Introduction to Performative Prediction

Tutorial @ UAI 2024

Celestine Mendler-Dünner

MPI for Intelligent Systems
ELLIS Institute Tübingen
Tübingen AI Center

Tijana Zrnic

Stanford Data Science &
Dept of Statistics
Stanford University

Schedule

1st hour:

Introduction

Framework

Performative stability and retraining

2nd hour:

Performative optimality

Performative optimality vs performative stability

Model-free and model-based optimization

Coffee break

3rd hour:

Extensions of the framework and connections

Power, incentives, digital activism

Discussion

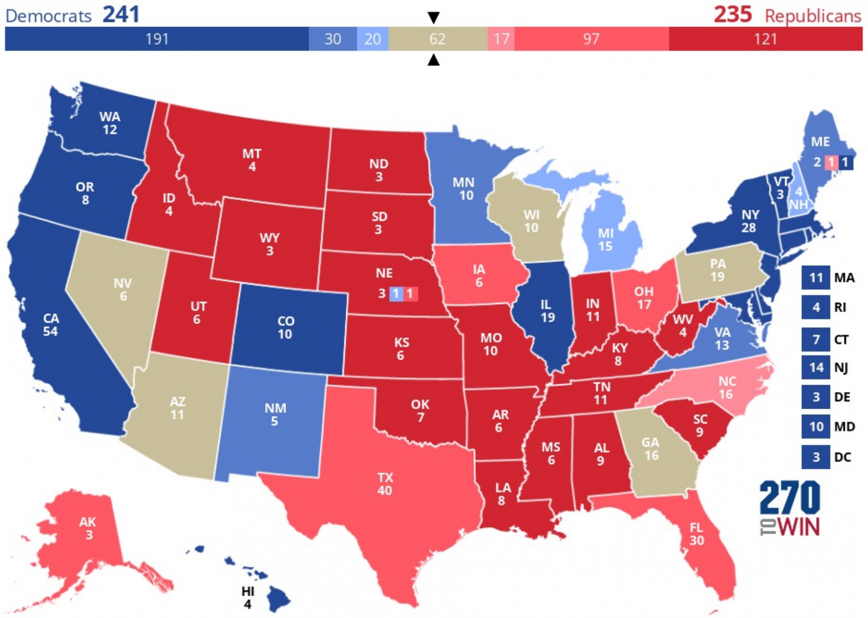
Different facets of prediction

Prediction in the social world is different from prediction in physical systems

Prediction in astronomy:
Detect regularities and laws in nature, purely explanatory and descriptive

Prediction in social context:
Predictions are an intrinsic part of the system, they inform decisions, beliefs and outcomes

Different facets of prediction



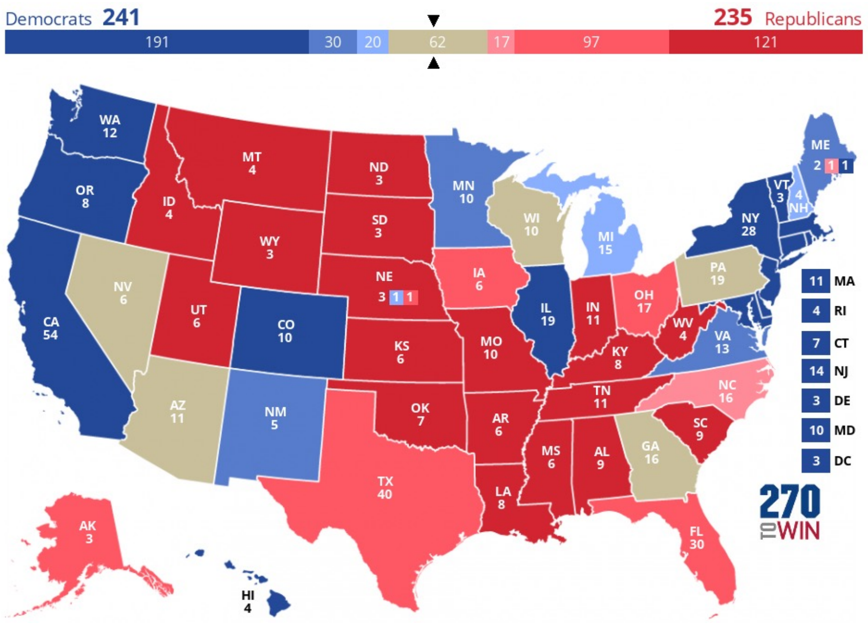
prediction of
election outcomes

FiveThirtyEight publishes predictions of US election outcome

Predictions change expectations and beliefs of individuals

They impact voter turnout and election outcome

Different facets of prediction



prediction of
election outcomes

FiveThirtyEight publishes predictions of US election outcome

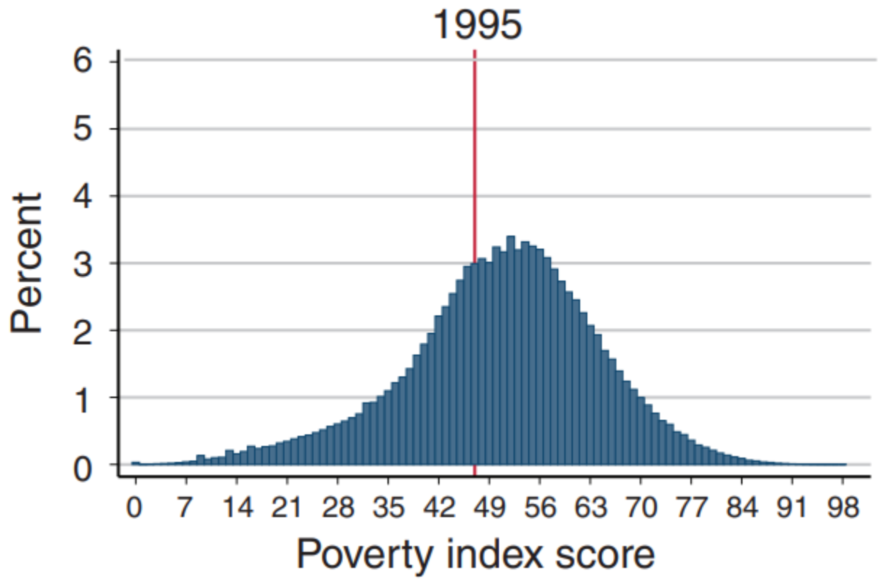
Predictions change expectations and beliefs of individuals

They impact voter turnout and election outcome

Herbert Simon 1954.

"Bandwagon and Underdog Effects and the Possibility of Election Predictions"

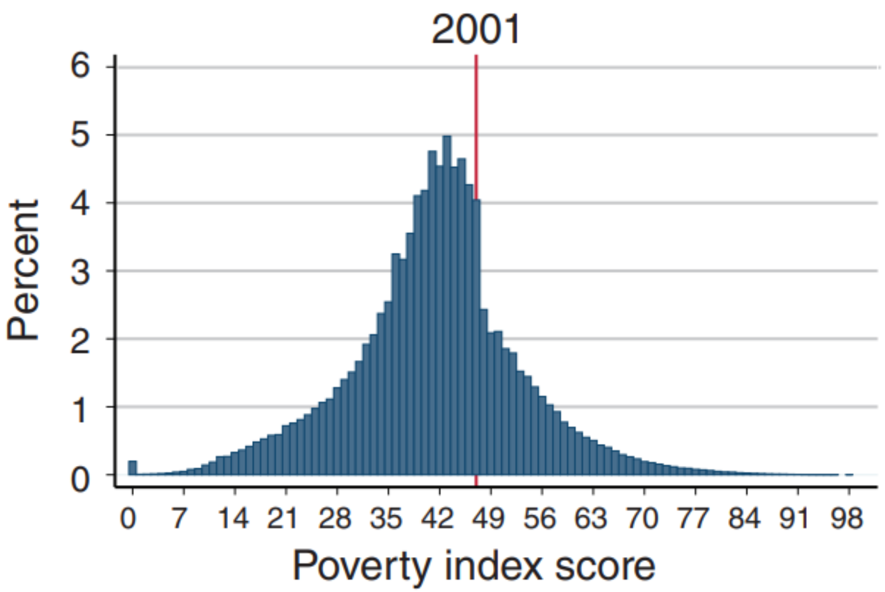
Different facets of prediction



Government agencies make predictions about socioeconomic status

Predictions are used to allocate benefits

Different facets of prediction



Camacho & Conover,
American Economic Journal, 2011

Government agencies make predictions about socioeconomic status

Predictions are used to allocate benefits

People adapt and statistical regularities in the population collapse

Different facets of prediction

“Option pricing theory—a “crown jewel” of neoclassical economics—succeeded empirically not because it discovered preexisting price patterns but because it pushed the market to conform to its predictions […].”

MacKenzie & Millo, American Journal of Sociology, 2003

Oskar Morgenstern

Habilitation, 1928

In the physical world - unlike the social world - there is no causal relationship between the prediction of an event and its occurrence.

