Practical Performative Policy Learning

Qianyi Chen, Bo Li

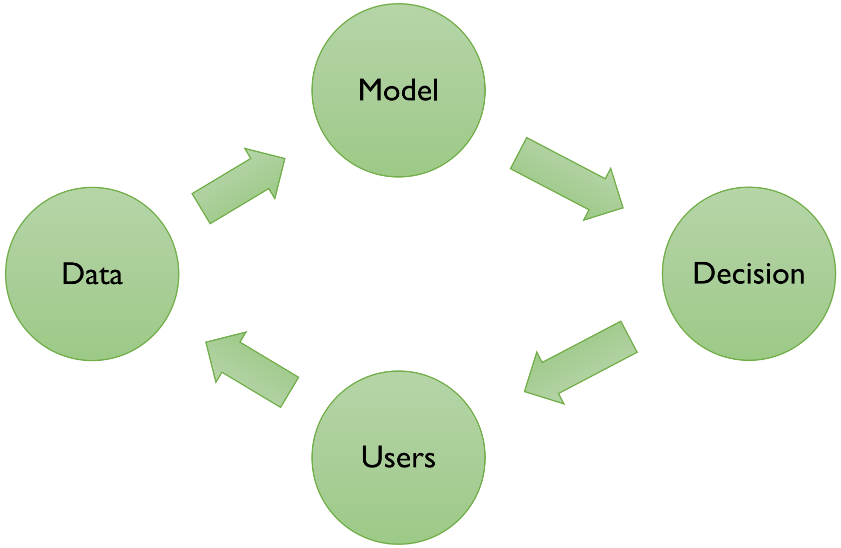
School of Economics and Management, Tsinghua University

QR code of paper

Machine Learning for Decision-Making

There are a variety of applications of machine learning used for decision-making, where users interact with the model with decisions of the ML model.

- Route Suggestion
- Ads/Coupon
- Pricing
- College Admission

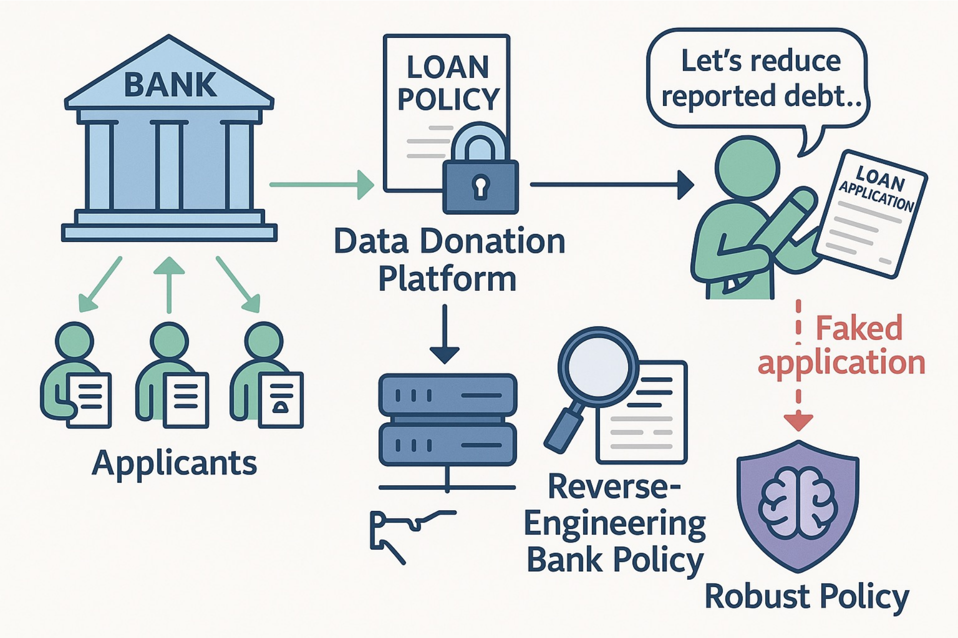


We call predictions performative when they impact the population they aim to predict.

Strategic Behavior: Faking

Example 1 (Loan application)

- Bank provides loan to applicants constantly according to its policy.
- There is crowdsource effort provides reverse-engineering of policy.
- Applicants will refer to such third-party organization to decide how to falsify the application materials, e.g. falsely report the debt.
- The target is learning a policy that's robust to such faking behavior.



Strategic Behavior: Improving

Example 2 (Loan application)

- …
- Applicants will refer to such third-party organization to decide the scheme of background improvement (e.g. sell the high-risk assets; improve liquidity) to uplift the probability of receiving the loan.

- In this context, we believe the authenticity of submitted materials. (Given a powerful checking system, we are free from falsifying.)
- The target here is learning a policy that incentivizing improvement.

Extend the Example: Impact of Prediction

Example 2 (Loan application)

- Bank may predict whether certain applicant will finally default, and charge higher interest rate for those applicants with higher default risk.
- Such high interest rate will increase the probability of default in turn.

A natural question is that how to model and tackle with such response mechanism of user?

Strategic Classification and Performative Prediction

Strategic Agents Respond to Prediction

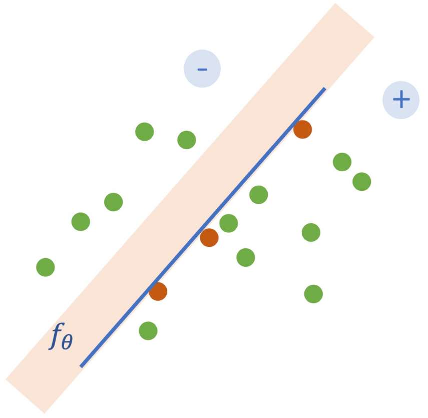
- Example of strategic classification: microfoundation of every agent

$$x(\theta) = \underset{x}{\operatorname{argmax}} \gamma f_\theta(x) - \operatorname{cost}(x_0, x)$$

Reward: probability of being classified to positive label

Cost: feature manipulation

- Distribution $D(f_\theta)$ comes from strategic behavior of individuals trying to adapt to decision rule.



Latent Assumptions on Former Example

- Rethink the strategic classification

$$x(\theta) = \underset{x}{\operatorname{argmax}} \underbrace{\gamma f_\theta(x)}_{\text{Reward of classified to positive label}} - \underbrace{\operatorname{cost}\left(x_0, x\right)}_{\text{Cost of feature manipulation}}$$

- Parametric utility function: how can we know the manipulation cost?
- Homogeneous feature manipulation cost: individual heterogeneity?
- …

Performative Prediction: Distributional Perspective

- Alternatively, performative prediction targets model the distribution shift induced by deployed model.

- Turn to consider the influence of predictive model f_θ on the whole data distribution D, which induces the distribution map $D(\theta)$.
- Example: $D(\theta) = \mathcal{N}(f(\theta), \Sigma)$

- Distribution $D(\theta)$ also comes from strategic behavior of individuals trying to adapt to decision rule.

Framework of Performative Prediction

- Risk in supervised learning:

$$\mathrm{Risk}(\theta, D) = \mathrm{E}_{z \sim D}[\ell(z; \theta)]$$

- Performative risk:

$$\mathrm{PR}(\theta) = \mathrm{Risk}(\theta, D(\theta))$$

- Performative optimality:

$$\theta_{\mathrm{PO}} \in \mathrm{argmin}_{\theta}\, \mathrm{PR}(\theta)$$

- Performative stability:

$$\theta_{\mathrm{PS}} \in \mathrm{argmin}_{\theta}\, \mathrm{Risk}\left(\theta, D\left(\theta_{\mathrm{PS}}\right)\right)$$

Framework of Performative Prediction

- Performative optimality (minimizer of performative risk)

$$\theta_{\mathrm{PO}} \in \operatorname{argmin}_\theta \mathrm{PR}(\theta)$$

- Performative stability (fixed point)

$$\theta_{\mathrm{PS}} \in \operatorname{argmin}_\theta \mathrm{Risk}\left(\theta, D\left(\theta_{\mathrm{PS}}\right)\right)$$

- Two concepts of solutions
  - Performative optimality: minimizing the risk after model deployment
  - Performative stability: a natural equilibrium notion

Retraining for Performative Stability

- Performative stability:

$$\theta_{\mathrm{PS}} \in \mathrm{argmin}_{\theta}\, \mathrm{Risk}\left(\theta, D\left(\theta_{\mathrm{PS}}\right)\right)$$

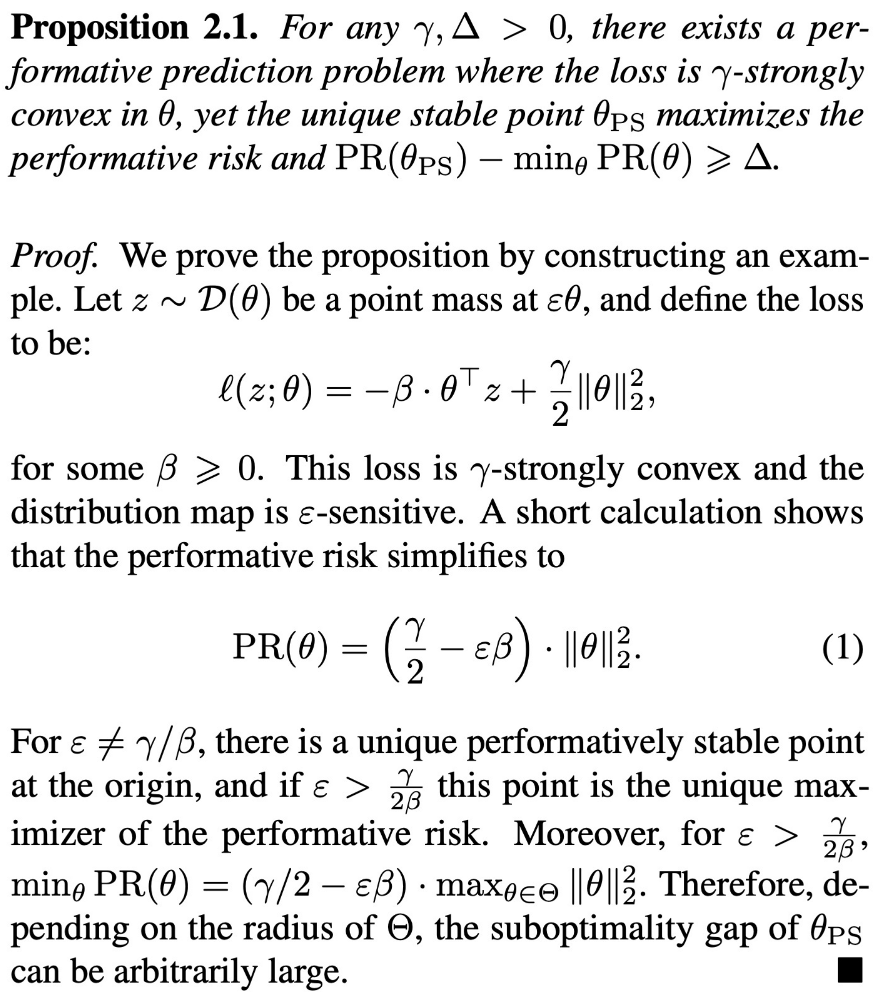
- We can reach such fixed point by repeated retraining.

- Repeated (empirical) risk minimization guarantees the convergence to stable points.

- But what we really want is performative optimality.

Gap between PS/PO Solutions Can be Large

- Here, we present a classic example (Miller et al. 2021) to show that the gap between PS and PO solution can be arbitrarily large, and PS solution is very bad.



Proposition 2.1. *For any $\gamma, \Delta > 0$, there exists a performative prediction problem where the loss is γ-strongly convex in θ, yet the unique stable point θ_{PS} maximizes the performative risk and* $\text{PR}(\theta_{\text{PS}}) - \min_\theta \text{PR}(\theta) \geqslant \Delta$.

