Strategyproof Linear Regression in High Dimensions: An Overview

YILING CHEN
Harvard University
and
CHARA PODIMATA
Harvard University
and
ARIEL D. PROCACCIA
Carnegie Mellon University
and
NISARG SHAH
University of Toronto

In this letter, we outline some of the results from our recent work, which is part of an emerging line of research at the intersection of machine learning and mechanism design aiming to avoid noise in training data by correctly aligning the incentives of data sources. Specifically, we focus on the ubiquitous problem of *linear regression*, where *strategyproof* mechanisms have previously been identified in two dimensions. In our setting, agents have single-peaked preferences and can manipulate only their response variables. Our main contribution is the discovery of a family of *group strategyproof* linear regression mechanisms in any number of dimensions, which we call *generalized resistant hyperplane* mechanisms. The game-theoretic properties of these mechanisms — and, in fact, their very existence — are established through a connection to a discrete version of the Ham Sandwich Theorem.

Categories and Subject Descriptors: I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems; J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Economics*

General Terms: Algorithms, Economics, Theory

1. INTRODUCTION

Even for the most powerful machine learning algorithms, the quality of their learned models is dependent upon the *quality* of the training data. This dependency has given rise to the study of machine learning algorithms that are *robust* to noise in the training data. A large body of work addresses stochastic noise, while on the other extreme, another branch of the literature focuses on adversarial noise, where errors are introduced by an adversary with the explicit purpose of sabotaging the algorithm. The latter approach is often too pessimistic, and generally leads to negative results.

More recently, some researchers have taken a game-theoretic viewpoint suggesting

Authors' addresses: yiling@seas.harvard.edu, podimata@g.harvard.edu, arielpro@cs.cmu.edu, nisarg@cs.toronto.edu

a model of *strategic noise* that can be seen as occupying the middle ground of noise models. Specifically, training data is provided by strategic sources — hereinafter *agents* — that may intentionally introduce errors *to maximize their own benefit*. Compared to adversarial noise, the advantage of this model (when its underlying assumptions hold true) is that, if we aligned the agents' incentives correctly, it would be possible to obtain uncontaminated data. From this viewpoint, the ideal is the design of learning algorithms that, in addition to being statistically efficient, are *strategyproof*, i.e., where supplying pristine data is a dominant strategy for each agent. Subscribing to this agenda, in our recent work [Chen et al. 2018] we analyzed strategyproof mechanisms for high dimensional linear regression.

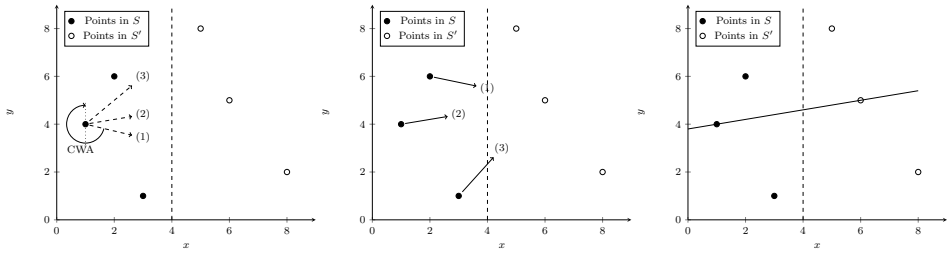
But when does this type of strategic regression problem arise? Dekel et al. [2010] give the real-world example of the global fashion chain Zara, whose distribution process relies on regression [Caro and Gallien 2010]. Specifically, the demand for each product at each store is predicted based on historical data, as well as information provided by store managers. Since the supply of popular items is limited, store managers may strategically manipulate requested quantities so that the output of the regression process would better fit their needs, and, indeed, there is ample evidence that many of them have done so [Caro et al. 2010]. More generally, as discussed in detail by Perote and Perote-Pena [2004], this type of setting is relevant whenever "data could come from surveys composed by agents interested in not being perceived as real outliers if the estimation results could be used in the future to change the economic situation of the agents that generate the sample."