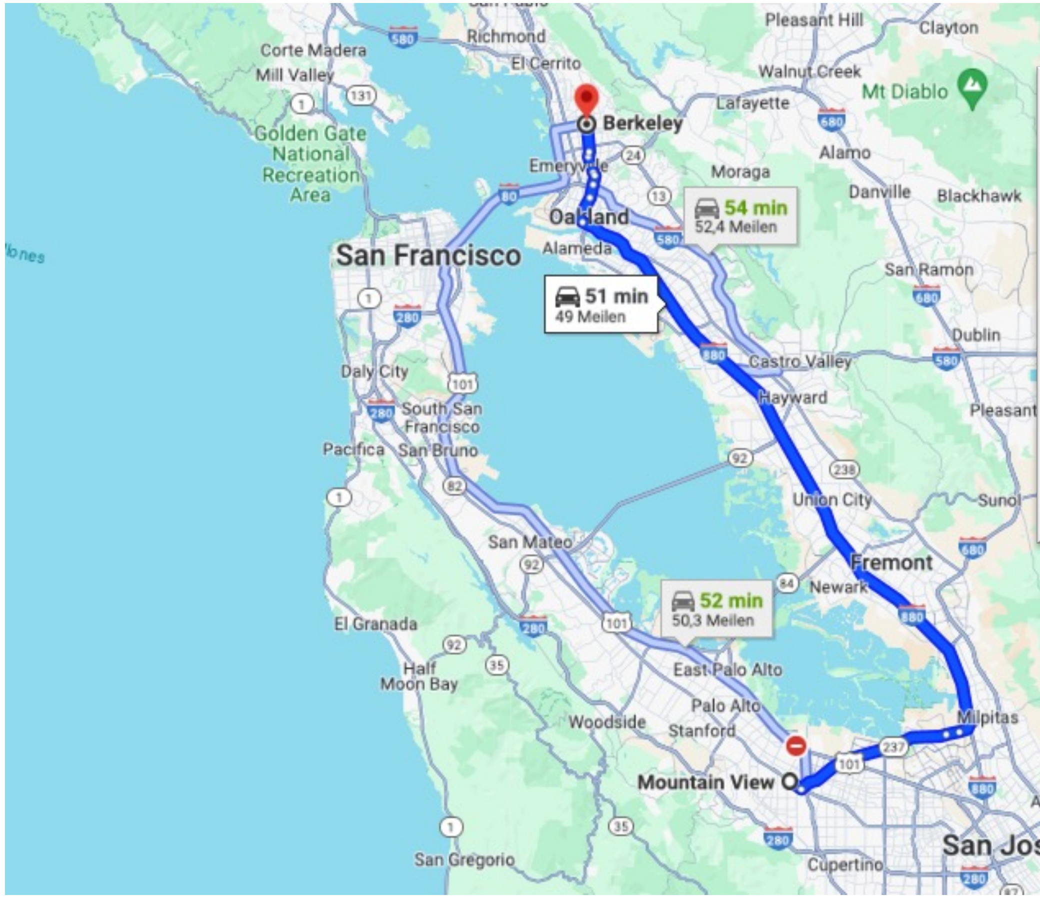
Forecasts that can impact the predicted event constitute one of the most central problems in the theory of economic forecasting

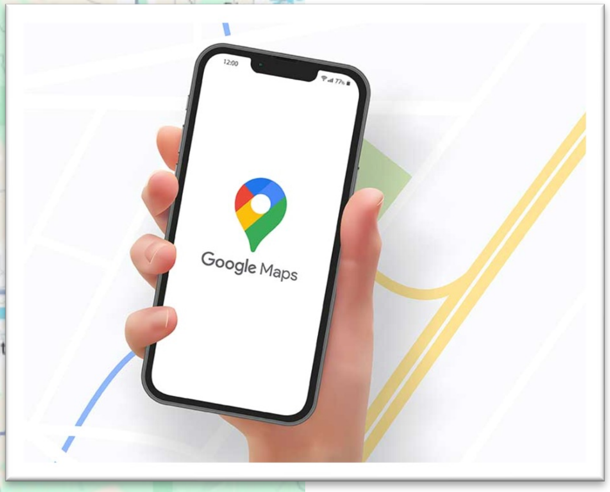
Habilitation titled “Wirtschaftsprognose”. 1928.

What about machine learning?

We routinely make predictions in economic and social contexts!

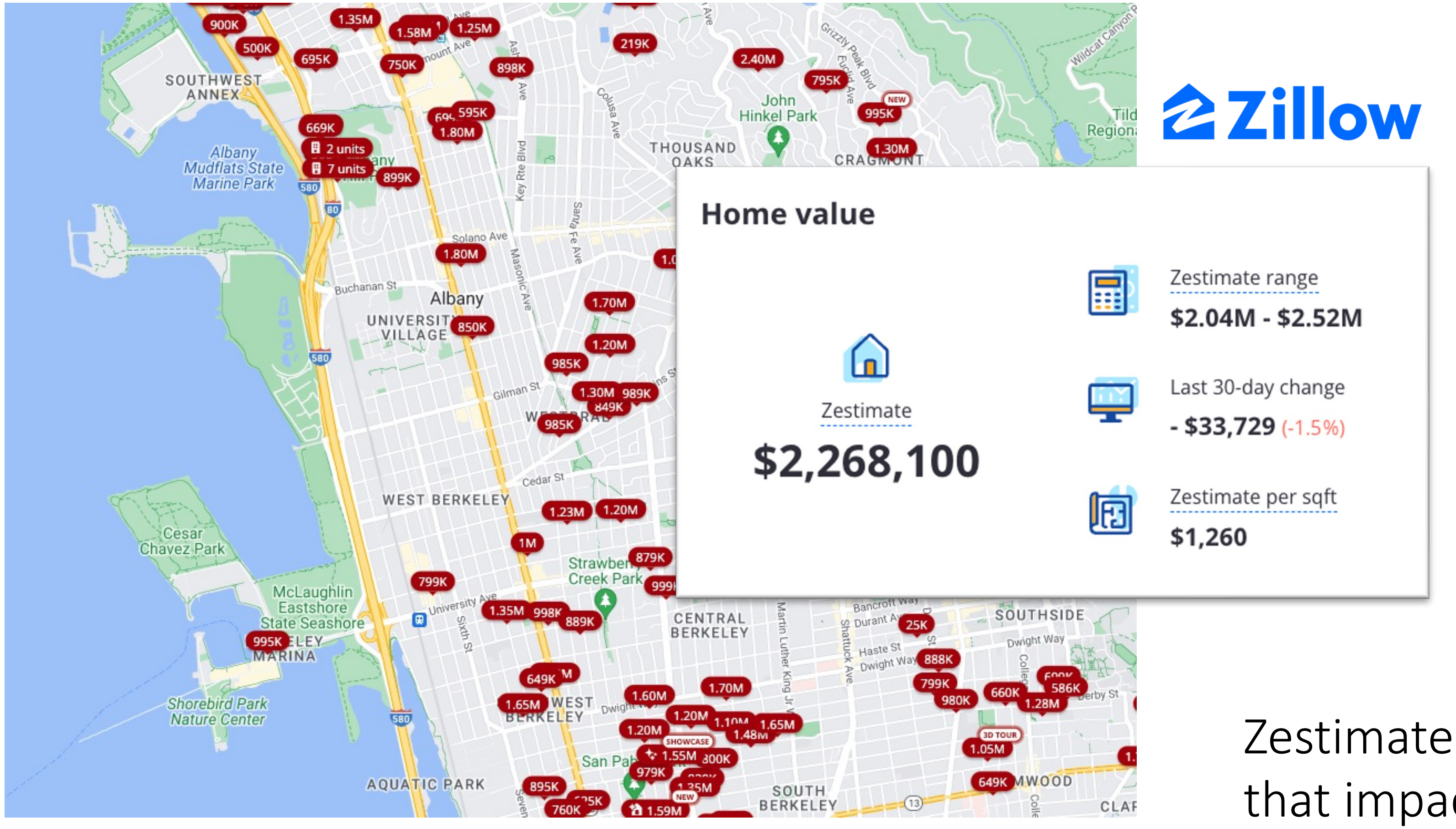
Prediction in machine learning





Traffic predictions impact routing decisions and hence traffic

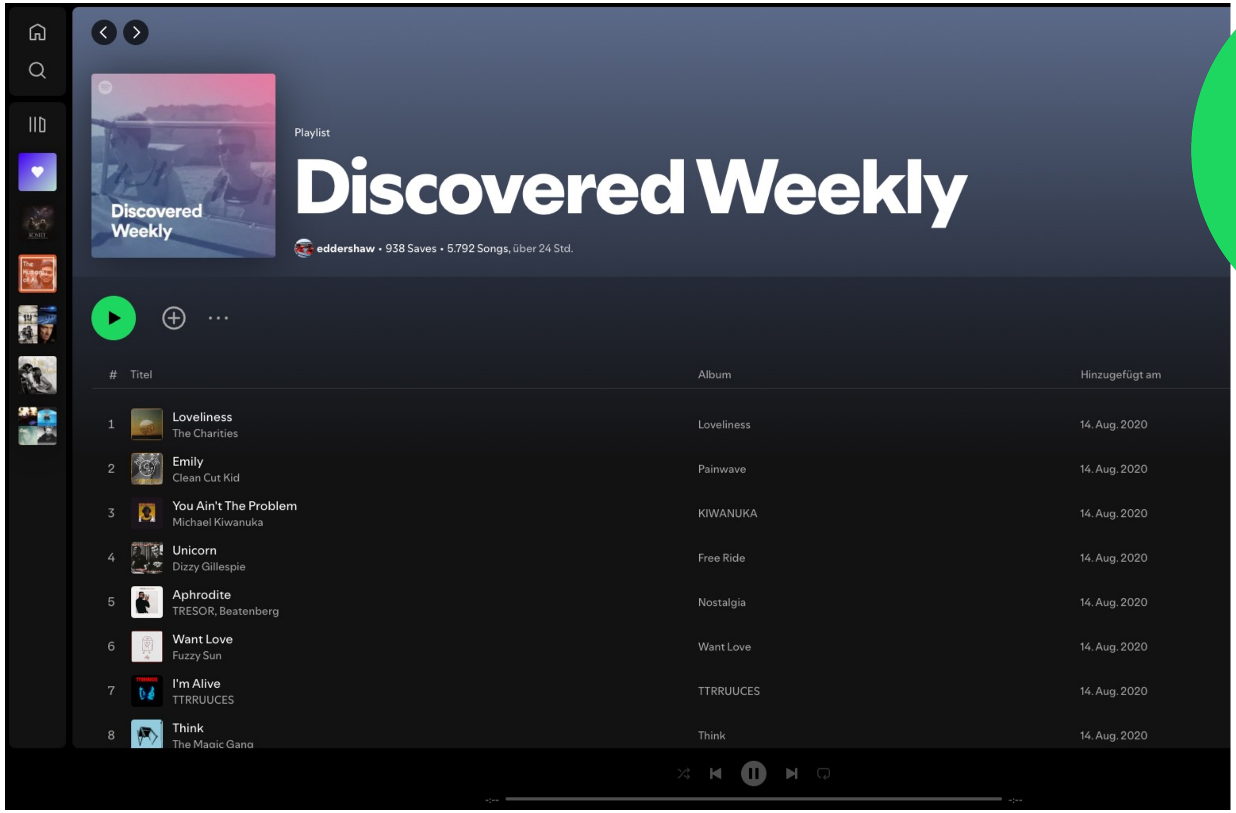
Prediction in machine learning



Zillow

Zestimates set expectations that impact sales prices

Prediction in machine learning



Recommender systems filter information and shape consumption

Supervised learning

- We represent the population as a distribution D over data instances (X, Y)
- Predictive model given by a parameter vector θ
- Find a good predictive model through risk minimization

$$\mathrm{Risk}(\theta, D) = \mathrm{E}_{(x,y)\sim D}\ [\,\mathrm{loss}((x, y); \theta)\,]$$