Proof. We prove the proposition by constructing an example. Let $z \sim \mathcal{D}(\theta)$ be a point mass at $\varepsilon\theta$, and define the loss to be:

$$\ell(z; \theta) = -\beta \cdot \theta^\top z + \frac{\gamma}{2}\|\theta\|_2^2,$$

for some $\beta \geqslant 0$. This loss is γ-strongly convex and the distribution map is ε-sensitive. A short calculation shows that the performative risk simplifies to

$$\text{PR}(\theta) = \left(\frac{\gamma}{2} - \varepsilon\beta\right) \cdot \|\theta\|_2^2. \qquad (1)$$

For $\varepsilon \neq \gamma/\beta$, there is a unique performatively stable point at the origin, and if $\varepsilon > \frac{\gamma}{2\beta}$ this point is the unique maximizer of the performative risk. Moreover, for $\varepsilon > \frac{\gamma}{2\beta}$, $\min_\theta \text{PR}(\theta) = (\gamma/2 - \varepsilon\beta) \cdot \max_{\theta\in\Theta} \|\theta\|_2^2$. Therefore, depending on the radius of Θ, the suboptimality gap of θ_{PS} can be arbitrarily large. ■

Targeting Performative Optimality

- Performative optimality

$$\theta_{\mathrm{PO}} \in \operatorname{argmin}_{\theta} \mathrm{Risk}\left(\theta, D\left(\theta\right)\right)$$

- The insight of repeated retrain is aware of existence of performativity, but doesn’t model the performativity, which make it unable to break the echo chamber.

- The distance between PO and PS solution can be arbitrarily large.

Bandit Exploration

- An immediate way is Bandit-style exploration:
  - We deploy a model with parameter θ_t
  - Collect samples and calculate a performative risk $PR(\theta_t)$
  - Zero-order optimization/ bandit algorithm (e.g. zooming) is implemented to decide θ_{t+1}

- However, zero-order optimization is too inefficient for high-dimensional model parameter θ. (Imagine zero-order search with 1,000 dimensions)

Parametric Performative Gradient

- Another way to pursue performative optimality is parametric assumptions on data distribution.
- The reaction of data distribution is summarized into reaction of parameter of such distribution.

- For example, if we assume the data subjects to Gaussian distribution, then we can assume that the distribution map (data distribution after the deployment of model with parameter θ) is:

$$D(\theta) = \mathcal{N}(\mu(\theta), \Sigma)$$

Parametric Performative Gradient

- Then we can construct the performative gradient

$$\theta_{\mathrm{PO}} \in \mathrm{argmin}_\theta \, \mathrm{Risk}\left(\theta, D\left(\theta\right)\right) \qquad D(\theta) = \mathcal{N}(\mu(\theta), \Sigma)$$

- Then we can learn/estimate $\mu(\theta)$ to plug-in the gradient w.r.t. θ.

- Cons of such way:
  - Restricted to parametric assumption on distribution
  - Low sample efficiency (since the response of model deployment is still summarized as statistics, such as mean vector in this example)

From Bandit Feedback to Batch Feedback

- Both former methods compress the responses into one feedback, i.e. a performative risk/ a parameter estimate. We call this bandit feedback.
- However, we actually collect a batch of data points in each round.
- Our idea: exploiting every data point, combining the micro-level response mechanism.
- To utilize every data point, we choose to learn the post-manipulation feature distribution given the information a specific agent would use.

Make Performative Learning Practical

Augment Performative Learning with micro-level behavior learning

Beyond Prediction: Policy is What We Want

- From the angle of bank, we actually want a policy on loan assignment, and binary label is only intermediate results.

$$\pi(a \mid x) = \mathbb{P}(\text{Take action } a \mid \text{Context } x)$$

- In the literature of classical policy learning in causal inference, we usually want to maximize the policy value (binary treatment).

$$\pi^* = \arg\max_{\pi} \int \pi(x) \mathbb{E}\left[Y_i(1) - Y_i(0) \mid X_i = x\right] dF(x)$$

- Optimal policy: treat user with positive CATE $\tau(x)$

$$\pi^*(x) = 1\left(\tau(x) > 0\right)$$

Beyond Prediction: Policy is What We Want

- New policy value: distribution shift induced by strategic agent

$$V'(\pi) = \int \pi(x)\mathbb{E}\left[Y_i(1) - Y_i(0) \mid X_i^\pi = x\right] dF^\pi(x)$$

- What's the optimal policy now?

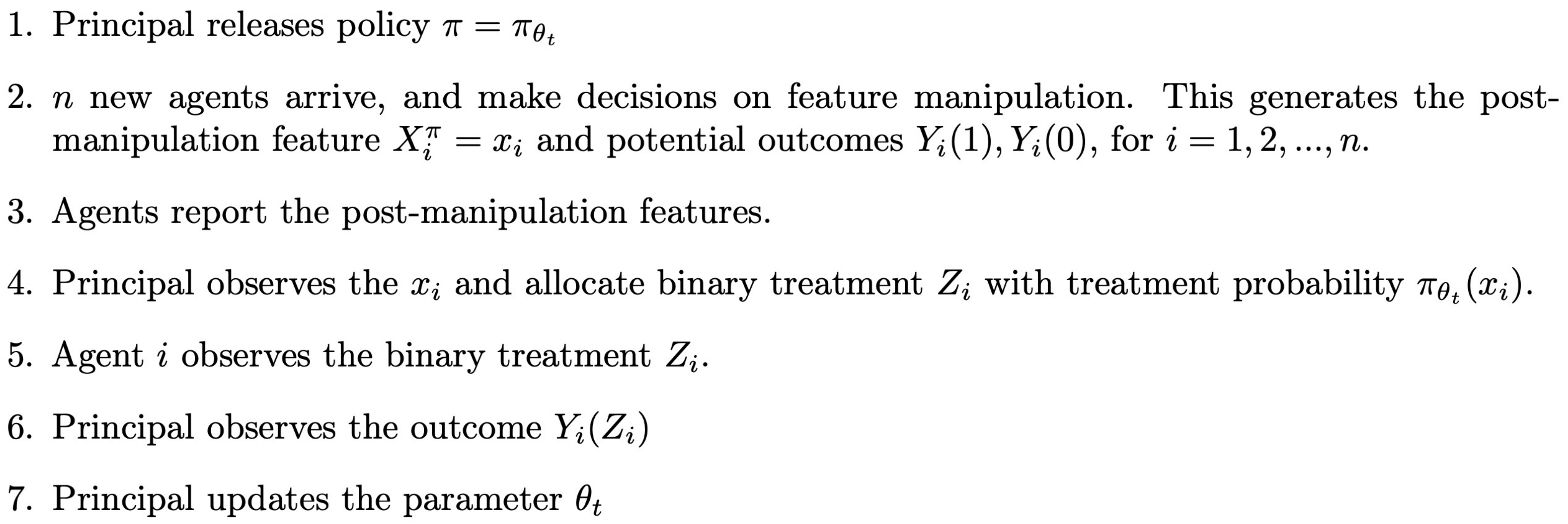
CATE changed

Covariates shift

- Munro 2025 analyze the structure of optimal policy through directional derivative of $V'(\pi)$ in π through functional analysis, and main result there is that cutoff rule isn't optimal anymore.
- We need a random policy to counter strategic agents.

Event Flow

We consider following event flow for period $t = 1,2, \ldots, T$



1. Principal releases policy $\pi = \pi_{\theta_t}$
2. n new agents arrive, and make decisions on feature manipulation. This generates the post-manipulation feature $X_i^{\pi} = x_i$ and potential outcomes $Y_i(1), Y_i(0)$, for $i = 1, 2, ..., n$.
3. Agents report the post-manipulation features.
4. Principal observes the x_i and allocate binary treatment Z_i with treatment probability $\pi_{\theta_t}(x_i)$.
5. Agent i observes the binary treatment Z_i.
6. Principal observes the outcome $Y_i(Z_i)$
7. Principal updates the parameter θ_t

CATE and Policy Value

- We define the conditional average treatment effect (CATE) here as the expected difference between treated and controlled outcomes for an individual given its submitted covariate x.

$$\tau(x) = \mathrm{E}\left[Y_i(1) - Y_i(0) \mid X_i^{\pi} = x\right]$$

- The policy value in this setting can be written as

$$\begin{aligned} V\left(\pi_\theta\right) &= \int \pi_\theta(x)\tau(x)p\left(x; \pi_\theta\right) \mathrm{d}x \\ &= \mathbb{E}_{x \sim p(\cdot; \pi_\theta)}\left[\pi_\theta(x)\tau(x)\right] \end{aligned}$$

Model π_θ as Intervention on Distribution

- One of main conceptual contributions of our work is clarifying the treatment: it's the model π_θ, instead of the parameter θ.

- The notation of distribution map $D(\theta)$ is somewhat misleading since it ignores the function form π.

- To illustrate it, we can breezily imagine that how can the parameter of a neural network own causal power on data distribution, separate from its architecture (functional form)?

Model π_θ as Intervention on Distribution

- However, such clarification brings a new challenge:
  The gradient now changes into the probability density function with regard to prediction/decision function, i.e. function-to-function gradient.
- To see this, we check the performative gradient:

$$\nabla_\theta V(\pi_\theta) = \mathbb{E}_{x \sim p(\cdot;\pi_\theta)}[\nabla_\theta \pi_\theta(x)\tau(x) + \pi_\theta(x)\tau(x)\nabla_\theta \log p(x;\pi_\theta)]$$

The core term.

- The left problem is simplifying such scenario with practical assumptions.

Bounded Rationality

- To touch the practical scenario, let’s revisit the decision process in strategic classification:

$$x(\theta) = \underset{x}{\operatorname{argmax}} \gamma f_\theta(x) - \operatorname{cost}(x_0, x)$$

Reward of classified to positive label

Cost of feature manipulation

- If f_θ is a complex function in practice, and x is high-dimensional, an individual is probably unable to implement such a complex optimization.
- What agent can do? We believe that an agent can only implement the optimization: Selecting the maximum among finite numbers!

Limited Manipulation

- Hence, we consider following scenario: there is high-dimensional feature x, but only a finite discrete type u can be manipulated (improved) by agents.

- We split the feature x into manipulatable feature u and fixed feature v.

- This split not only practical but also benefit us greatly in characterizing the impact of deployed policy.