Before we move on to presenting our results, we remark that the research agenda of machine learning algorithms that are robust to strategic noise can be described using three key axes: (i) manipulable information (i.e., whether the dependent variables are private information [Dekel et al. 2010; Meir et al. 2012] or the independent variables are [Hardt et al. 2016; Dong et al. 2017]), (ii) goal of the agents (i.e., whether they are motivated by privacy concerns [Cummings et al. 2015; Cai et al. 2015] or by accurate assessments of the algorithm on their own sample [Dekel et al. 2010]) and (iii) potential use of payments [Cai et al. 2015] and other incentive guarantees. In our work, dependent variables are private information, agents wish to make the regression accurate on their true datapoint, and our goal is to incentivize agents to report truthfully without the use of monetary payments.

2. MODEL

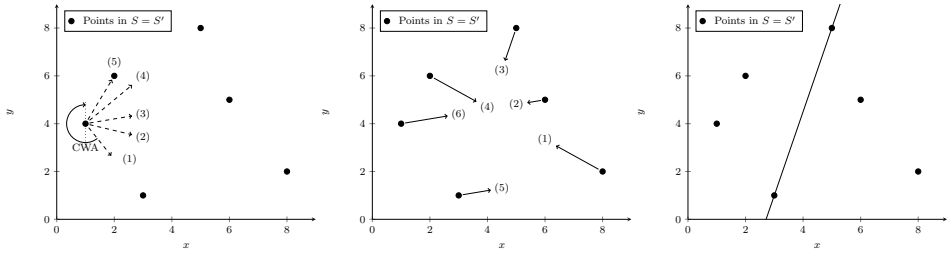
Given a natural number $k \in \mathbb{N}$, let $[k] = \{1, \dots, k\}$. Given a set of numbers $\{a_i : i \in T\}$ and $k \in [|T|]$, let $\min^k_{i \in T} x_i$ be the k^{th} smallest number in the set.

Let N be the set of agents. Each agent $i \in N$ controls one datapoint $(\boldsymbol{x}_i, y_i)$, where $\boldsymbol{x}_i \in \mathbb{R}^d$ is *publicly verifiable* while $y_i \in \mathbb{R}$ is *private* to the agent. Agent i reports $(\boldsymbol{x}_i, \widetilde{y}_i)$ to the mechanism, where potentially $\widetilde{y}_i \neq y_i$. The mechanism outputs a hyperplane $M^{\boldsymbol{x}}(\widetilde{\boldsymbol{y}}) = (\boldsymbol{\beta}_1, \beta_0) \in \mathbb{R}^{d+1}$, where $\widetilde{\boldsymbol{y}} = (\widetilde{y}_i)_{i \in N}$ and by slightly abusing notation $\boldsymbol{x} = (\boldsymbol{x}_i)_{i \in N}$. The *outcome* for agent i is $\widehat{y}_i = \langle \boldsymbol{\beta}_1, \boldsymbol{x}_i \rangle + \beta_0$, and the *residual* of agent i is $r_i = \widehat{y}_i - y_i$. Agent i has *single-peaked* preferences over $\widehat{y}_i$, with a peak at y_i. Formally, this means that given any $a \geq b > y_i$ or $a \leq b < y_i$, agent i strongly prefers y_i to b and weakly prefers b to a. We wish to design mechanisms that are *strategyproof* (in which no *individual* agent can benefit from



(a) Computing CWA (b) Median-of-medians (c) Final line

Fig. 1: The CRM mechanism when S and S' are separable.



(a) Computing CWA (b) Median-of-medians (c) Final line

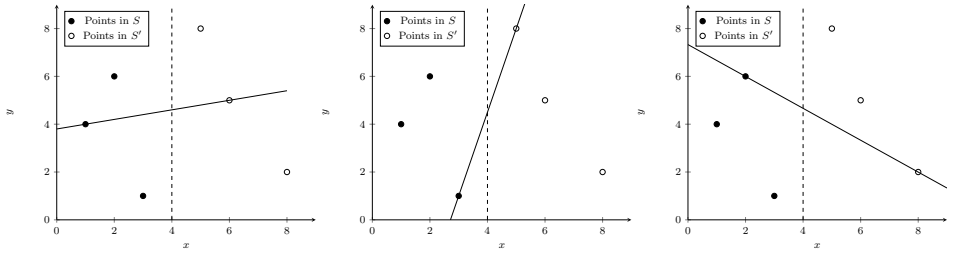
Fig. 2: The CRM mechanism when $S = S'$.

misreporting regardless of the reports of other agents) or *group-strategyproof* (in which no *coalition* of agents can benefit from simultaneously misreporting regardless of the reports of other agents).