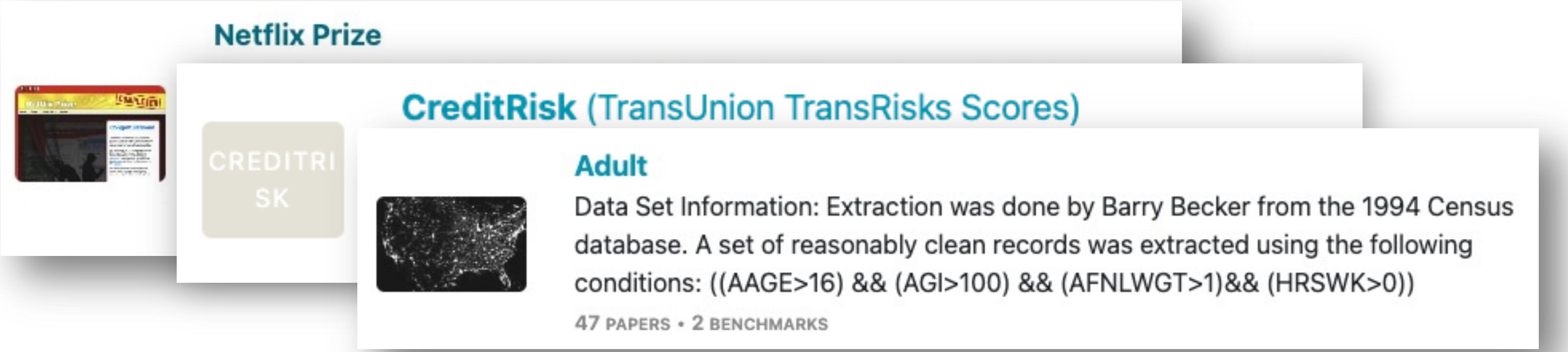
static description of the world

Supervised learning

- We represent the population as a distribution D over data instances (X, Y)
- Predictive model given by a parameter vector θ
- Find a good predictive model through risk minimization

$$\mathrm{Risk}(\theta, D) = \mathrm{E}_{(x,y)\sim D}\ [\, \mathrm{loss}((x,y);\theta)\,]$$

static description of the world



Supervised learning

- We represent the population as a distribution D over data instances (X, Y)
- Predictive model given by a parameter vector θ
- Find a good predictive model through risk minimization

$$\mathrm{Risk}(\theta, D) = \mathrm{E}_{(x,y)\sim D}\ [\,\mathrm{loss}((x, y); \theta)\,]$$

static description of the world

No language to articulate Morgenstern's argument

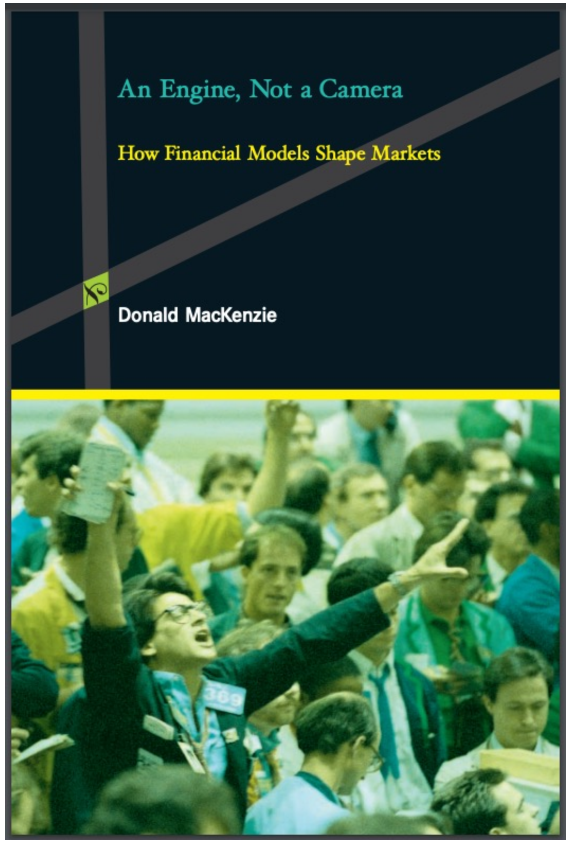
Performative prediction

An extension of the classical risk minimization framework that accounts for the causal effect of predictions on the target of prediction

Performative prediction

An extension of the classical risk minimization framework that accounts for the causal effect of predictions on the target of prediction

Performativity is an established concept in economics, finance, public policy, and social science
see, e.g., M. Callon, D. MacKenzie



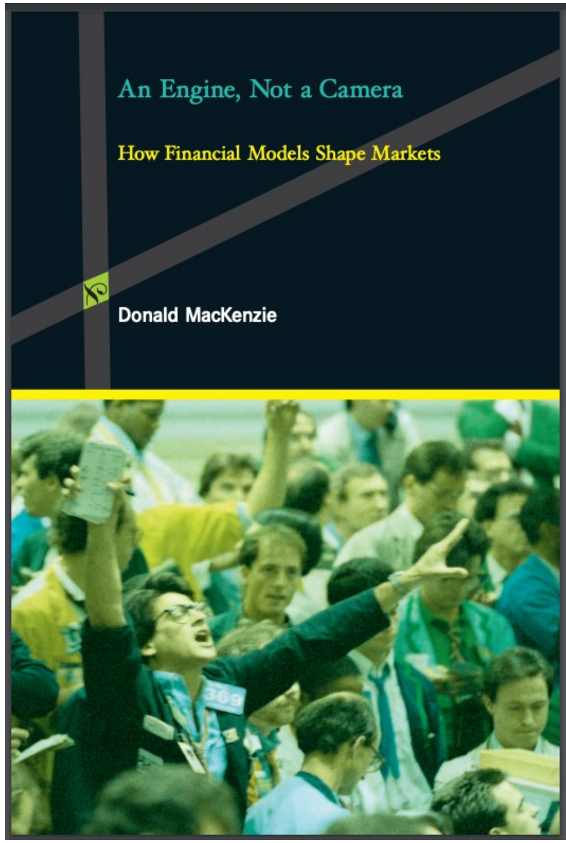
Donald MacKenzie. 2006.
"How Financial Models Shape Markets"

Performative prediction

An extension of the classical risk minimization framework that accounts for the causal effect of predictions on the target of prediction

Performativity is an established concept in economics, finance, public policy, and social science
see, e.g., M. Callon, D. MacKenzie

Goal: bring performativity as a concept into the foundations of machine learning



Donald MacKenzie. 2006.
"How Financial Models Shape Markets"

Framework

[PZMH20]

Performative prediction framework

Performativity thesis:

Predictions can have a causal influence on the world they aim to predict

[PZMH20]

Performative prediction framework

Performativity thesis:

Predictions can have a causal influence on the world they aim to predict

Lens to the world is the data

→ Data distribution $D(\theta)$ changes in response to a deployed model θ

[PZMH20]

Performative prediction framework

Performativity thesis:

Predictions can have a causal influence on the world they aim to predict

Lens to the world is the data

→ Data distribution $D(\theta)$ changes in response to a deployed model θ

$$\text{Risk}(\theta, D(\theta)) = \mathrm{E}_{(x,y)\sim D(\theta)}\ [\,\text{loss}((x,y);\theta)\,]$$

the observed loss of a model is the loss on the distribution
that surfaces after its deployment

[PZMH20]

Performative prediction framework

Performativity thesis:

Predictions can have a causal influence on the world they aim to predict

Lens to the world is the data

→ Data distribution $D(\theta)$ changes in response to a deployed model θ

$$\text{Risk}(\theta, D(\theta)) = \mathrm{E}_{(x,y)\sim D(\theta)}\left[\,\text{loss}((x,y);\theta)\,\right]$$

the model impacts the risk in two ways:
through the loss and the distribution

Performative stability

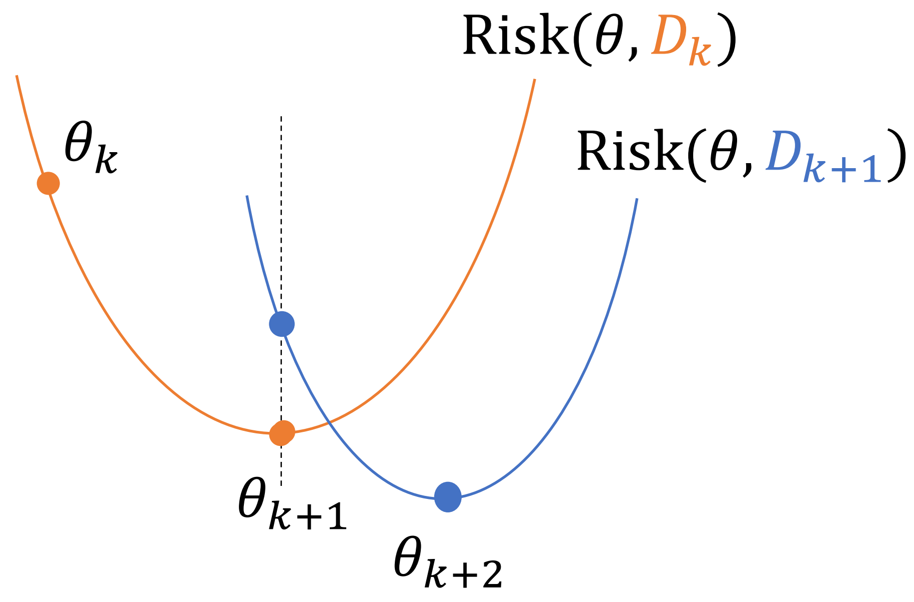
Retraining and stability

Collect data and update your model given data

Repeated risk minimization (RRM):