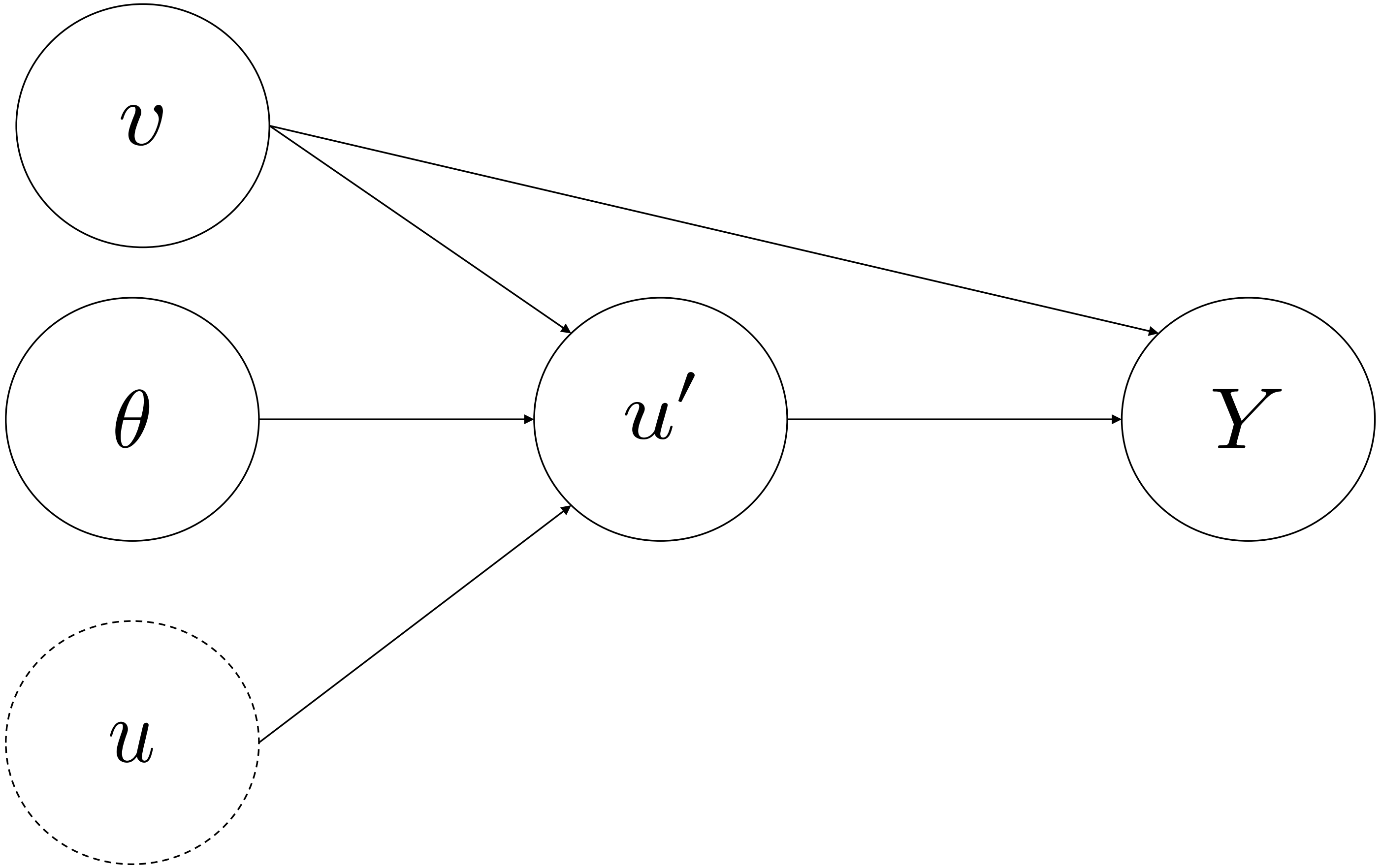
Further Motivating Limited Manipulation

Example 3 (College Admission)

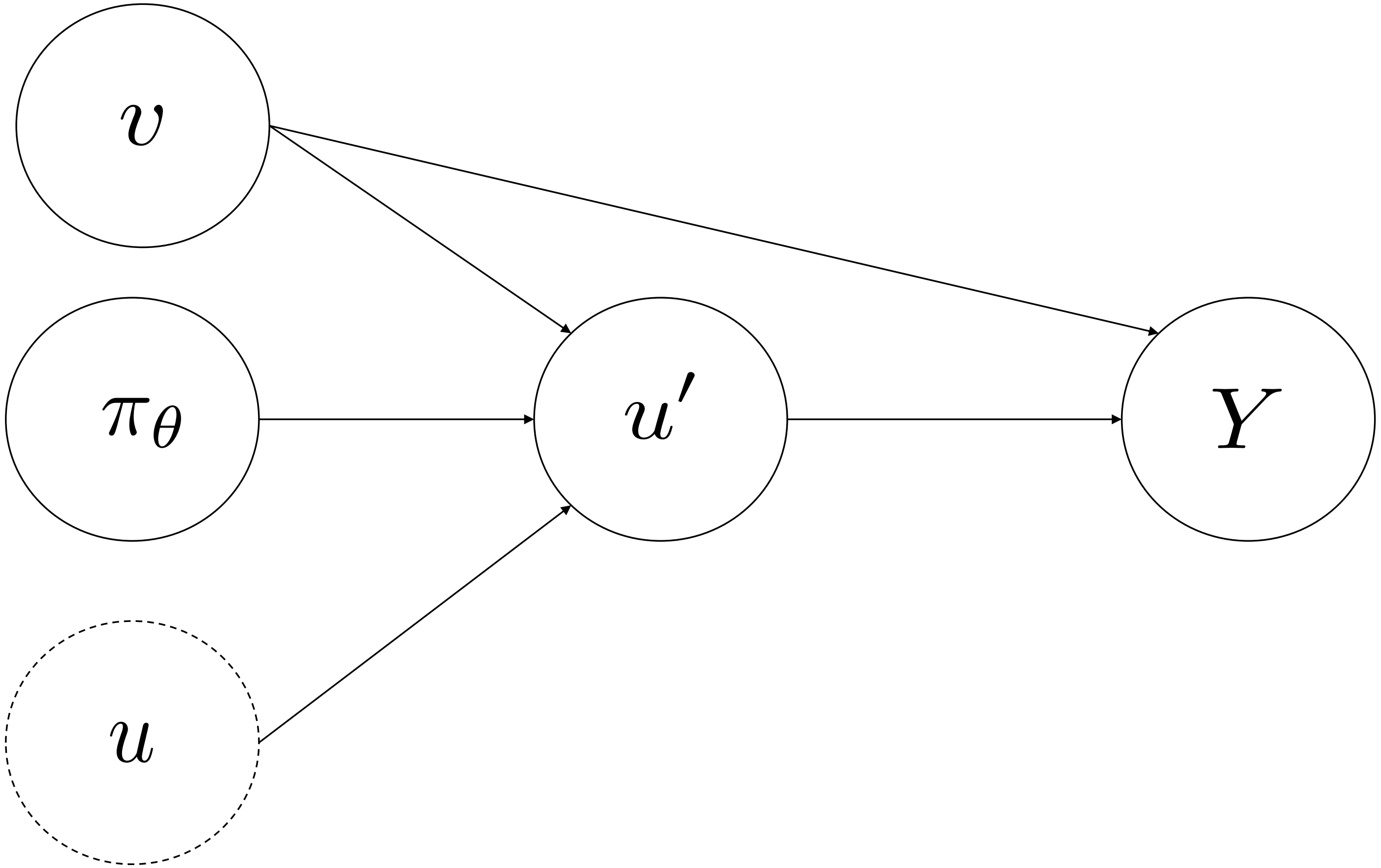
- College reveals part of their policies, e.g. higher weight on the language grades, TOEFL/GRE
- Student is informed of that at Junior year, and most of their features are fixed in the application year, e.g. school, major, demographics, grades in first two years, etc.

- The precondition is that college owns mechanism to guarantee the authenticity of submitted materials.

Traditional Perspective: θ as intervention



Our Perspective: π_θ as intervention



Abstract from Microfoundation

- We only keep the information structure in the classic strategic classification:

$$x(\theta) = \underset{x}{\operatorname{argmax}}\ \pi_\theta(x) - \operatorname{cost}(x_0, x)$$

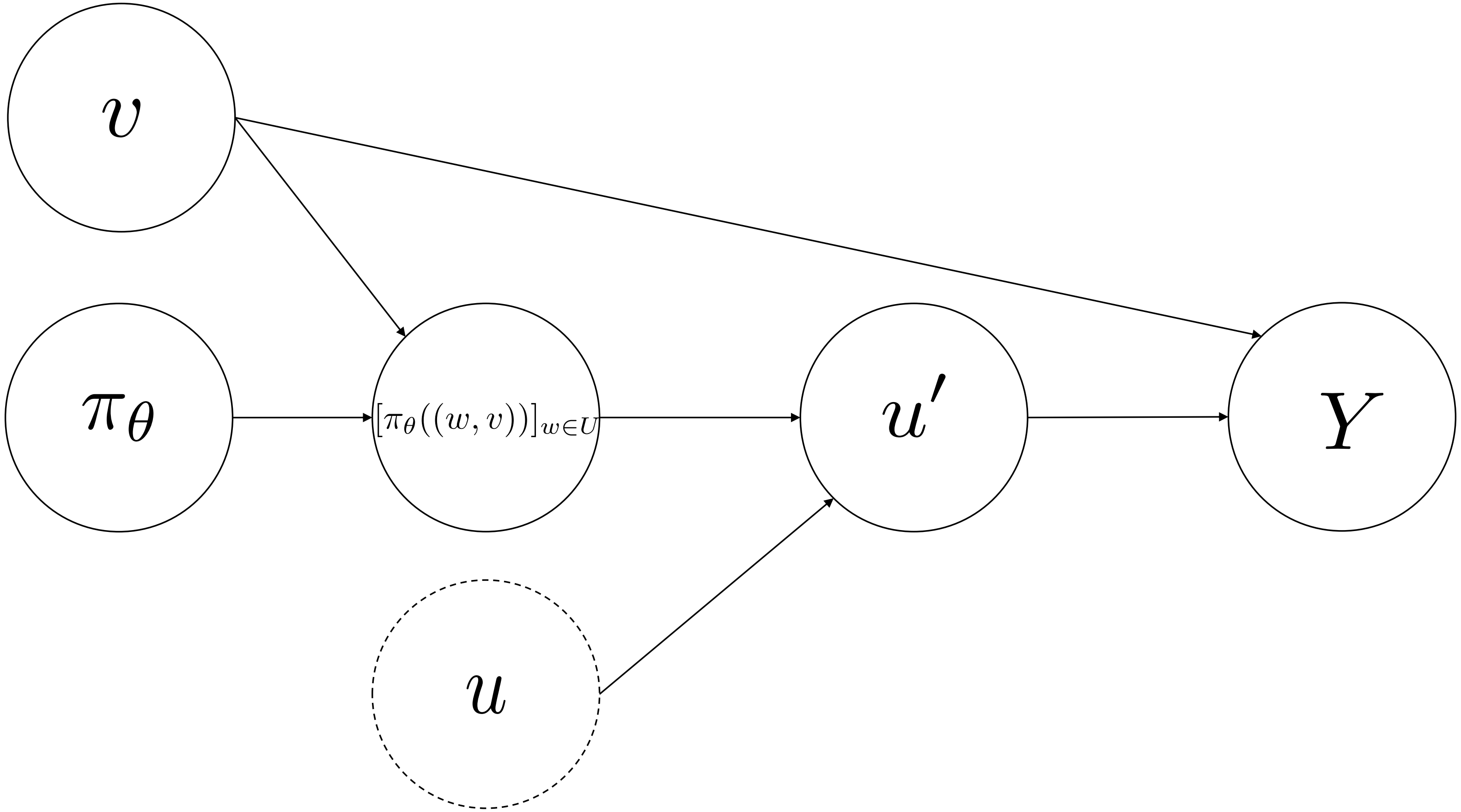
Reward of classified to positive label

Cost of feature manipulation

- The information about π_θ used by agent with covariate (u, v) is only the $\pi_\theta(u', v)$ for all possible u' (he manipulates from u to u').

