have different interests and different ideas of a good output
- Manipulation to improve result, leads to bias in the training data

- Machine learning and (algorithmic) mechanism design
- Two *interrelated* issues: *sampling* and *manipulation*
- This talk: regression learning

Outline

Regression Learning

- *Input space* $\mathcal{X}$
- *Target function* $o : \mathcal{X} \to \mathbb{R}$, distribution ρ over $\mathcal{X}$ (both unknown)
- *Hypothesis space* $\mathcal{F}$ of functions $f : \mathcal{X} \to \mathbb{R}$
- Loss function $\ell : \mathbb{R}^2 \to \mathbb{R}$
  - Absolute loss $\ell(a, b) = |a - b|$
  - Squared loss $\ell(a, b) = (a - b)^2$
- Risk (disutility) $R(f) = \mathbb{E}_{\mathbf{x}\sim\rho}[\ell(f(\mathbf{x}), o(\mathbf{x}))]$ associated with $f \in \mathcal{F}$
- Regression learning

  Given training set $S = \{(\mathbf{x}_j, o(\mathbf{x}_j))\}_{j=1,\ldots,m}$, $\mathbf{x}_j$ sampled i.i.d. from ρ, find $h \in \operatorname{argmin}_{f\in\mathcal{F}} R(f)$

Regression Learning with Strategic Agents

- *Input space* $\mathcal{X}$
- Set N of strategic agents
- *Target functions* $o_i : \mathcal{X} \to \mathbb{R}$, $i \in N$
- Distributions ρ_i over $\mathcal{X}$, $i \in N$
- *Hypothesis space* $\mathcal{F}$
- Risk $R_i(f) = \mathbb{E}_{\mathbf{x}\sim\rho_i}[\ell(f(\mathbf{x}), o_i(\mathbf{x}))]$
- Goal: minimize $\mathbb{E}_J[R_J(h)]$, where J is a random variable distributed uniformly over N (*i.e.*, maximize social welfare)

- Training set?

The Learning Game

- Agent i controls $\mathbf{x}_{ij}$, $j = 1, \dots, m$, sampled i.i.d. from ρ_i
- $y_{ij} = o_i(\mathbf{x}_{ij})$ is private information
- Agent i reveals $\hat{y}_{ij}$, $j = 1, \dots, m$
- True sample $S = \{S_i : i \in N\}$, $S_i = \{(\mathbf{x}_{ij}, y_{ij}) : 1 \le j \le m\}$
- Training set $\hat{S} = \{\hat{S}_i : i \in N\}$, $\hat{S}_i = \{(\mathbf{x}_{ij}, \hat{y}_{ij}) : 1 \le j \le m\}$
- Empirical Risk $\hat{R}(f, S) = \frac{1}{|S|} \sum_{(\mathbf{x},y) \in S} \ell(f(\mathbf{x}), y)$
- Empirical Risk Minimization (ERM): minimize $\hat{R}(f, \hat{S})$

- Two issues: *sampling* and *manipulation*

Mechanism Design Terminology

- ERM is
  - a *social choice function*
  - economically *efficient*, *i.e.*, maximizes social welfare
- *Mechanism*: social choice function plus a payment function
  - *strategyproof* if agent i cannot increase payoff (*i.e.*, decrease sum of risk and payment) by revealing a $\hat{y}_{ij} \neq y_{ij}$
  - *group strategyproof* if no group can (weakly) increase the payoff of all members
- *Why strategyproofness?*

  Otherwise: no well-defined input to the learning problem, no theoretical guarantees

Degenerate Distributions: ERM with Absolute Loss

- Distribution ρ_i degenerate at $\mathbf{x}_i$
- Agent i holds $y_i = o_i(\mathbf{x}_i)$ and reveals $\hat{y}_i$
- $\hat{S} = \{(\mathbf{x}_i, \hat{y}_i) : i \in N\}$
- $h = \operatorname{argmin}_{f \in \mathcal{F}} \hat{R}(f, \hat{S}) = \operatorname{argmin}_{f \in \mathcal{F}} \sum_{i \in N} \ell(f(\mathbf{x}_i), \hat{y}_i)$
- Agent i incurs cost $R_i(h) = \hat{R}(h, S_i) = \ell(h(\mathbf{x}_i), y_i)$

- **Theorem:** If ℓ is the *absolute loss* and $\mathcal{F}$ is convex, then ERM is group strategyproof.