1. observe data distribution D_k
2. let θ_{k+1} be the risk minimizer on D_k
3. deploy $\theta_{k+1} \rightarrow D_{k+1} = D(\theta_{k+1})$
4. repeat

$$\theta_{k+1} \leftarrow \mathrm{argmin}_\theta \, \mathrm{Risk}(\theta, D_k)$$



[PZMH20]

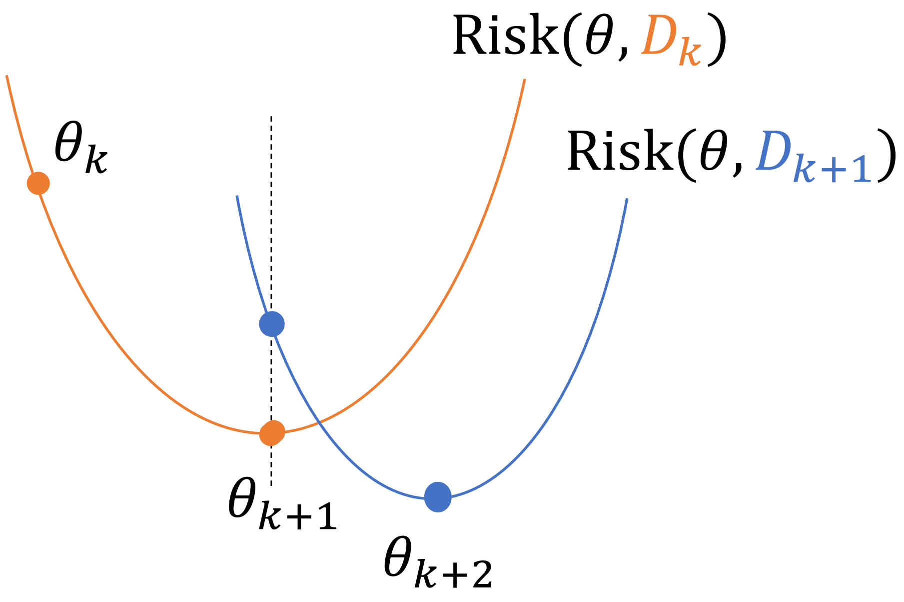
Retraining and stability

Collect data and update your model given data

Repeated risk minimization (RRM):

1. observe data distribution D_k
2. let θ_{k+1} be the risk minimizer on D_k
3. deploy $\theta_{k+1} \rightarrow D_{k+1} = D(\theta_{k+1})$
4. repeat

$$\theta_{k+1} \leftarrow \mathrm{argmin}_\theta \ \mathrm{Risk}(\theta, D_k)$$



Performative stability:

$$\theta^* = \mathrm{argmin}_\theta \ \mathrm{Risk}(\theta, D(\theta^*))$$

Model remains optimal after deployment

A natural equilibrium concept

When does retraining converge?

Definition: We say the distribution map $D(\theta)$ is ϵ-sensitive if for all θ, θ'

$$W\big(D(\theta), D(\theta')\big) \leq \epsilon \big|\big|\theta - \theta'\big|\big|_2$$

"Similar models lead to similar distributions"

When does retraining converge?

Definition: We say the distribution map $D(\theta)$ is ϵ-sensitive if for all θ, θ'

$$W\big(D(\theta), D(\theta')\big) \leq \epsilon\big||\theta - \theta'|\big|_2$$

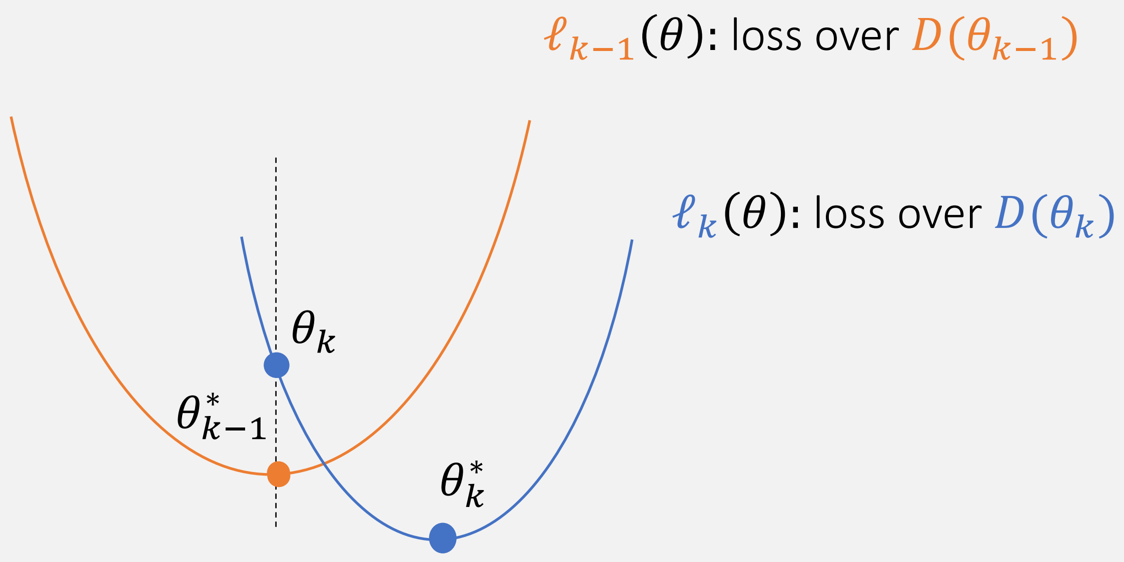
"Similar models lead to similar distributions"

Theorem [PZMH20]: Suppose the loss function is γ-strongly convex in θ and β-smooth in the data*. Then retraining converges to a unique stable point as long as $D(\theta)$ is not too sensitive: $\epsilon < \gamma/\beta$. The rate of convergence is linear:

$$\big||\theta_k - \theta^*|\big| \leq \left(\frac{\epsilon\beta}{\gamma}\right)^k ||\theta_0 - \theta^*||$$

* $\nabla_\theta \ell(z, \theta)$ is β-Lipschitz in z

Proof sketch

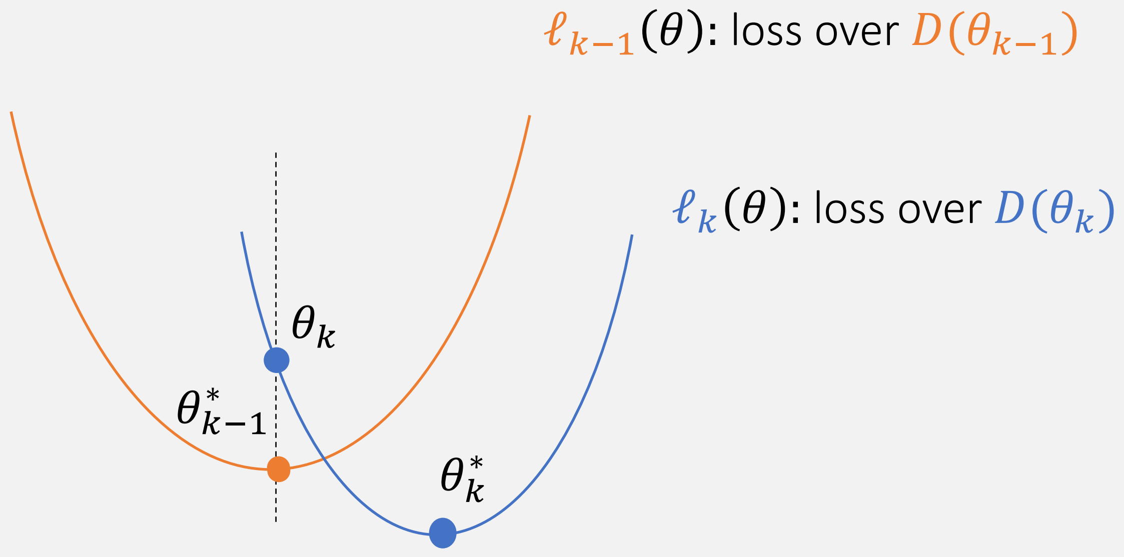


- γ-strong convexity of the loss in θ: $\quad [\nabla\ell_k(\theta_k) - \cancelto{0}{\nabla\ell_k(\theta_k^*)}]^T(\theta_k - \theta_k^*) \geq \gamma||\theta_k - \theta_k^*||^2$
- β-smoothness of the loss in the data: $\quad [\nabla\ell_k(\theta_k) - \cancelto{0}{\nabla\ell_{k-1}(\theta_k)}]^T(\theta_k - \theta_k^*) \leq \beta\big||\theta_k - \theta_k^*|\big|\, W(D(\theta_{k-1}), D(\theta_k))$

<u>Kantorovich-Rubinstein duality theorem:</u> for L-Lipschitz functions g:

$$\mathrm{E}_{x\sim D_1} g(x) - \mathrm{E}_{x\sim D_2} g(x) \leq L\, W(D_1, D_2)$$