- We summarize the used information as: $[\pi_\theta(u, v)]_{u \in \mathcal{U}}$

Encapsulate the Impact of Policy



Encapsulate Impact with Evaluation Vector

- Based on the this causal mechanism, we derive

$$\begin{aligned}
p(x_i;\pi_\theta) &= p(u_i|v_i,\pi_\theta)p(v_i|\pi_\theta) \\
&= p(u_i|v_i,\pi_\theta)p(v_i) && (\pi_\theta \text{ only influence } u_i) \\
&= p(u_i|v_i,[\pi_\theta((u,v_i))]_{u\in\mathcal{U}})p(v_i) && (\pi_\theta \text{ is summarized as evaluations}) \\
&= p(u_i|[\pi_\theta((u,v_i))]_{u\in\mathcal{U}})p(v_i) && (\text{conditional independence})
\end{aligned}$$

- Hence, we can derive following performative gradient

$$\nabla_\theta \log p\left(x_i;\pi_\theta\right) = \nabla_\theta \log p\left(u_i \mid \left[\pi_\theta\left(\left(u,v_i\right)\right)\right]_{u\in\mathcal{U}}\right)$$

Behavior Model

- We use another neural network to model such conditional distribution, which actually characterizes the response pattern of agent reaction:

$$q\left(u_i \mid \left[\pi_\theta\left(\left(u, v_i\right)\right)\right]_{u \in \mathcal{U}}\right)$$

- For training this conditional mass function, we utilize the reported covariate X_i as the supervision to implement classification at each epoch, corresponding to the currently released policy network π_{θ_t}.

- Neural network and Gaussian process classifier are both qualified.

Strategic Policy Gradient

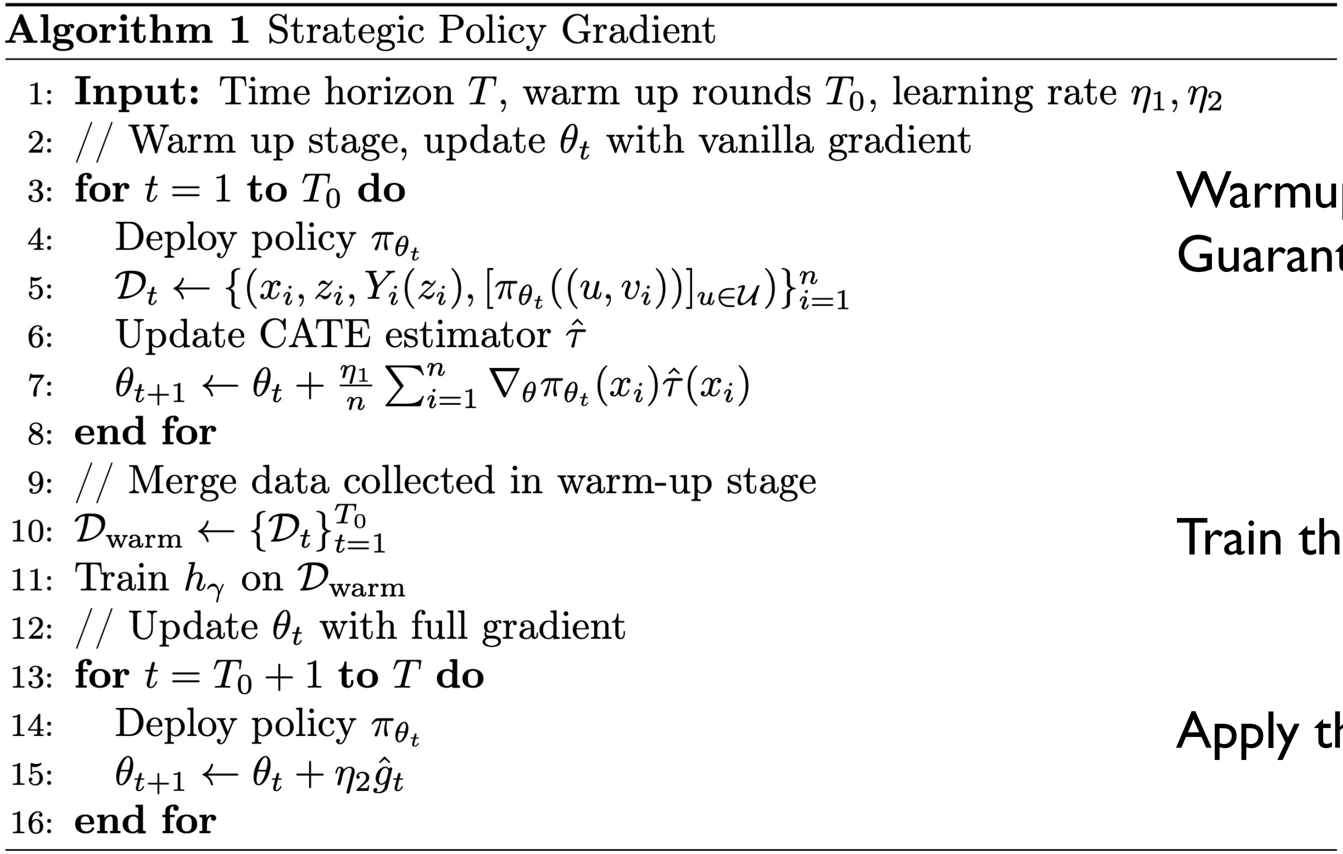
- The policy gradient in general form is

$$\begin{aligned}\nabla_\theta V\left(\pi_\theta\right)=\mathbb{E}_{x\sim p(\cdot;\pi_\theta)}&\left[\nabla_\theta\pi_\theta(x)\tau(x)\right.\\&\left.+\pi_\theta(x)\tau(x)\nabla_\theta\log p\left(x;\pi_\theta\right)\right]\end{aligned}$$

- Through macro-level feature space exploitation and micro-level behavior model, we can approximate such gradient by

$$\begin{aligned}\hat{g}=\nabla_\theta\hat{V}\left(\pi_\theta\right)=\frac{1}{n}\sum_{i=1}^{n}&\left[\nabla_\theta\pi_\theta\left(x_i\right)\hat{\tau}\left(x_i\right)\right.\\&\left.+\pi_\theta\left(x_i\right)\hat{\tau}\left(x_i\right)\nabla_\theta\log\left(q\left(u_i\mid\left[\pi_\theta\left(\left(u,v_i\right)\right)\right]_{u\in\mathcal{U}}\right)\right)\right]\end{aligned}$$

Algorithm: Strategic Policy Gradient



Algorithm 1 Strategic Policy Gradient

```
**Input:** Time horizon $T$, warm up rounds $T_0$, learning rate $\eta_1, \eta_2$
// Warm up stage, update $\theta_t$ with vanilla gradient
**for** $t = 1$ **to** $T_0$ **do**
    Deploy policy $\pi_{\theta_t}$
    $\mathcal{D}_t \leftarrow \{(x_i, z_i, Y_i(z_i), [\pi_{\theta_t}((u, v_i))]_{u \in \mathcal{U}})\}_{i=1}^{n}$
    Update CATE estimator $\hat{\tau}$
    $\theta_{t+1} \leftarrow \theta_t + \frac{\eta_1}{n} \sum_{i=1}^{n} \nabla_\theta \pi_{\theta_t}(x_i) \hat{\tau}(x_i)$
**end for**
// Merge data collected in warm-up stage
$\mathcal{D}_{\text{warm}} \leftarrow \{\mathcal{D}_t\}_{t=1}^{T_0}$
Train $h_\gamma$ on $\mathcal{D}_{\text{warm}}$
// Update $\theta_t$ with full gradient
**for** $t = T_0 + 1$ **to** $T$ **do**
    Deploy policy $\pi_{\theta_t}$
    $\theta_{t+1} \leftarrow \theta_t + \eta_2 \hat{g}_t$
**end for**
```

Warmup stage, we implement repeated retraining. Guarantee stable performance and collect data.

Train the behavior model after warmup stage.

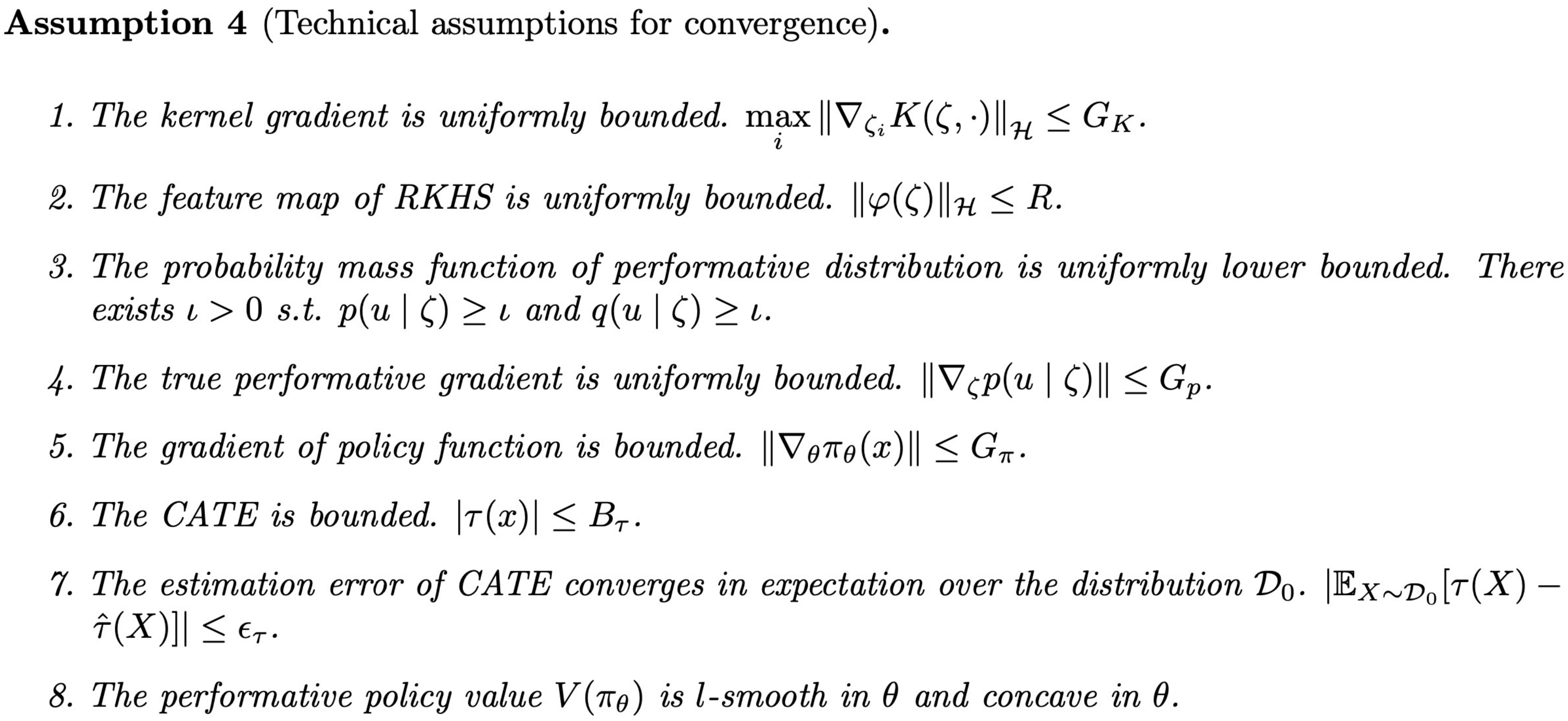
Apply the approximated performative gradient.

Convergence Guarantee

- Before introducing the theoretical result, we first provide some facts:
  - Convergence to the performative optimal solution is extremely hard.
  - Convergence guarantees in pursuit of performative optimal solution are often built for achieving local minima of performative risk.
  - There are usually sacrifice in algorithm for building such a guarantee, e.g., adopting finite difference method to construct gradient estimate, which scales poorly.

- In this paper, we build a convergence guarantee through realizability in Reproducing Kernel Hilbert Space (RKHS) and do not distort our algorithm for theoretical result.