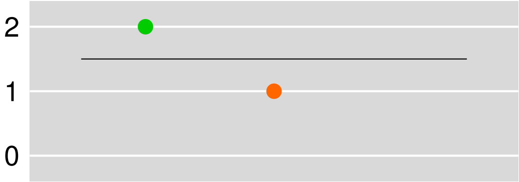
  Actually: if we don't get the (real) best fit, then somebody must have lied and lost

ERM with Superlinear Loss

- **Theorem:** If ℓ is "superlinear", $\mathcal{F}$ is convex, not "full" on $\mathbf{x}_1, \ldots . \mathbf{x}_n$, and contains at least two functions, then ERM is *not* strategyproof.

ERM with Superlinear Loss

- **Theorem:** If ℓ is "superlinear", $\mathcal{F}$ is convex, not "full" on $\mathbf{x}_1, \ldots .\mathbf{x}_n$, and contains at least two functions, then ERM is *not* strategyproof.
- Example: $\mathcal{X} = \mathbb{R}$, $\mathcal{F}$ are the constant functions, $\ell(a, b) = (a - b)^2$, $N = \{1, 2\}$

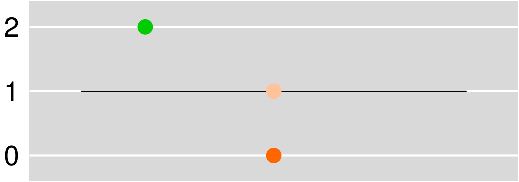


ERM with Superlinear Loss

- **Theorem:** If ℓ is "superlinear", $\mathcal{F}$ is convex, not "full" on $\mathbf{x}_1, \ldots . \mathbf{x}_n$, and contains at least two functions, then ERM is *not* strategyproof.
- Example: $\mathcal{X} = \mathbb{R}$, $\mathcal{F}$ are the constant functions, $\ell(a, b) = (a - b)^2$, $N = \{1, 2\}$

ERM with Superlinear Loss

- **Theorem:** If ℓ is "superlinear", $\mathcal{F}$ is convex, not "full" on $\mathbf{x}_1, \ldots .\mathbf{x}_n$, and contains at least two functions, then ERM is *not* strategyproof.
- Example: $\mathcal{X} = \mathbb{R}$, $\mathcal{F}$ are the constant functions, $\ell(a, b) = (a - b)^2$, $N = \{1, 2\}$



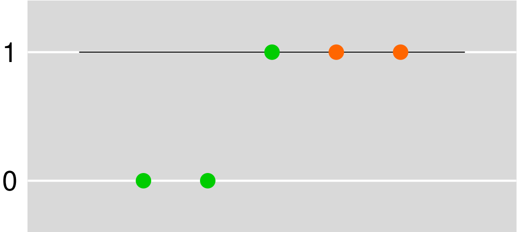
Uniform Distributions

- Distribution ρ_i discrete, uniform over $\{\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im}\}$
- Agent i holds $y_{ij} = o_i(\mathbf{x}_{ij})$ and reveals $\hat{y}_{ij}$, $j = 1, \ldots, m$
- ERM computes $h = \operatorname{argmin}_{f \in \mathcal{F}} \hat{R}(h, \hat{S})$
- Agent i incurs cost $R_i(h) = \hat{R}(h, S_i) = \frac{1}{m} \sum_{j=1}^{m} \ell(h(\mathbf{x}_{ij}), y_{ij})$

Uniform Distributions

- Distribution ρ_i discrete, uniform over $\{\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im}\}$
- Agent i holds $y_{ij} = o_i(\mathbf{x}_{ij})$ and reveals $\hat{y}_{ij}$, $j = 1, \ldots, m$
- ERM computes $h = \text{argmin}_{f \in \mathcal{F}} \hat{R}(h, \hat{S})$
- Agent i incurs cost $R_i(h) = \hat{R}(h, S_i) = \frac{1}{m} \sum_{j=1}^{m} \ell(h(\mathbf{x}_{ij}), y_{ij})$

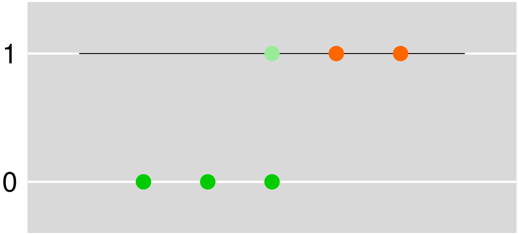
- ERM with absolute loss no longer strategyproof



Uniform Distributions

- Distribution ρ_i discrete, uniform over $\{\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im}\}$
- Agent i holds $y_{ij} = o_i(\mathbf{x}_{ij})$ and reveals $\hat{y}_{ij}$, $j = 1, \ldots, m$
- ERM computes $h = \text{argmin}_{f \in \mathcal{F}} \hat{R}(h, \hat{S})$
- Agent i incurs cost $R_i(h) = \hat{R}(h, S_i) = \frac{1}{m} \sum_{j=1}^{m} \ell(h(\mathbf{x}_{ij}), y_{ij})$