Proof sketch



- γ-strong convexity of the loss in θ: $[\nabla\ell_k(\theta_k) - \cancelto{0}{\nabla\ell_k(\theta_k^*)}]^T(\theta_k - \theta_k^*) \geq \gamma||\theta_k - \theta_k^*||^2$

- β-smoothness of the loss in the data: $[\nabla\ell_k(\theta_k) - \cancelto{0}{\nabla\ell_{k-1}(\theta_k)}]^T(\theta_k - \theta_k^*) \leq \beta\big||\theta_k - \theta_k^*|\big| W(D(\theta_{k-1}), D(\theta_k))$

$$\Rightarrow ||\theta_k - \theta_k^*|| \leq \frac{\beta}{\gamma} W(D(\theta_{k-1}), D(\theta_k))$$

- ϵ-sensitivity of $D(\cdot)$:

$$\leq \frac{\beta}{\gamma}\epsilon\,\big||\theta_{k-1} - \theta_k|\big|$$

$$= \frac{\beta}{\gamma}\epsilon\,\big||\theta_{k-1} - \theta_{k-1}^*|\big| \qquad \text{contraction for } \epsilon < \gamma/\beta \qquad \square$$

Fixed-point argument

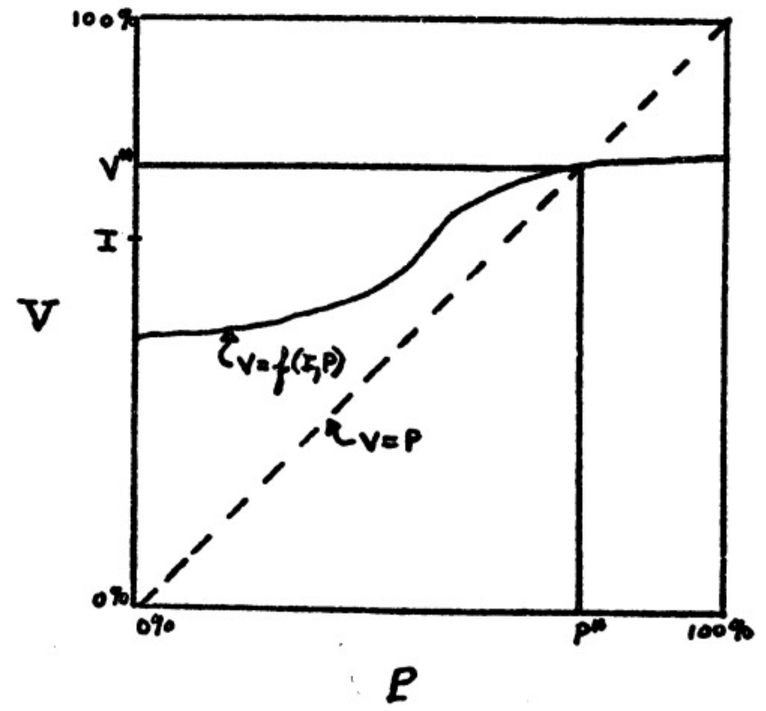
Historical arguments about the possibility of public prediction using Brouwer's fixed point theorem

- Simon 1954
- Grunberg, Modigliani 1954

If the response funtion $Y = R(\hat{Y})$ is continuous then perfect public prediction is possible

Performative stability is a natural analogue of these fixed points in parametric prediction settings

FIGURE 2



A formal proof of the theorem will not be given here. It is a classical theorem of topology due to Brouwer (the "fixed-point" theorem), and a non-technical exposition may be found in *What is Mathematics?*.[2] The reader who does not demand a rigorous proof may satisfy himself of the correctness of the theorem by graphical means. Construct a figure like Figure 2, but omit the solid curve. Mark any point on the y-axis between V = 0 per cent and V = 100 per cent; and a second arbitrary point on the vertical line, P = 100 per cent, within the same limits. Now try to connect these two points, without lifting the pencil from the paper, without going outside the limits 0 per cent to 100 per cent for V and P (that is, without going outside the square), and without intersecting the broken line. Since this is impossible, *any* continuous curve relating V and P for the whole range of values $0\% \leq P \leq 100\%$ must intersect the line V = P in at least one point.

Herbert Simon, 1954

Fixed-point argument

Historical arguments about the possibility of public prediction using Brouwer’s fixed point theorem

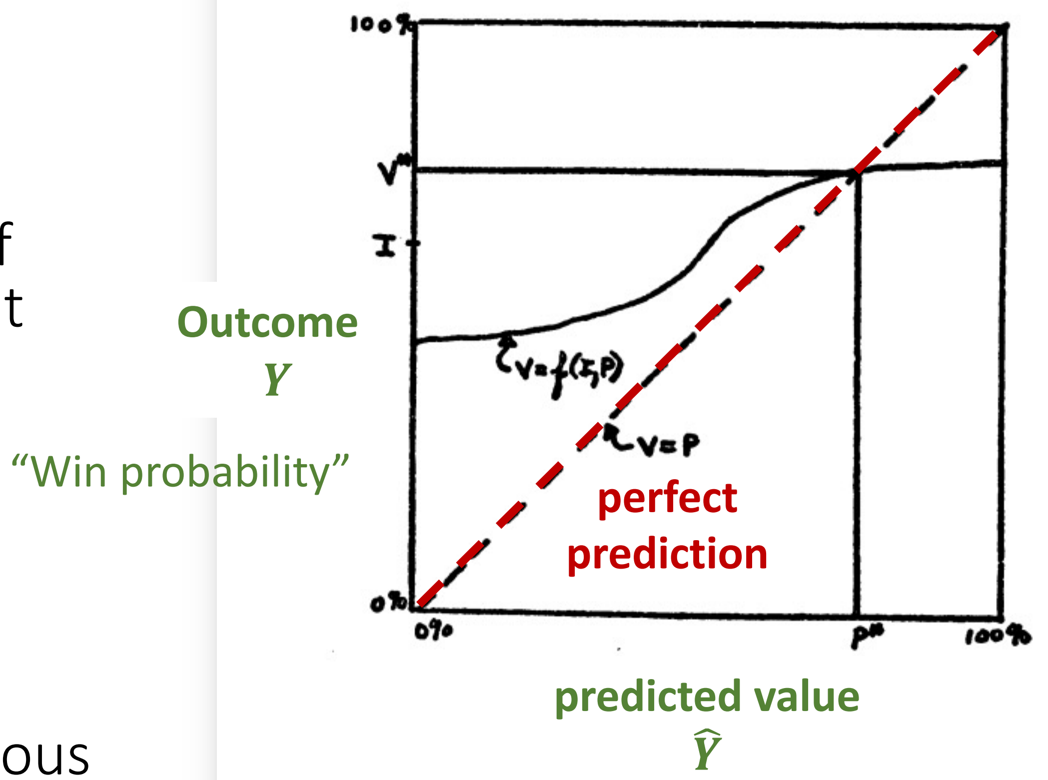
- Simon 1954
- Grunberg, Modigliani 1954

If the response funtion $\mathrm{Y} = R(\hat{Y})$ is continuous then perfect public prediction is possible

Performative stability is a natural analogue of these fixed points in parametric prediction settings

POSSIBILITY OF ELECTION PREDICTIONS 249

FIGURE 2



A formal proof of the theorem will not be given here. It is a classical theorem of topology due to Brouwer (the “fixed-point” theorem), and a non-technical exposition may be found in *What is Mathematics?*.[2] The reader who does not demand a rigorous proof may satisfy himself of the correctness of the theorem by graphical means. Construct a figure like Figure 2, but omit the solid curve. Mark any point on the y-axis between V = 0 per cent and V = 100 per cent; and a second arbitrary point on the vertical line, P = 100 per cent, within the same limits. Now try to connect these two points, without lifting the pencil from the paper, without going outside the limits 0 per cent to 100 per cent for V and P (that is, without going outside the square), and without intersecting the broken line. Since this is impossible, *any* continuous curve relating V and P for the whole range of values 0% ≤ P ≤ 100% must intersect the line V = P in at least one point.

Herbert Simon, 1954

[PZMH20]

When is sensitivity satisfied?

Definition: We say the distribution map $D(\theta)$ is **ϵ-sensitive** if for all θ, θ'

$$W\big(D(\theta), D(\theta')\big) \leq \epsilon \big|\big|\theta - \theta'\big|\big|_2$$

"Similar models lead to similar distributions"

Estimate a binary outcome

$$\hat{Y} = f_\theta(X) = \theta X \qquad X \in [0,1]$$

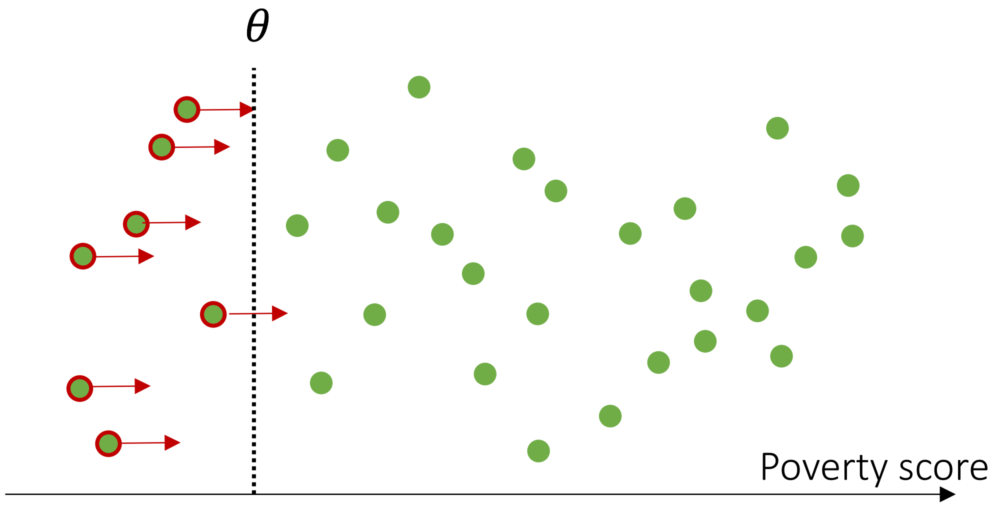
Prediction is self-fulfilling:

sensitivity

$$P(Y = 1|X) = \text{Bernoulli}(\mu(X) + \epsilon\, f_\theta(X))$$

$\epsilon \propto$ strength of effect

Subsidize individuals below threshold



$\epsilon \propto$ fraction of individuals impacted by unit change in θ

[PZMH20]

When is sensitivity satisfied?