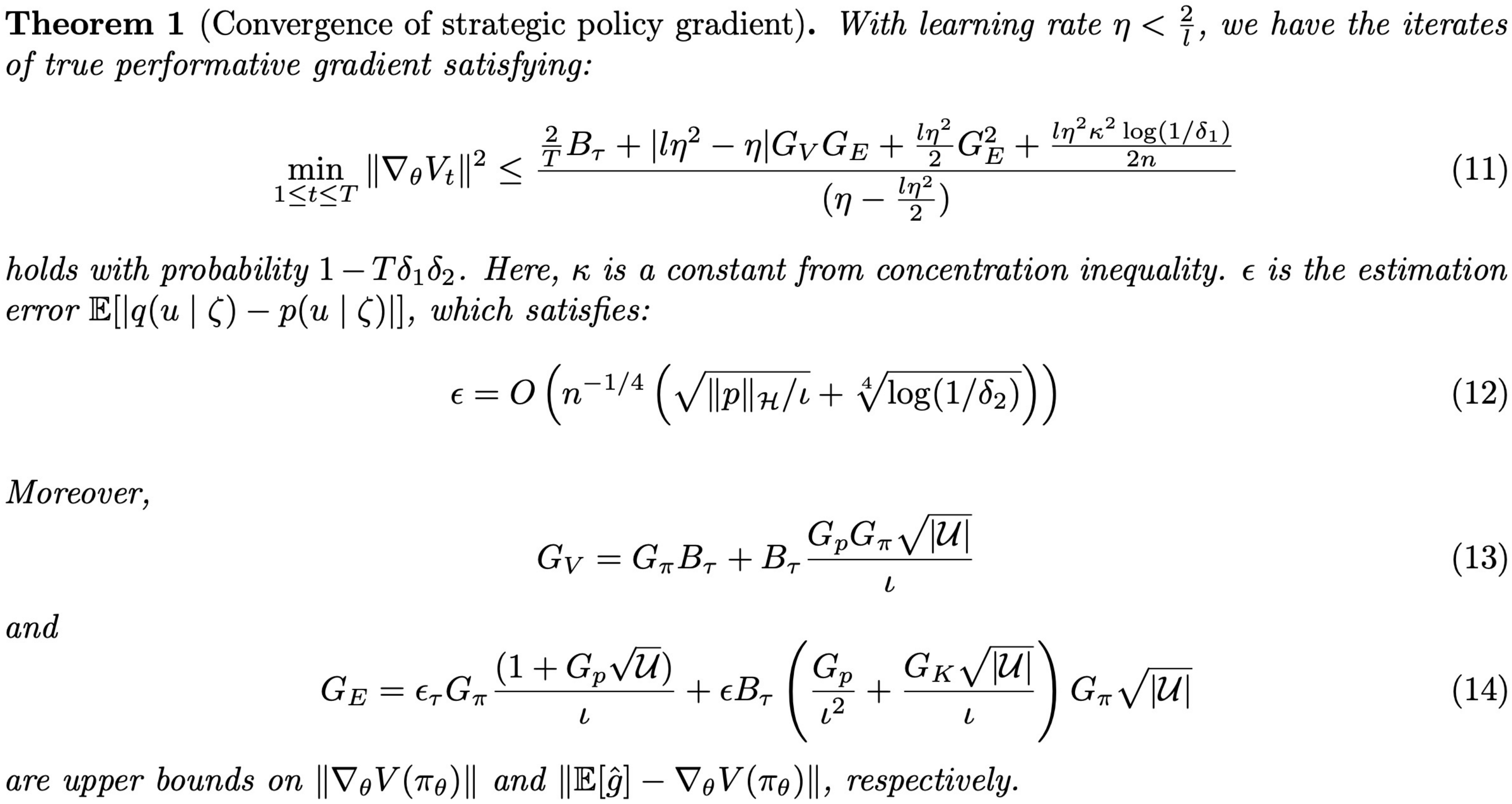
Convergence Guarantee



Assumption 4 (Technical assumptions for convergence)**.**

1. *The kernel gradient is uniformly bounded.* $\max_i \|\nabla_{\zeta_i} K(\zeta, \cdot)\|_{\mathcal{H}} \leq G_K$.
2. *The feature map of RKHS is uniformly bounded.* $\|\varphi(\zeta)\|_{\mathcal{H}} \leq R$.
3. *The probability mass function of performative distribution is uniformly lower bounded. There exists* $\iota > 0$ *s.t.* $p(u \mid \zeta) \geq \iota$ *and* $q(u \mid \zeta) \geq \iota$.
4. *The true performative gradient is uniformly bounded.* $\|\nabla_\zeta p(u \mid \zeta)\| \leq G_p$.
5. *The gradient of policy function is bounded.* $\|\nabla_\theta \pi_\theta(x)\| \leq G_\pi$.
6. *The CATE is bounded.* $|\tau(x)| \leq B_\tau$.
7. *The estimation error of CATE converges in expectation over the distribution* $\mathcal{D}_0$. $|\mathbb{E}_{X \sim \mathcal{D}_0}[\tau(X) - \hat{\tau}(X)]| \leq \epsilon_\tau$.
8. *The performative policy value* $V(\pi_\theta)$ *is* l*-smooth in* θ *and concave in* θ.

Convergence Guarantee



Theorem 1 (Convergence of strategic policy gradient). *With learning rate* $\eta < \frac{2}{l}$, *we have the iterates of true performative gradient satisfying:*

$$\min_{1 \le t \le T} \|\nabla_\theta V_t\|^2 \le \frac{\frac{2}{T} B_\tau + |l\eta^2 - \eta| G_V G_E + \frac{l\eta^2}{2} G_E^2 + \frac{l\eta^2 \kappa^2 \log(1/\delta_1)}{2n}}{(\eta - \frac{l\eta^2}{2})} \tag{11}$$

holds with probability $1 - T\delta_1\delta_2$. *Here,* κ *is a constant from concentration inequality.* ϵ *is the estimation error* $\mathbb{E}[|q(u \mid \zeta) - p(u \mid \zeta)|]$, *which satisfies:*

$$\epsilon = O\left(n^{-1/4}\left(\sqrt{\|p\|_{\mathcal{H}}/\iota} + \sqrt[4]{\log(1/\delta_2)}\right)\right) \tag{12}$$

Moreover,

$$G_V = G_\pi B_\tau + B_\tau \frac{G_p G_\pi \sqrt{|\mathcal{U}|}}{\iota} \tag{13}$$

and

$$G_E = \epsilon_\tau G_\pi \frac{(1 + G_p \sqrt{\mathcal{U}})}{\iota} + \epsilon B_\tau \left(\frac{G_p}{\iota^2} + \frac{G_K \sqrt{|\mathcal{U}|}}{\iota}\right) G_\pi \sqrt{|\mathcal{U}|} \tag{14}$$

are upper bounds on $\|\nabla_\theta V(\pi_\theta)\|$ *and* $\|\mathbb{E}[\hat{g}] - \nabla_\theta V(\pi_\theta)\|$, *respectively.*

Insight: Do not Fully Personalize

- Given some weak regular assumptions on the microfoundation of agent, we can prove that:
- If we release a piecewise constant policy with knots $a_1, \dots a_k$, then all agents that decide to manipulate their feature will move to these knots.

- This means: we can incentivize discretized through deploying a partially personalized policy

- If principal do not extremely scheme and intrigue with agent, then performative learning would become practically feasible.

Simulation Study

High-dimensional data and policy function

Synthetic Experiment: 3 Baselines

1. Cutoff Policy, which only assigns treatments to those agents with positive CATE. It's optimal without strategic behavior. Allocates treatment to all agents with positive CATE.

2. Vanilla Policy Gradient, which ignores the impact of policy on data distribution, and updates the model parameters with gradient w.r.t. CATE only.

3. End2end Policy Gradient, which models $p(u \mid v, \pi_\theta)$ with $p(u \mid v, \theta)$. We construct this baseline method to demonstrate the benefit of taking f_θ as intervention instead of θ.

Synthetic Experiment: Data Geneartion

- We construct fixed feature v from 20 dimensional standard multi-variate Gaussian distribution, and generate u from transformation

$$\mathbb{P}(u \mid v) = \mathrm{Softmax}(Wv)$$

- The cardinality of U is 5, then we design strategic behavior
  - Best response

$$u' = \mathrm{argmax}_u \, \pi_\theta(u, v) - c\,|u - u_0|$$

  - Softmax response

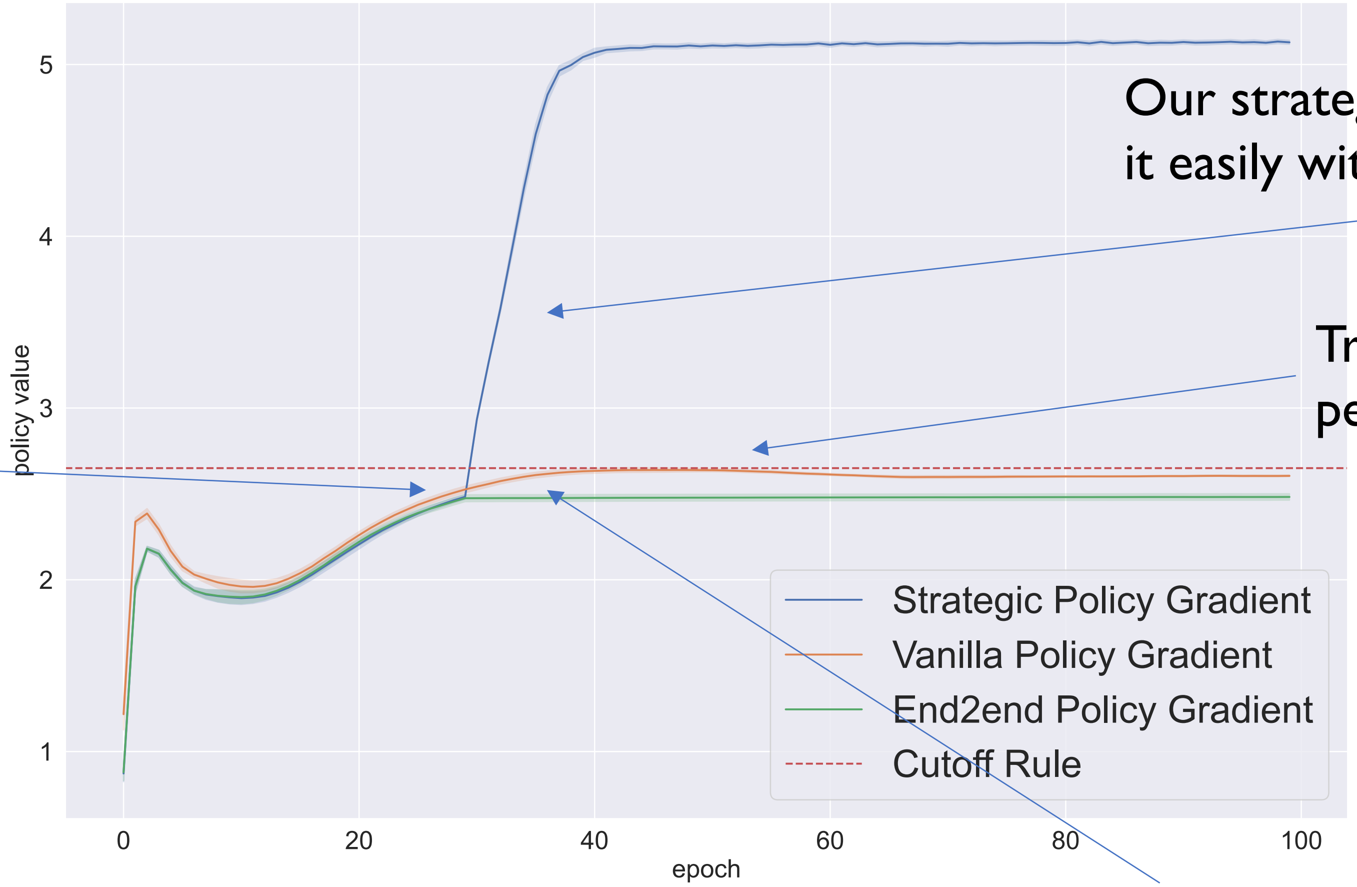
$$u' \sim \mathrm{Softmax}\left(5 \times \left(\pi_\theta(u, v) - c\,|u - u_0|\right)\right)$$