Two notes are in order here. First, we make no assumptions on the data generation process (e.g., whether the data comes from a distribution). Second, strategyproofness or group-strategyproofness are not the sole desiderata; a constant function (e.g., the flat hyperplane $y = 0$) is group-strategyproof, but not necessarily desirable. We would like our mechanisms to also have good statistical efficiency.

3. STRATEGYPROOF MECHANISMS

CRM mechanisms. Perote and Perote-Pena [2004] were the first to study strategyproof mechanisms for simple linear regression (i.e., where $d = 1$) in the foregoing setting. They proposed a novel family of (allegedly) strategyproof mechanisms, which they termed *Clockwise Repeated Median (CRM)* estimators. This family is parametrized by two subsets of agents $S, S' \subseteq N$. These subsets are chosen based on the public information $\boldsymbol{x}$, and therefore can be treated as fixed. Informally, given $S, S' \subseteq N$, the (S, S')-CRM mechanism first computes the median *clockwise angle* (CWA), with respect to the y axis, from each point $i \in S$ to points in S' (Fig. 1a and 2a). Then, it chooses the point $i^* \in S$ whose median CWA is the median of the median CWAs from all points in S (Fig. 1b and 2b). If the median CWA from point i^* is towards point $j^* \in S'$, the mechanism returns the straight line passing through points i^* and j^* (Fig. 1c and 2c).



(a) $(S, S', 2, 2)$-GRL (b) $(S, S', 1, 3)$-GRL (c) $(S, S', 3, 1)$-GRL

Fig. 3: Examples of (S, S', k, k')-GRL mechanisms.

Perote and Perote-Pena [2004] claimed that the (S, S')-CRM mechanism is strategyproof when $S \subseteq S'$ or $S \cap S' = \emptyset$. However, after identifying a mistake in their proof, we are able to recover strategyproofness (and in fact, establish group-strategyproofness) of a more restricted family of CRM mechanisms, namely, when a) $S = S'$, b) S and S' are *separable* (i.e., when the two sets can be separated by a vertical line), or c) $S \cap S' = \emptyset$ and $\min(|S|, |S'|) = 1$. Given this correction, one might wonder: *Can we generalize the corrected CRM family to high dimensions? What would be a generalization of the clockwise angle?* We provide such a generalization by first generalizing the geometric CRM family to a more algebraic family of *Generalized Resistant Line* (GRL) mechanisms in two dimensions, and then generalizing GRL mechanisms to high dimensions.

Generalizing in two dimensions. While sets S and S' in our corrected CRM family are not separable in two of three cases, our generalization below uses only separable sets S and S', and yet incorporates all three cases of our CRM family.

***Definition* 3.1** (*Generalized Resistant Line (GRL) Mechanisms)*. Given separable sets $S, S' \subseteq N$, $k \in [|S|]$, and $k' \in [|S'|]$, the (S, S', k, k')-generalized resistant line (GRL) mechanism returns the line $\boldsymbol{\beta} = (\beta_1, \beta_0)$ given by

$$\min^{k}_{i \in S} y_i - \beta_1 x_i - \beta_0 = \min^{k'}_{j \in S'} y_j - \beta_1 x_j - \beta_0 = 0. \tag{1}$$

In words, the GRL mechanism returns the line which makes both the k^{th} minimum residual from S and the $(k')^{\text{th}}$ minimum residual from S' zero (see Fig. 3 for some examples). This family owes its name to the fact that it is a direct generalization of the *resistant line* mechanisms [Johnstone and Velleman 1985] that were proposed in the statistics literature as robust-to-outliers methods. These methods make the *median* residuals from S and S' zero (i.e., use $k = \lceil |S|/2 \rceil, k' = \lceil |S'|/2 \rceil$).