<u>Definition</u>: We say the distribution map $D(\theta)$ is **ϵ-sensitive** if for all θ, θ'

$$W\left(D(\theta), D(\theta')\right) \leq \epsilon \left\|\theta - \theta'\right\|_2$$

"Similar models lead to similar distributions"

<u>Estimate a binary outcome</u>

$$\hat{Y} = f_\theta(X) = \theta X \qquad X \in [0,1]$$

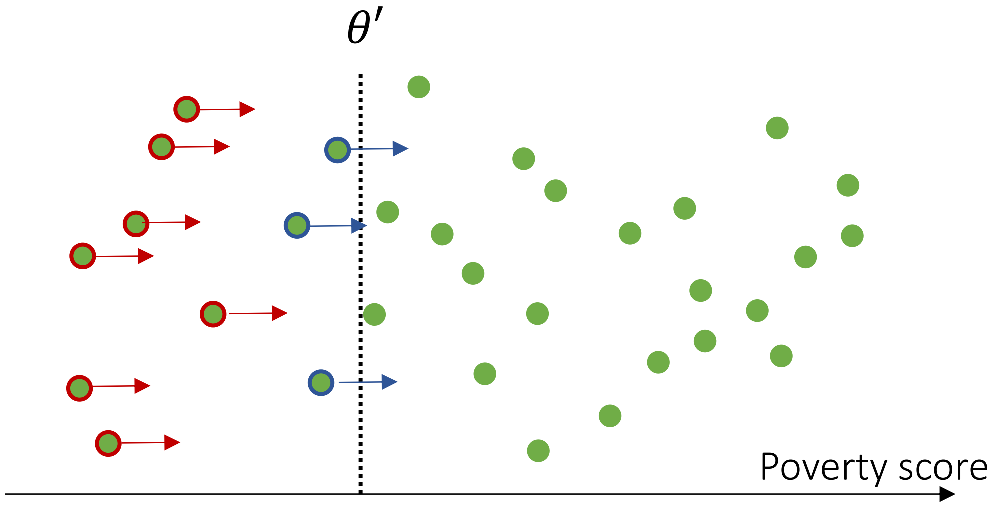
Prediction is self-fulfilling:

sensitivity

$$P(Y = 1|X) = \text{Bernoulli}(\mu(X) + \epsilon\, f_\theta(X))$$

$\epsilon \propto$ strength of effect

<u>Subsidize individuals below threshold</u>



$\epsilon \propto$ fraction of individuals impacted by unit change in θ

[PZMH20]

When is sensitivity satisfied?

<u>Definition</u>: We say the distribution map $D(\theta)$ is **ϵ-sensitive** if for all θ, θ'

$$W\big(D(\theta), D(\theta')\big) \leq \epsilon \big|\big|\theta - \theta'\big|\big|_2$$

"Similar models lead to similar distributions"

<u>Estimate a binary outcome</u>

$\hat{Y} = f_\theta(X) = \theta X \qquad X \in [0,1]$

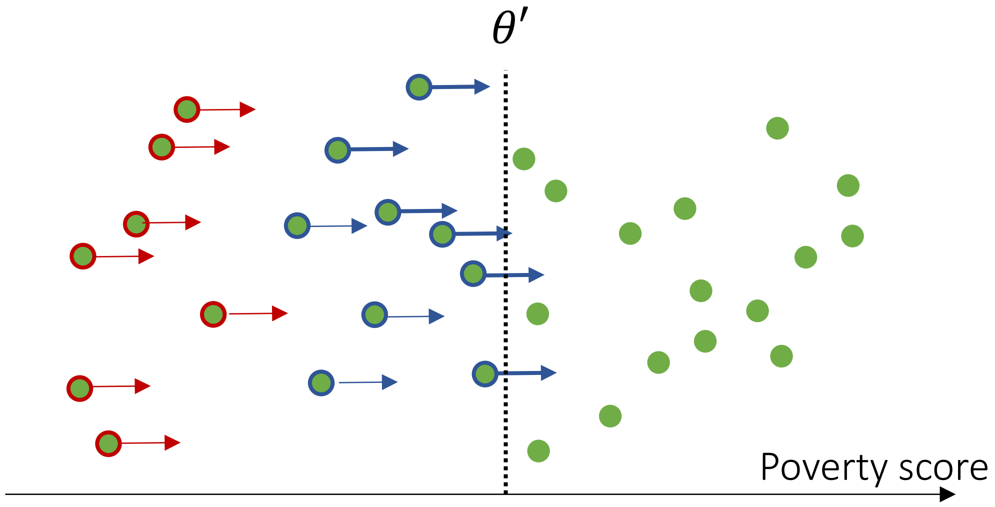
Prediction is self-fulfilling:

sensitivity

$$P(Y = 1|X) = \text{Bernoulli}(\mu(X) + \epsilon\, f_\theta(X))$$

$\epsilon \propto$ strength of effect

<u>Subsidize individuals below threshold</u>



$\epsilon \propto$ fraction of individuals impacted by unit change in θ

Retraining heuristics

Beyond risk minimization

Retraining heuristics as natural fixed point dynamics under performativity

$$\theta_{k+1} \leftarrow \text{argmin}_\theta \text{ Risk}(\theta, D(\theta_k))$$

Empirical risk using samples of $D(\theta_k)$

gradient update $\theta_k - \eta \, \mathrm{E}_{z\sim D(\theta_k)} [\nabla \ell(z; \theta_k)]$

- ERM and repeated gradient descent [PZMH20]
- Stochastic gradient descent [MPZH20, DX23]
- Proximal point methods [DX23]
- Projected gradient descent [WBD21]

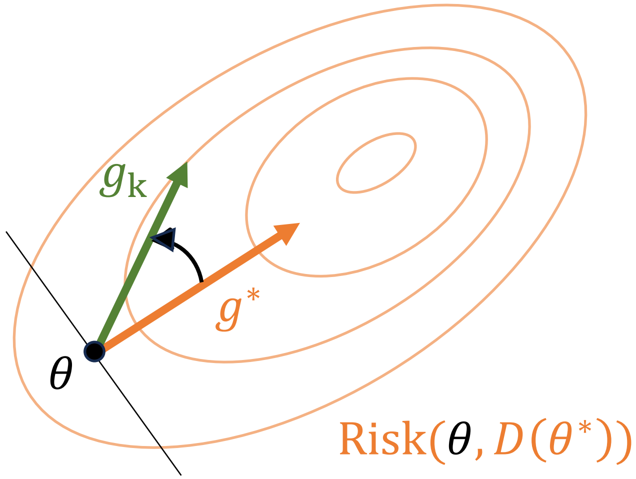
[DX23]

Stochastic gradient descent

- SGD update uses unbiased estimate of the gradient: $g_k(\theta) := \mathrm{E}_{z\sim D_k}[\nabla \ell_\theta(z;\theta)]$

- For small $\epsilon < \gamma/\beta$ the gradient on problem $D(\theta_k)$ is aligned with the gradient on problem $D(\theta^*)$ and never points against the gradient flow:

$$\left|\left|g_k(\theta) - \nabla g^*(\theta)\right|\right| \leq \epsilon\beta\left|\left|\theta_k - \theta^*\right|\right| \quad \rightarrow \quad \cos\left(\measuredangle\left(g_k(\theta), g^*(\theta)\right)\right) \leq \sqrt{1 - \left(\frac{\epsilon\beta}{\gamma}\right)^2}$$



Stochastic gradient descent

- SGD update uses unbiased estimate of the gradient: $g_k(\theta) := \mathrm{E}_{z \sim D_k}[\nabla \ell_\theta(z;\theta)]$

- For small $\epsilon < \gamma/\beta$ the gradient on problem $D(\theta_k)$ is aligned with the gradient on problem $D(\theta^*)$ and never points against the gradient flow:

$$||g_k(\theta) - \nabla g^*(\theta)|| \leq \epsilon\beta||\theta_k - \theta^*|| \quad \rightarrow \quad \cos\left(\measuredangle\left(g_k(\theta), g^*(\theta)\right)\right) \leq \sqrt{1 - \left(\frac{\epsilon\beta}{\gamma}\right)^2}$$

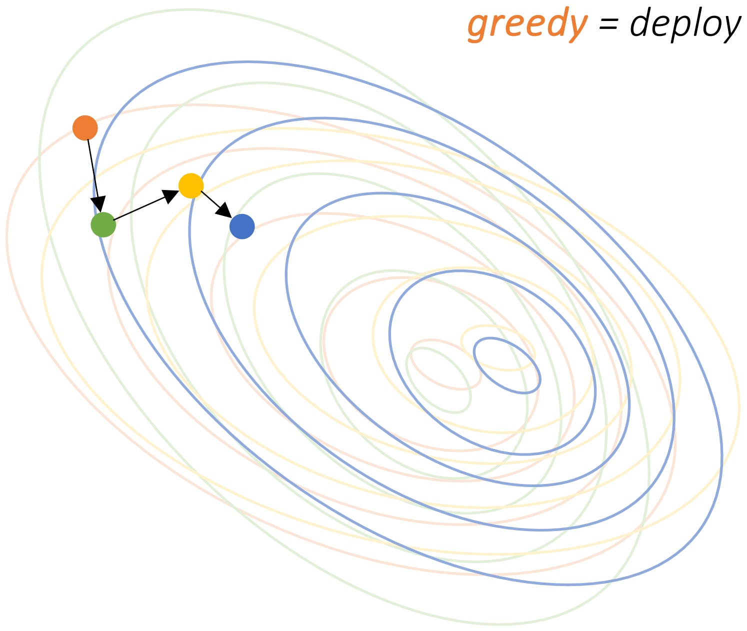
- Choosing step size such that gradient variance decreases sufficiently quickly as we approach stability implies classical $O\left(\frac{1}{k}\right)$ convergence rate