  - Noisy utility

$$u' = \mathrm{argmax}_u \, \pi_\theta(u, v) + \epsilon_\pi - c\,|u - u_0|$$

Part of Result in Synthetic Experiment

Warmup stage (first 30 epochs), all methods are performativity-agnostic

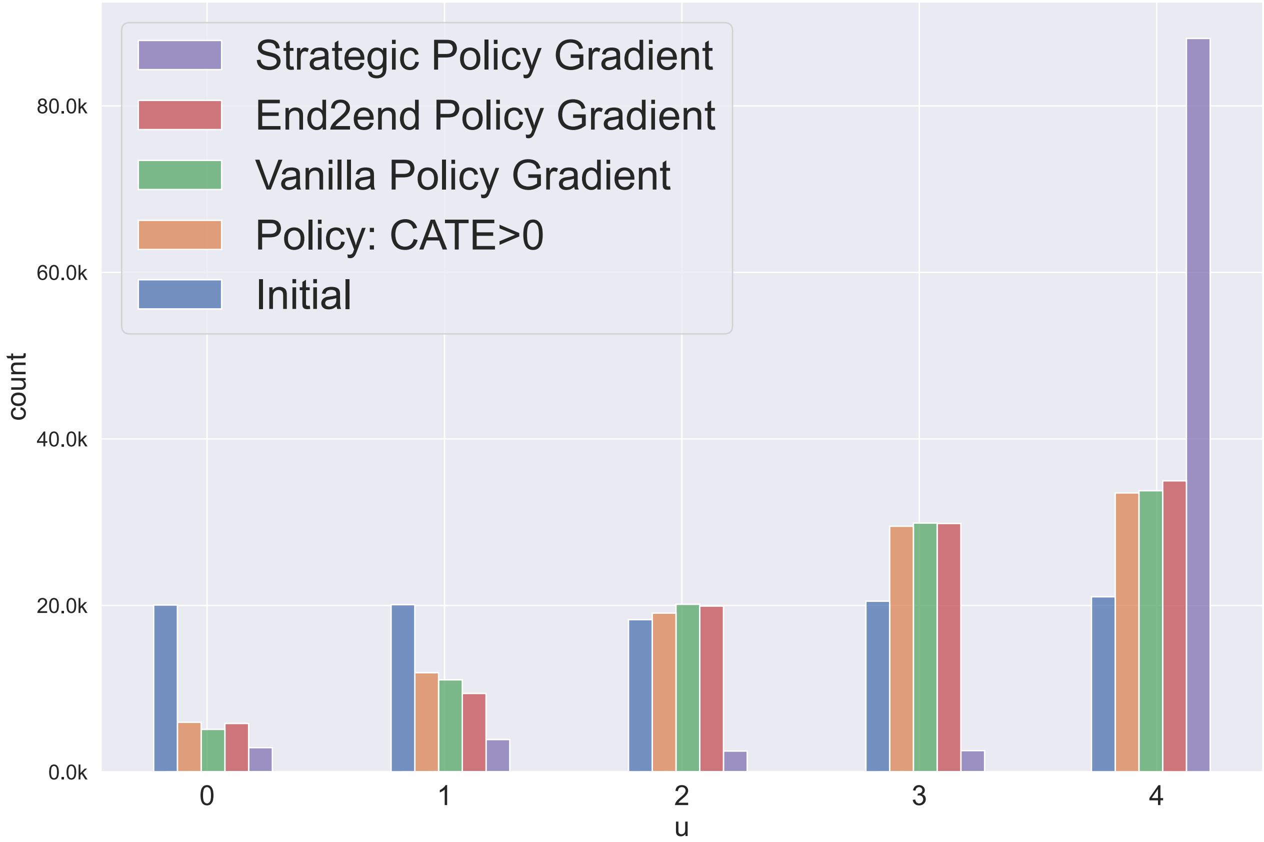


Our strategic policy gradient break it easily with stable convergence

Traditional cutoff policy is a performative stable solution

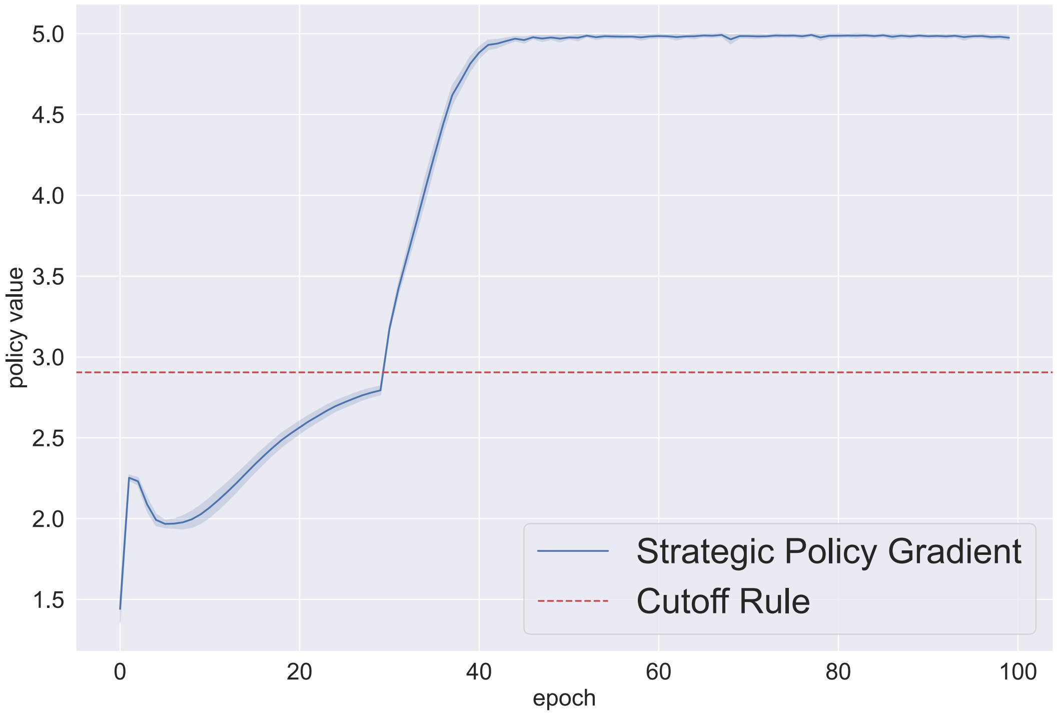
Other baselines can't break it, even it's performativity-aware

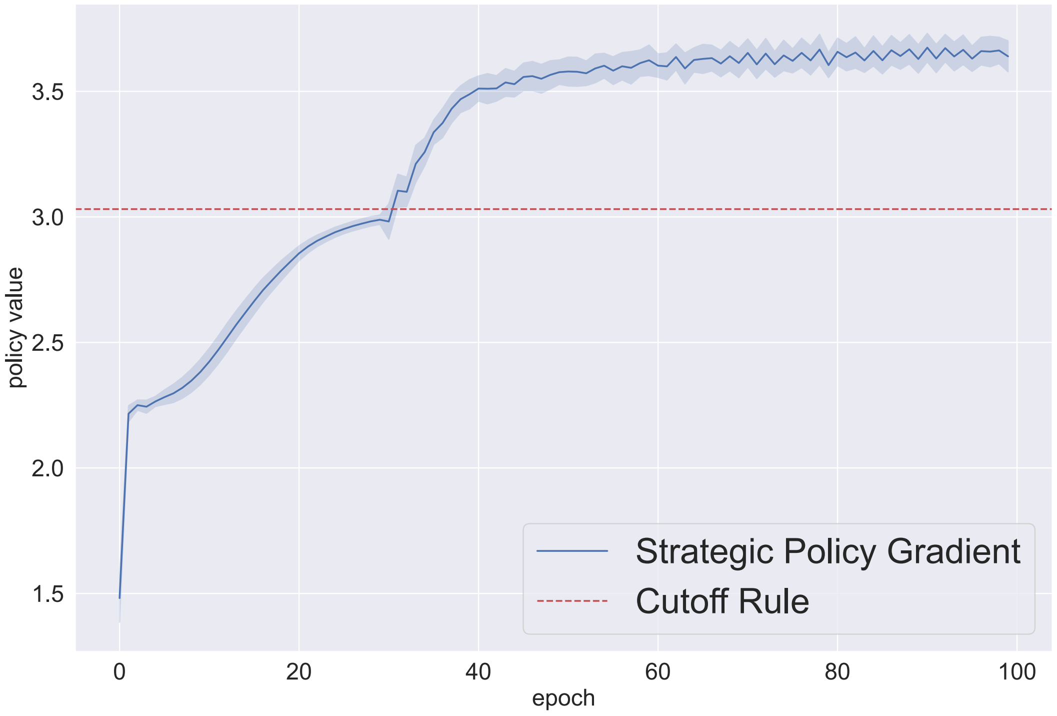
Manipulation Induced by Policies



Our method incentivize much higher proportion of units to increase their u.