We show that GRL mechanisms are well-defined as Equation (1) is guaranteed to have a unique solution; they include our corrected family of CRM mechanisms; and every GRL mechanism is group-strategyproof.

Existence of a unique solution to Equation (1) in two dimensions uses the separability of S and S'. While this equation naturally generalizes to high dimensions, it is not immediately clear what conditions would be required to ensure a unique outcome. This is where the literature on the *Ham Sandwich Theorem* comes to the rescue.

Generalizing to higher dimensions. Given a hyperplane H, let H^+ and H^- be its positive and negative closed half-spaces, respectively. A basic version of the Ham Sandwich theorem due to Stone and Tukey [1942] states that given k continuous measures $\mu_1, \ldots, \mu_k$ on $\mathbb{R}^k$, there exists a hyperplane H such that $\mu_i(H^+) = 1/2$ for each $i \in [k]$. A discrete version of this result due to Elton and Hill [2011] states that given k finite sets $S_1, \ldots, S_k \subseteq \mathbb{R}^k$, there exists a hyperplane H such that for each $i \in [k]$, $H \cap S_i \neq \emptyset$ and H "bisects" S_i (i.e., $\min(|H^+ \cap S_i|, |H^- \cap S_i|) \geq \lceil |S|/2 \rceil$).

For linear regression, this implies that given $S_1, \ldots, S_{d+1} \subseteq N$, there exists a "resistant hyperplane" which makes the median residual from S_t zero, for each $t \in [d+1]$. While this seems like a natural generalization of GRL mechanisms, it is easy to check that a) such a hyperplane is not always unique, and b) the existence is not guaranteed if median is replaced by other percentiles.[1]

Steiger and Zhao [2010], building upon previous work [Bárány et al. 2008; Breuer 2010], provide a generalization that *almost* perfectly fits our needs. They show that under a certain separability condition on $S_1, \ldots, S_{d+1}$ and a mild assumption, there exists a unique hyperplane H which contains a prescribed number of points from each set in its negative closed half-space. However, their condition uses the private information $\boldsymbol{y}$, whereas we would like sets $S_1, \ldots, S_{d+1}$ to be defined based only on the public information $\boldsymbol{x}$ for our game-theoretic desiderata. We provide such a condition which still yields a unique hyperplane as well as group-strategyproofness. We also eliminate the mild assumption imposed by Steiger and Zhao [2010]. Our results closely mirror but do not make use of the results of Steiger and Zhao [2010].

***Definition* 3.2** (*Generalized Resistant Hyperplane (GRH) Mechanisms)*. We say that a family $\mathcal{S} = (S_1, \ldots, S_{d+1})$ of nonempty pairwise disjoint subsets of N is *publicly separable* if for any $I, J \subseteq [d+1]$, there exists a hyperplane in $\mathbb{R}^d$ separating $\bigcup_{t \in I} \{\boldsymbol{x}_i : i \in S_t\}$ from $\bigcup_{t \in J} \{\boldsymbol{x}_i : i \in S_t\}$. Given a publicly separable family $\mathcal{S} = (S_1, \ldots, S_{d+1})$ of subsets of agents, and $\boldsymbol{k} = (k_1, \ldots, k_{d+1})$ with $k_t \in [|S_t|]$ for $t \in [d+1]$, the $(\mathcal{S}, \boldsymbol{k})$-generalized resistant hyperplane (GRH) mechanism returns a hyperplane $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \beta_0)$ such that for each $t \in [d+1]$,

$$\min_{i \in S_t}^{k_t} y_i - \langle \boldsymbol{\beta}_1, \boldsymbol{x}_i \rangle - \beta_0 = 0. \tag{2}$$

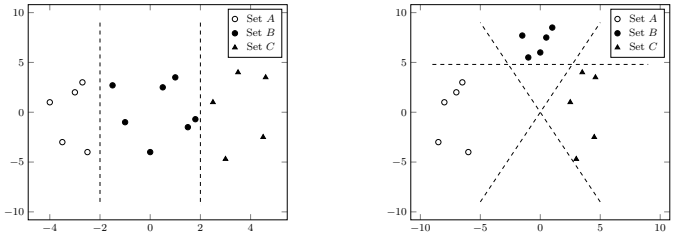
That is, it makes the k_t^{th} smallest residual from each set $S_t \in \mathcal{S}$ zero.