SGD under performativity $\approx$ perturbed SGD at equilibrium distribution $D(\theta^*)$

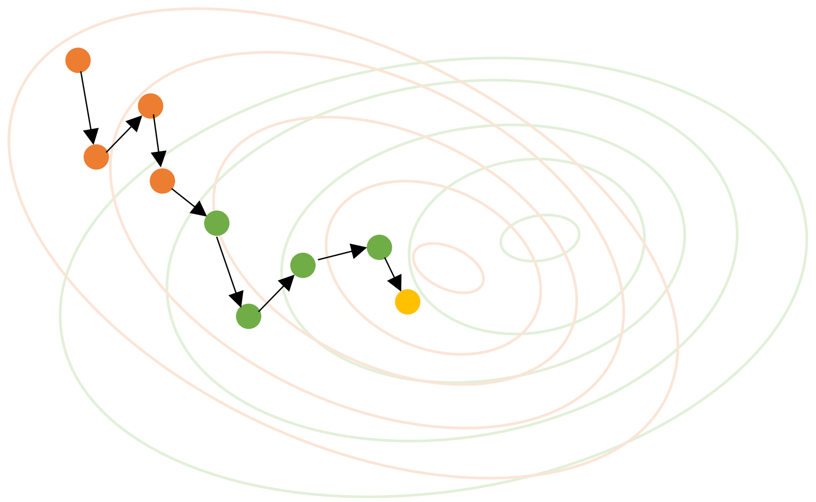
[MPZH20]

Greedy vs lazy

greedy = deploy every step

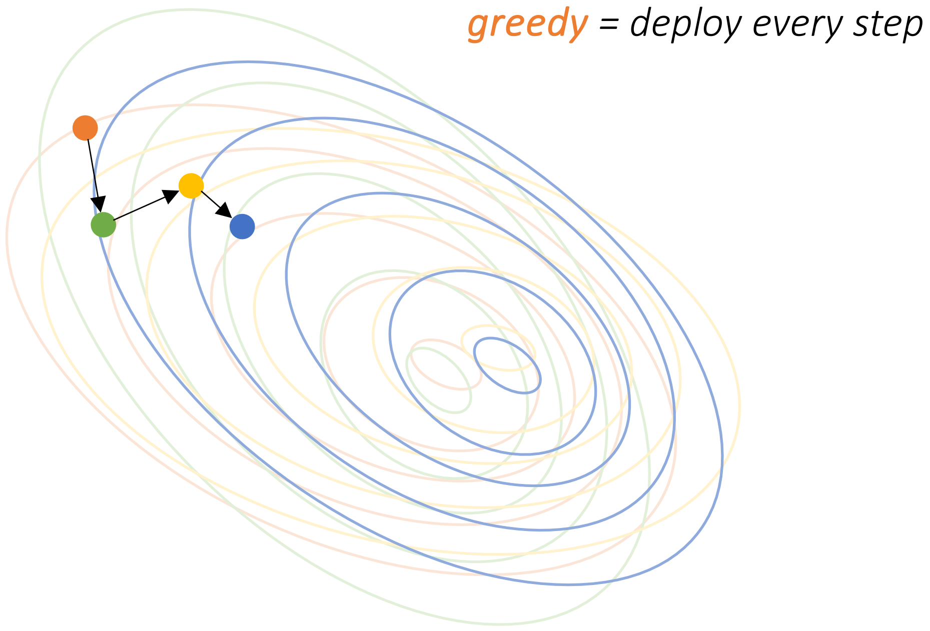


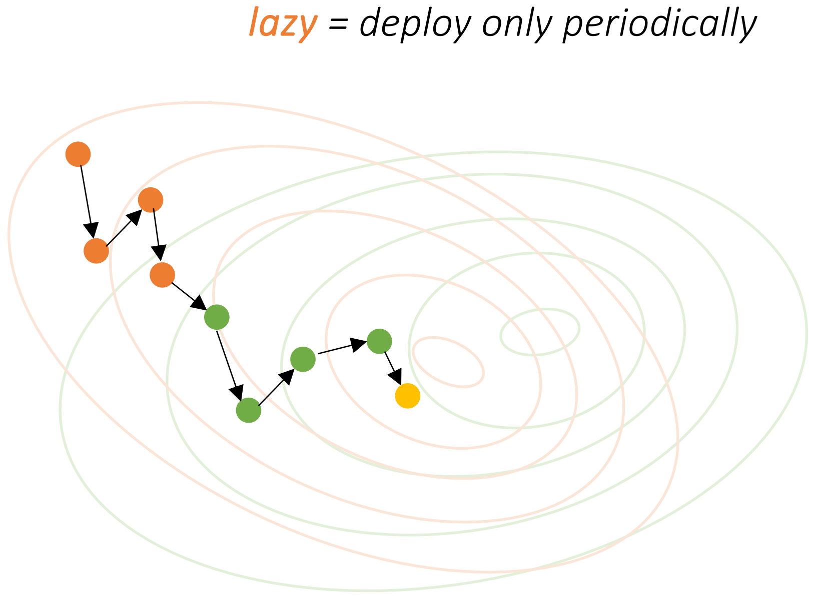
lazy = deploy only periodically



[MPZH20]

Greedy vs lazy



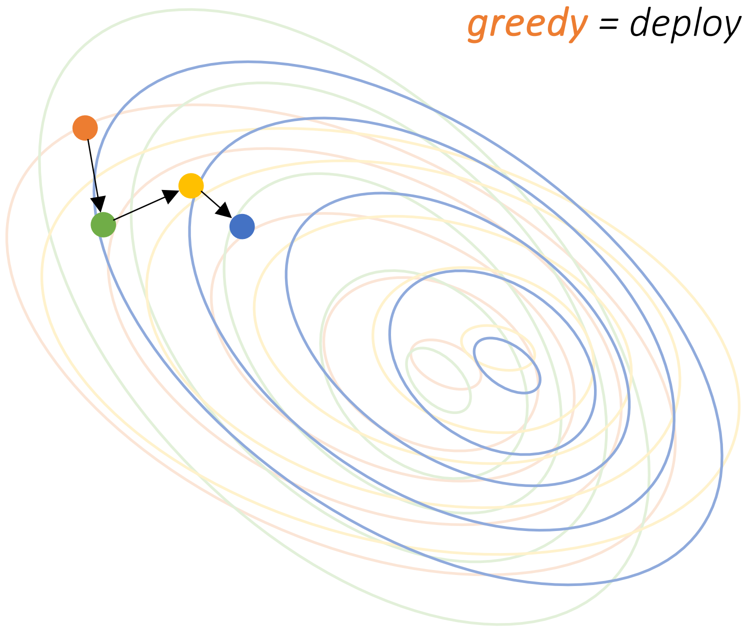


Step size for greedy deploy globally decreasing and more conservative as ϵ grows

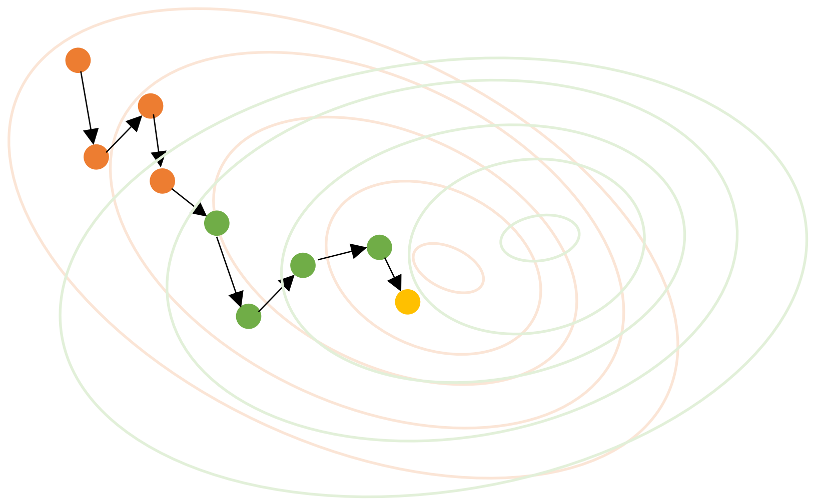
Step size for lazy deploy locally decreasing between deployments and independent of ϵ

Greedy vs lazy

greedy = deploy every step



lazy = deploy only periodically



Step size for greedy deploy globally decreasing and more conservative as ϵ grows

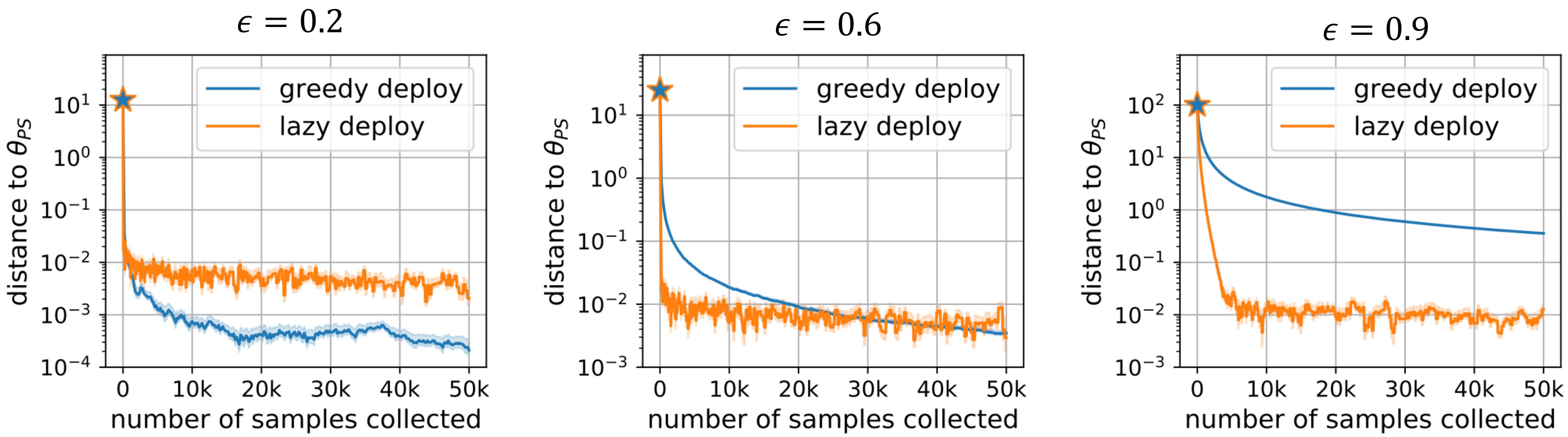
Step size for lazy deploy locally decreasing between deployments and independent of ϵ

Which one works better?