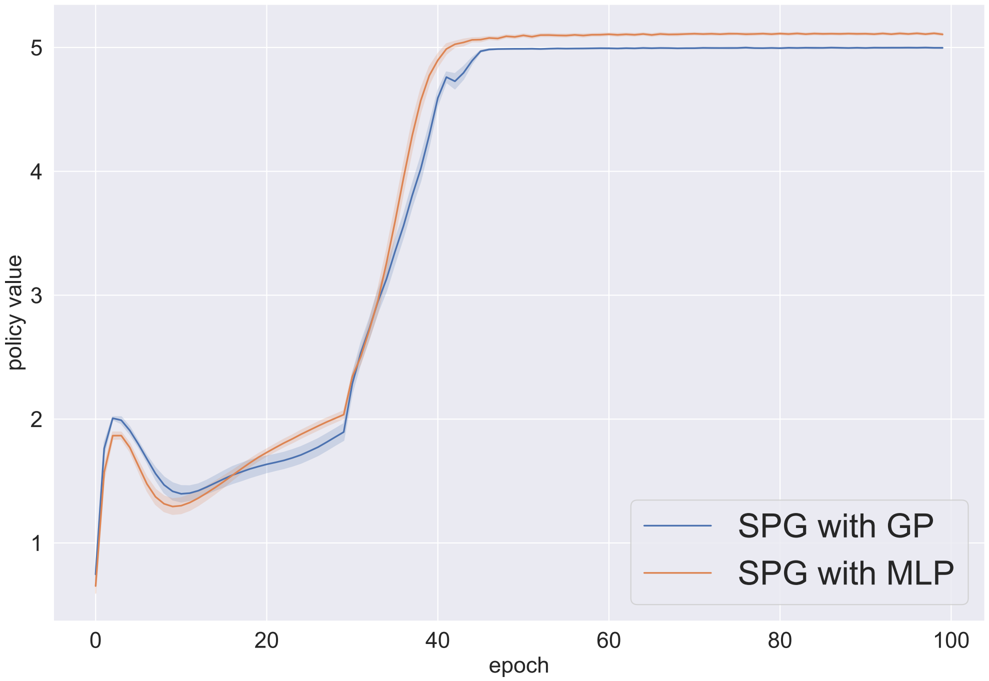
Performance with Noisy/Softmax Utility

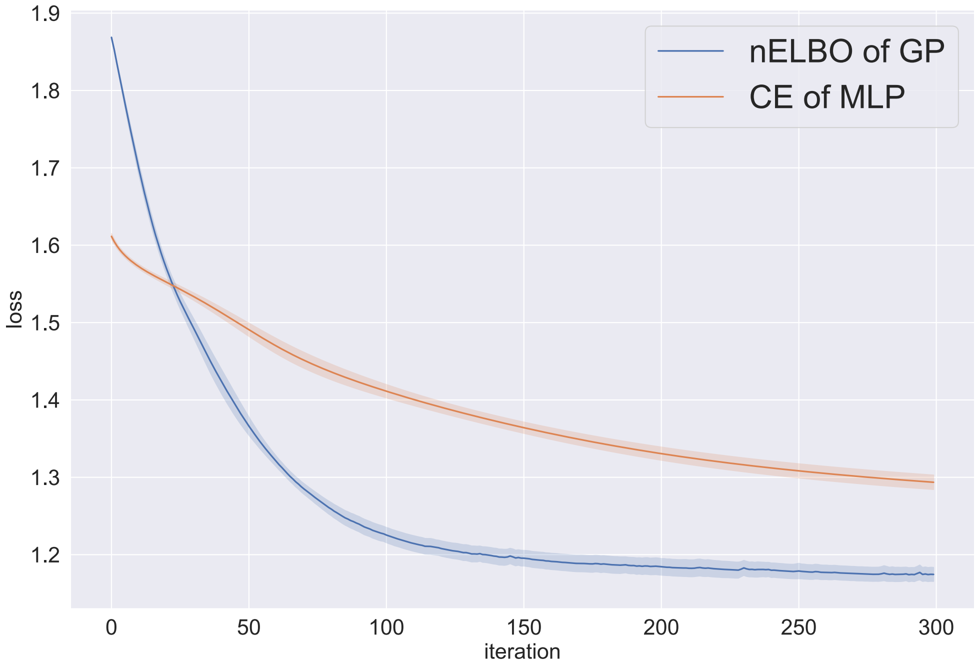




Our method can easily incorporate other special manipulation mechanisms, beyond the classic best-response.

Gaussian Process Classifier as Behavior Model

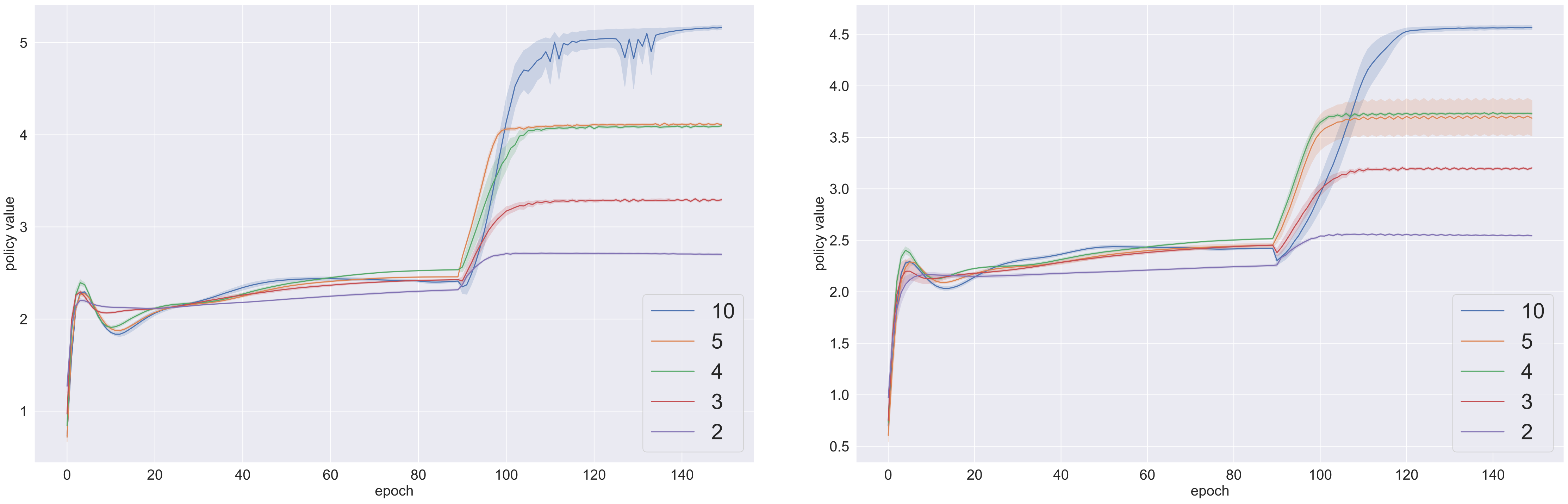




Any powerful classifier with differentiability w.r.t. its input is qualified to serve as our behavior model.

Coarser Discretization of Manipulatable Feature

- True levels of u are 15 (left) and 50 (right), while we deem it as 2, 3, 4, 5, 10.



The factual number of u levels has tiny impact on the performance. The gist lies in the used/deemed levels.

Basic Setting of Semi-synthetic Experiment

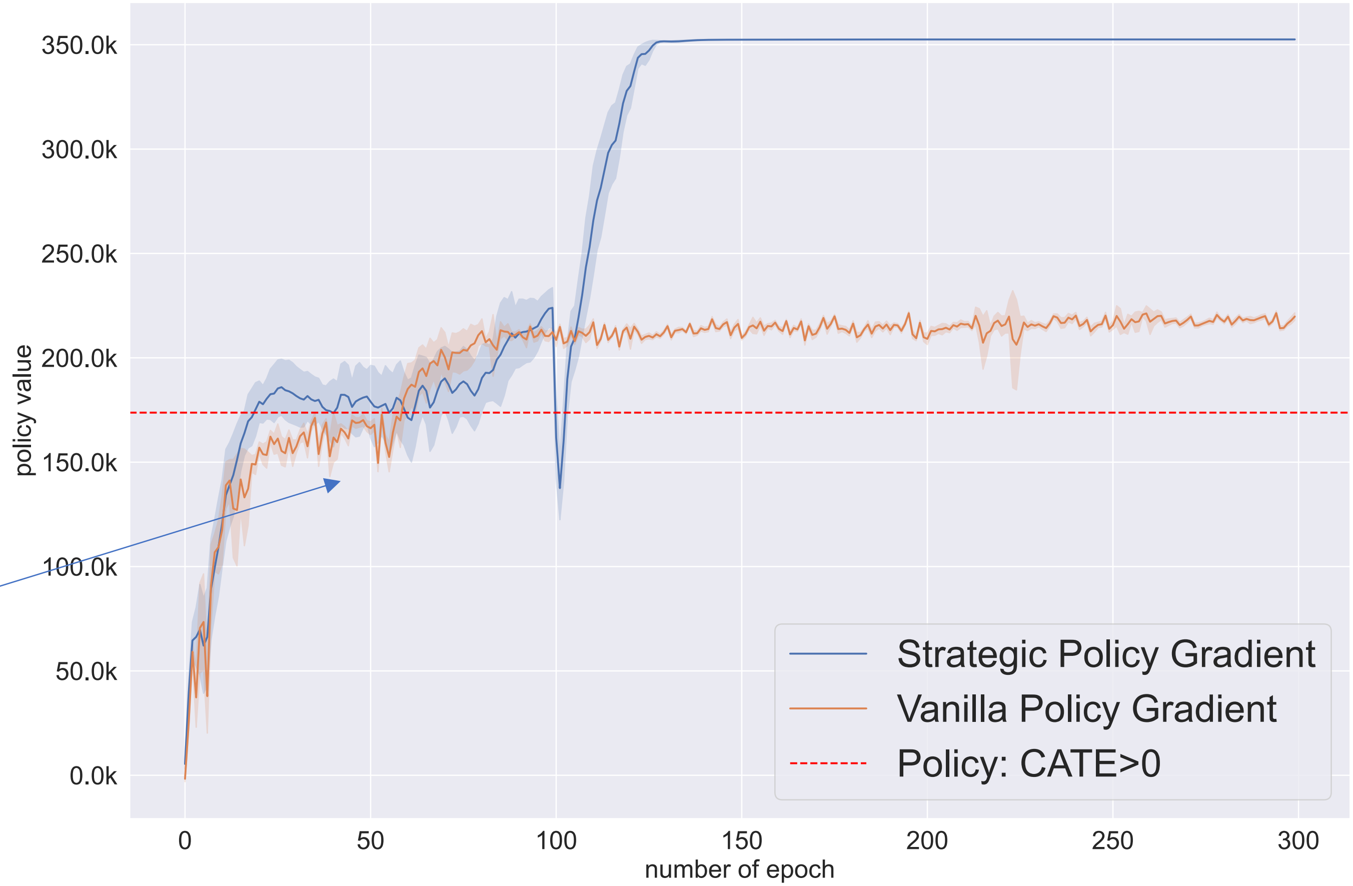
- Scenario: policy learning for lending (loan), 307,508 records

- Fixed feature: loan information (amount, duration, interest rate etc.), demographics (age, profession, credits, etc.) of loan applicant

- Manipulatable feature: external credit score

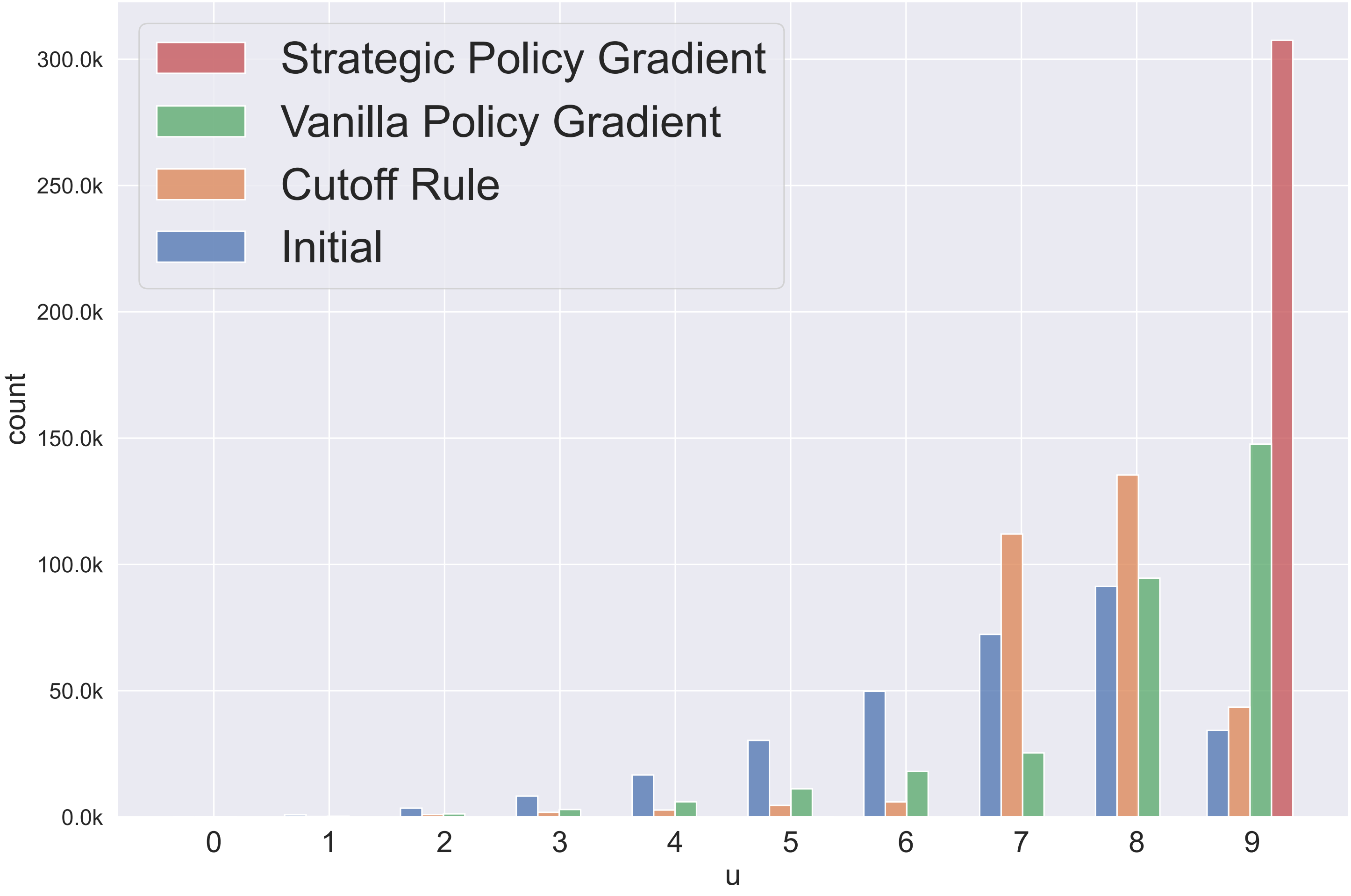
- Outcome: amount of interest, with probability of loss of capital because of default. ($Y(0) = 0$, control indicates rejection)