Fig. 4 provides a pictorial intuition of public separability in two dimensions. As mentioned above, we show that GRH mechanisms are not only *well defined* (i.e., Equation (2) has a unique solution), they are also *group-strategyproof*. It is easy to check that in two dimensions, they coincide with GRL mechanisms.

Another family of mechanisms in high dimensions. Prior to our work, the only known (non-trivial) strategyproof mechanism for linear regression in high dimensions was due to Dekel et al. [2010], who proved that the *empirical risk minimizer (ERM)* of the L_1 loss[2] — hereinafter, the L_1-ERM — is group-strategyproof. We generalize this family in two ways: we allow a weighted L_1 loss, in which the loss

[1] Recall that even in two dimensions, we needed an additional condition, namely separability of S and S' by a vertical line.

[2] Formally, this mechanism finds a hyperplane $\boldsymbol{\beta}$ minimizing the L_1 loss $\sum_{i \in N} |y_i - \langle \boldsymbol{\beta}_1, \boldsymbol{x}_i \rangle - \beta_0|$, and breaks any ties by minimizing the norm of $\boldsymbol{\beta}$.



(a) Not separable (b) Separable

Fig. 4: Pictorial intuition for publicly separable family of sets of agents. The figures depict the case where $d = 2$. Public information $\boldsymbol{x}_i$ of each agent i is plotted. Fig. 4a shows a case where sets A, B and C are not publicly separable because no line can separate sets B and C from set A. Fig. 4b shows a case where the three sets are publicly separable.

to each agent i is multiplied by a weight $w_i^{\boldsymbol{x}}$, and we allow adding a convex function of $\boldsymbol{\beta}$ to the loss function, often known as regularization in the machine learning literature. Our mechanism still breaks any ties by minimizing the norm of $\boldsymbol{\beta}$. We show that every mechanism in this generalized family is still group-strategyproof.

Strategyproofness versus group-strategyproofness. All mechanisms for linear regression we outlined so far are group-strategyproof. One might wonder if there exist strategyproof mechanisms that are not group-strategyproof. For $d = 0$, Moulin [1980] proved that every strategyproof mechanism is also group-strategyproof. Interestingly, this does not hold for linear regression in two or more dimensions. A simple counterexample is the mechanism for two agents in two dimensions which returns the line passing through (x_1, y_2) and (x_2, y_1). The mechanism is trivially strategyproof because the residual of each agent is independent of the agent's report; this condition is known as *impartiality* in the literature. With more agents, it is not clear if impartial mechanisms exist. We show that there exists a large family of impartial mechanisms for any number of agents and in any dimension, and that all non-trivial impartial mechanisms violate group-strategyproofness.

Efficiency versus strategyproofness. The most popular mechanism for linear regression is the *Ordinary Least Squares* (OLS), which returns the hyperplane $\boldsymbol{\beta}$ minimizing the squared L_2 loss $\sum_{i \in N}(y_i - \langle \boldsymbol{\beta}_1, \boldsymbol{x}_i \rangle - \beta_0)^2$. The Gauss-Markov theorem establishes the OLS as the most efficient mechanism under mild assumptions, but it is known that the OLS is not strategyproof [Dekel et al. 2010]. This raises an important question: *Can we design strategyproof mechanisms that are arbitrarily close to the OLS?* We answer this *negatively* by showing that every strategyproof mechanism can cause a squared L_2 loss at least twice that of the OLS, in the worst case over inputs.