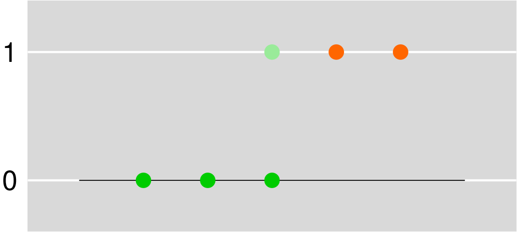
- ERM with absolute loss no longer strategyproof



Uniform Distributions

- Distribution ρ_i discrete, uniform over $\{\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im}\}$
- Agent i holds $y_{ij} = o_i(\mathbf{x}_{ij})$ and reveals $\hat{y}_{ij}$, $j = 1, \ldots, m$
- ERM computes $h = \text{argmin}_{f \in \mathcal{F}} \hat{R}(h, \hat{S})$
- Agent i incurs cost $R_i(h) = \hat{R}(h, S_i) = \frac{1}{m} \sum_{j=1}^{m} \ell(h(\mathbf{x}_{ij}), y_{ij})$

- ERM with absolute loss no longer strategyproof



VCG to the Rescue

- General Mechanism due to Vickrey (1961), Clarke (1971), and Groves (1973)
- Perform ERM (recall: it is economically efficient)
- Agent i pays $\sum_{j \neq i} \hat{R}_j(h, \hat{S})$,

  incurs total cost $\hat{R}(h, S_i) + \sum_{j \neq i} \hat{R}(h, \hat{S}_j)$
- Good news: strategyproof for any loss function
- Bad news
  - in general: not group strategyproof, payments problematic
  - in our case: payments not bounded, no obvious way to ensure individual rationality

- Is there a (group) strategyproof mechanism without payments?

Mechanisms without Payments

- Absolute loss function
- Relax efficiency requirement, consider α-*efficient* mechanisms

- **Theorem (upper bound):** There exists a 3-efficient, group strategyproof mechanism for constant functions over $\mathbb{R}^k$ and homogeneous linear functions over $\mathbb{R}$

Mechanisms without Payments

- Absolute loss function
- Relax efficiency requirement, consider α-*efficient* mechanisms

- **Theorem (upper bound):** There exists a 3-efficient, group strategyproof mechanism for constant functions over $\mathbb{R}^k$ and homogeneous linear functions over $\mathbb{R}$
- **Theorem (lower bound):** There is no $(3-\varepsilon)$-efficient, strategyproof mechanism for constant or homogeneous linear functions over $\mathbb{R}^k$ for any k and any $\varepsilon > 0$

Mechanisms without Payments

- Absolute loss function
- Relax efficiency requirement, consider α-*efficient* mechanisms

- **Theorem (upper bound):** There exists a 3-efficient, group strategyproof mechanism for constant functions over $\mathbb{R}^k$ and homogeneous linear functions over $\mathbb{R}$
- **Theorem (lower bound):** There is no $(3 - \varepsilon)$-efficient, strategyproof mechanism for constant or homogeneous linear functions over $\mathbb{R}^k$ for any k and any $\varepsilon > 0$

- **Conjecture:** There is no α-efficient, strategyproof mechanism without payments for homogeneous linear functions over $\mathbb{R}^k$ for $k \geq 2$ and *any* α

Generalization

- Assume that for all $f \in \mathcal{F}$,
  (a). for all $i \in N$, $|\hat{R}_i(f, S) - R_i(f)| \leq \frac{\varepsilon}{2}$ and
  (b). $|\hat{R}(f, S) - \frac{1}{n}\sum_{i \in N} R_i(f)| \leq \frac{\varepsilon}{2}$
- Then the following holds for any mechanism:
  - If (group) strategyproof under uniform distributions, then ε-(group) strategyproof under general distributions
  - If α-efficient under uniform distributions, then α-efficient under general distributions up to an additive factor of ε
- If $\mathcal{F}$ has bounded complexity and sample size is $\Theta(\frac{\log(1/\delta)}{\varepsilon^2})$, then (a) holds with probability $1 - \delta$
- (b) holds if (a) holds for all i, increasing the sample size by a $\log |N|$ factor

Discussion

- For m large enough, any loss function, any function class: VCG is ε-strategyproof with probability $1 - \delta$
- For m large enough, absolute loss, constant functions: there exists a mechanism without payments that is ε-group strategyproof and 3-efficient (up to additive ε)
- Future work:
  - Settle impossibility conjecture for homogeneous linear functions over $\mathbb{R}^k$
  - Extend to other settings, e.g. classification

Thank you for your attention!