Result in Semi-synthetic Experiment (c=0.1)

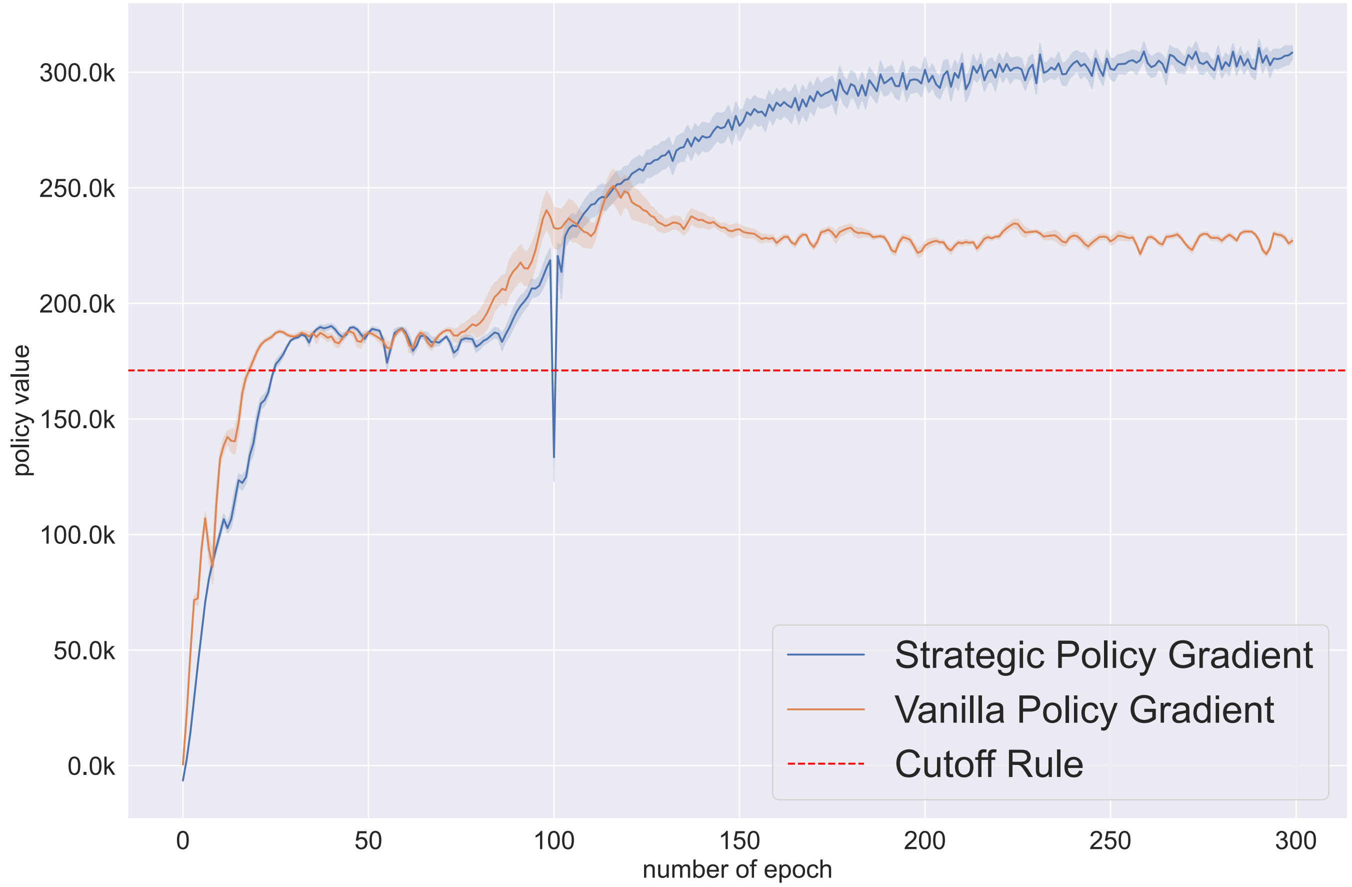


Warmup stage (first 50 epochs), all methods are performativity-agnostic

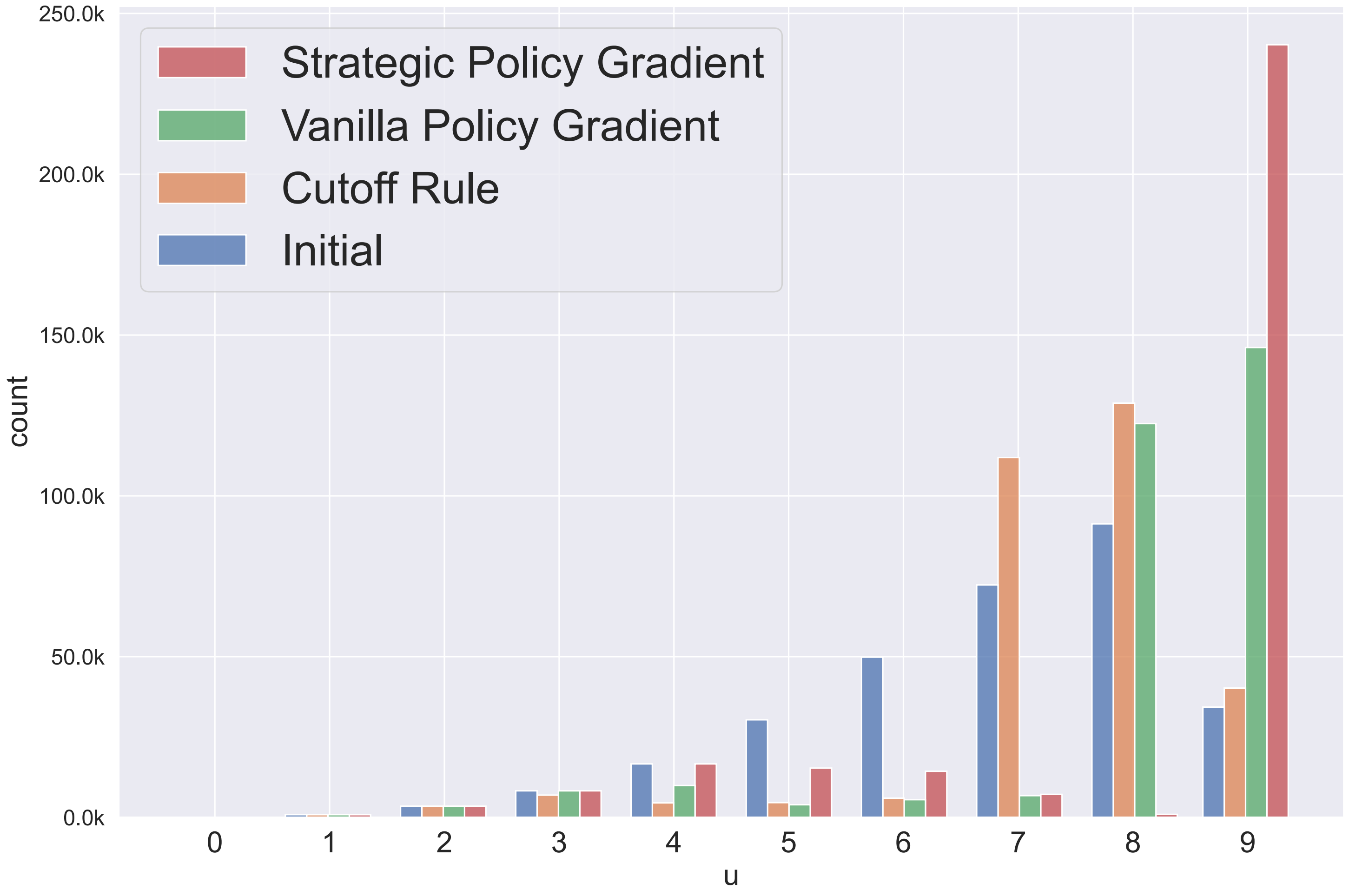
Manipulation Induced by Policies (c=0.1)



Result in Semi-synthetic Experiment (c=0.2)



Manipulation Induced by Policies (c=0.2)



Main Traits of Our Methodology

- No parametric assumption on utility or data distribution
- A new pattern of limited manipulation, in the light of bounded rationality
- Causal mechanism that supports the π_θ acts as intervention (versus θ)
- Use of batch feedback (versus bandit feedback)
- Targeting high-dimensional model parameters and data

Main Reference

Miller, John, Smitha Milli, and Moritz Hardt. "Strategic classification is causal modeling in disguise." International Conference on Machine Learning. PMLR, 2020.

Perdomo, Juan, et al. "Performative prediction." International Conference on Machine Learning. PMLR, 2020.

Izzo, Zachary, Lexing Ying, and James Zou. "How to learn when data reacts to your model: performative gradient descent." International Conference on Machine Learning. PMLR, 2021.

Mendler-Dünner, Celestine, Frances Ding, and Yixin Wang. "Anticipating performativity by predicting from predictions." Advances in neural information processing systems 35 (2022): 31171-31185.

Munro, Evan. "Treatment Allocation with Strategic Agents." Management Science. 2025.

Stratis Tsirtsis, Behzad Tabibian, Moein Khajehnejad, Adish Singla, Bernhard Schölkopf, and Manuel Gomez-Rodriguez. "Optimal decision making under strategic behavior. " Management Science. 2024.

Main Reference

Lechner, Tosca, Ruth Urner, and Shai Ben-David. "Strategic classification with unknown user manipulations." International Conference on Machine Learning. PMLR, 2023.

Ghalme, Ganesh, et al. "Strategic classification in the dark." International Conference on Machine Learning. PMLR, 2021.

Miller, John P., Juan C. Perdomo, and Tijana Zrnic. "Outside the echo chamber: Optimizing the performative risk." International Conference on Machine Learning. PMLR, 2021.

Jagadeesan, Meena, Tijana Zrnic, and Celestine Mendler-Dünner. "Regret minimization with performative feedback." International Conference on Machine Learning. PMLR, 2022.

Manski, Charles F. "Statistical treatment rules for heterogeneous populations." Econometrica 72.4 (2004): 1221-1246.

Athey, Susan, and Stefan Wager. "Policy learning with observational data." Econometrica 89.1 (2021): 133-161.