4. CONCLUSION AND OPEN QUESTIONS

Our work leaves several directions for future research. Perhaps the most ambitious direction is to fully characterize strategyproof (or group-strategyproof) mechanisms for linear regression, which might help us analytically find the *most efficient* strategyproof mechanism. In one dimension, such an analysis was done by Caragiannis et al. [2016] using the characterization result of Moulin [1980].

It is also interesting to consider generalizations of our model where each agent controls multiple data points [Dekel et al. 2010], or where only a small subset of data points are manipulated by strategic agents but the mechanism does not know which data points are manipulated [Charikar et al. 2017]. We hope that our work, which takes a step forward in developing a *theory of incentives in machine learning* [Procaccia 2008], can serve as a stepping stone to studying incentives in more realistic environments. With machine learning algorithms increasingly being used to make real-world decisions, we should be especially careful about the possibility of strategic manipulation leading the algorithm astray.

REFERENCES

Bárány, I., Hubard, A., and Jerónimo, J. 2008. Slicing convex sets and measures by a hyperplane. *Discrete & Computational Geometry 39,* 1-3, 67–75.

Breuer, F. 2010. Uneven splitting of ham sandwiches. *Discrete & Computational Geometry 43,* 4, 876–892.

Cai, Y., Daskalakis, C., and Papadimitriou, C. H. 2015. Optimum statistical estimation with strategic data sources. In *28th.* 280–296.

Caragiannis, I., Procaccia, A. D., and Shah, N. 2016. Truthful univariate estimators. In *Proceedings of the 33rd International Conference on Machine Learning.* 127–135.

Caro, F. and Gallien, J. 2010. Inventory management of a fast-fashion retail network. *Operations Research 58,* 2, 257–273.

Caro, F., Gallien, J., Miranda, M. D., Torralbo, J. C., Corras, J. M. C., Vazquez, M. M., Calamonte, J. A. R., and Correa, J. 2010. Zara uses operations research to reengineer its global distribution process. *Interfaces 40,* 1, 71–84.

Charikar, M., Steinhardt, J., and Valiant, G. 2017. Learning from untrusted data. In *Proceedings of the 49th ACM Symposium on Theory of Computing.* 47–60.

Chen, Y., Podimata, C., Procaccia, A. D., and Shah, N. 2018. Strategyproof linear regression in high dimensions. In *Proceedings of the 2018 ACM Conference on Economics and Computation.* ACM, 9–26.

Cummings, R., Ioannidis, S., and Ligett, K. 2015. Truthful linear regression. In *28th.* 448—-483.

Dekel, O., Fischer, F., and Procaccia, A. D. 2010. Incentive compatible regression learning. *Journal of Computer and System Sciences 76,* 8, 759–777.

Dong, J., Roth, A., Schutzman, Z., Waggoner, B., and Wu, Z. S. 2017. Strategic classification from revealed preferences. arXiv:1710.07887.

Elton, J. H. and Hill, T. P. 2011. A stronger conclusion to the classical ham sandwich theorem. *European Journal of Combinatorics 32,* 5, 657–661.

Hardt, M., Megiddo, N., Papadimitriou, C. H., and Wootters, M. 2016. Strategic classification. In *7th.* 111–122.

Johnstone, I. M. and Velleman, P. F. 1985. The resistant line and related regression methods. *Journal of the American Statistical Association 80,* 392, 1041–1054.

Meir, R., Procaccia, A. D., and Rosenschein, J. S. 2012. Algorithms for strategyproof classification. *Artificial Intelligence 186*, 123–156.

Moulin, H. 1980. On strategy-proofness and single-peakedness. *Public Choice 35*, 437–455.

Perote, J. and Perote-Pena, J. 2004. Strategy-proof estimators for simple regression. *Mathematical Social Sciences 47,* 2, 153–176.

Procaccia, A. D. 2008. Towards a theory of incentives in machine learning. *ACM SIGecom Exchanges 7,* 2, 6.

Steiger, W. and Zhao, J. 2010. Generalized ham-sandwich cuts. *Discrete & Computational Geometry 44,* 3, 535–545.

Stone, A. H. and Tukey, J. W. 1942. Generalized "sandwich" theorems. *Duke Mathematical Journal 9,* 2, 356–359.