[MPZH20]

Greedy vs lazy

Setup: Mean estimation $z \sim N(\mu + \epsilon\theta, \sigma^2)$ using $\ell(z, \theta) = \frac{1}{2}(z - \theta)^2$



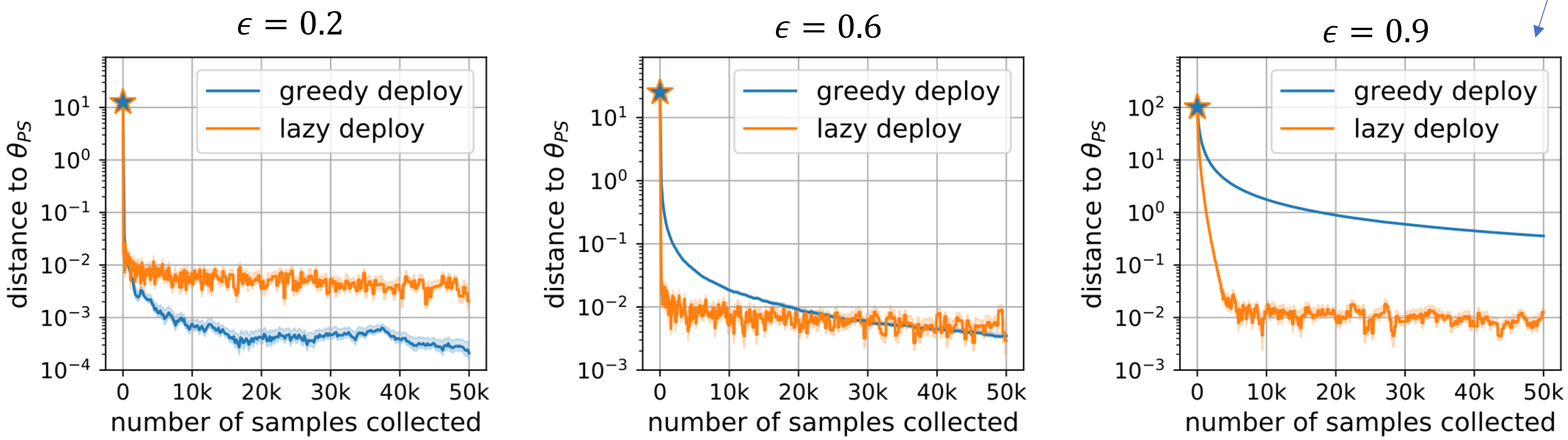
- Greedy deploy: Better if performativity is weak
- Lazy deploy: Better at dealing with strong shifts and poor initialization

[MPZH20]

Greedy vs lazy

deployments
greedy: 50K
lazy: 200

Setup: Mean estimation $z \sim N(\mu + \epsilon\theta, \sigma^2)$ using $\ell(z,\theta) = \frac{1}{2}(z-\theta)^2$



- Greedy deploy: Better if performativity is weak
- Lazy deploy: Better at dealing with strong shifts and poor initialization
- Practical tradeoff between sample collection and deployment costs

[MMG23, LW24]

Stochastic optimization under nonconvexity

Performatively stable points are stationary points: $\mathrm{E}_{z \sim D(\theta^*)}\,[\nabla \ell(z; \theta^*)] = 0$
Stationarity makes sense even with nonconvex losses!

More generally: θ^* is δ-stationary performatively stable if

$$|| \, \mathrm{E}_{z \sim D(\theta^*)}\,[\nabla \ell(z; \theta^*)] \, ||^2 \leq \delta$$

Stochastic optimization under nonconvexity

Performatively stable points are stationary points: $\mathrm{E}_{z\sim D(\theta^*)}\,[\nabla \ell(z;\theta^*)] = 0$
Stationarity makes sense even with nonconvex losses!

More generally: $\theta^*$is δ-stationary performatively stable if

$$\|\, \mathrm{E}_{z\sim D(\theta^*)}\,[\nabla \ell(z;\theta^*)]\,\|^2 \leq \delta$$

<u>Theorem</u> [LW24]:
Assume $D(\theta)$ is ϵ-sensitive and $\ell(z;\theta)$ is Lipschitz in θ and possibly nonconvex. Then,

- greedy deploy satisfies

$$\frac{1}{\mathrm{T}}\sum_{t=1}^{T} \|\, \mathrm{E}_{z\sim D(\theta_t)}\,[\nabla \ell(z;\theta_t)]\,\|^2 = O\left(\frac{1}{\sqrt{T}}\right) + O(\epsilon);$$

- lazy deploy with batch size K satisfies

$$\frac{1}{\mathrm{T}}\sum_{t=1}^{T} \|\, \mathrm{E}_{z\sim D(\theta_t)}\,[\nabla \ell(z;\theta_t)]\,\|^2 = O\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{K}}\right) + O\left(\frac{\epsilon}{K}\right).$$

Performative stability and retraining: recap

- Performative stability as a natural equilibrium concept of retraining
- Retraining heuristics converge to stable points if problem is close to static
- Online vs offline updates as a new design choice for stochastic optimization

Next: Performative optimality

Performative optimality

Performative optimality

Under performativity, after deploying θ the learner experiences performative risk

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

Performative optimality

Under performativity, after deploying θ the learner experiences performative risk

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

Performative optimality:

$$\theta_{\mathrm{PO}} = \mathrm{argmin}_{\theta}\ \mathrm{PR}(\theta) = \mathrm{argmin}_{\theta}\ \mathrm{Risk}(\theta, D(\theta))$$

lowest possible risk after deployment

Performative optimality

Under performativity, after deploying θ the learner experiences performative risk

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

Performative optimality:

$$\theta_{\mathrm{PO}} = \mathrm{argmin}_{\theta}\ \mathrm{PR}(\theta) = \mathrm{argmin}_{\theta}\ \mathrm{Risk}(\theta, D(\theta))$$

lowest possible risk after deployment

Performative stability:

$$\theta_{\mathrm{PS}} = \mathrm{argmin}_{\theta}\ \mathrm{Risk}(\theta, D(\theta_{\mathrm{PS}}))$$

Do stable points have low risk after deployment?

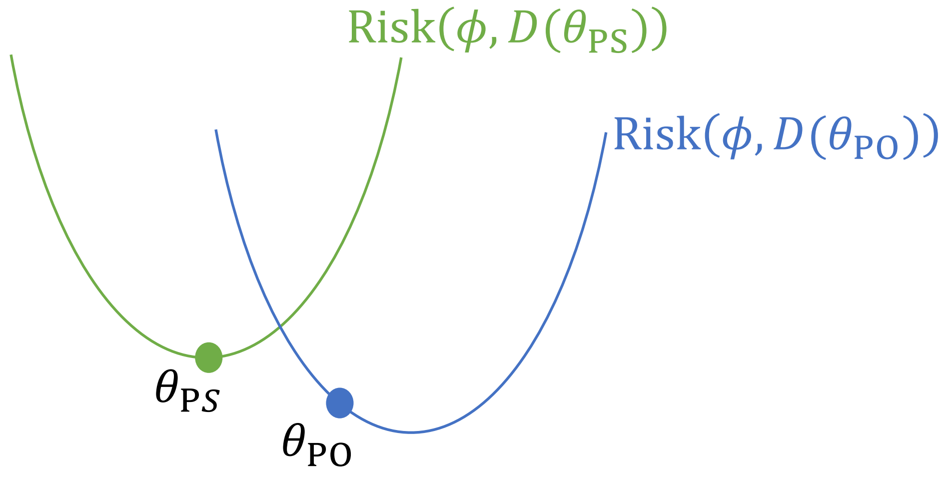
Stability and performative risk

Performative stability: on its own distribution $D(\theta_{\mathrm{PS}})$, θ_{PS} looks optimal

$$\mathrm{PR}(\theta) \coloneqq \underbrace{\mathrm{Risk}\big(\theta, D(\theta)\big) - \min_\phi \mathrm{Risk}\big(\phi, D(\theta)\big)}_{=\,0 \text{ for } \theta\,=\,\theta_{\mathrm{PS}}} + \underbrace{\min_\phi \mathrm{Risk}\big(\phi, D(\theta)\big)}_{\text{“easiness” of } D(\theta)}$$

→ θ_{PS} approximately optimal if it induces ”easy” distribution

Not always true! Stable points can even maximize $\mathrm{PR}(\theta)$



Stability and performative risk

ϵ-sensitive

Example:

$$D(\theta)\colon\ \mathrm{X} \sim \mathrm{Unif}(\{-1,+1\}),\ \mathrm{Y|X} \sim \mathrm{Bern}(0.5 + \epsilon\theta X)$$

$\frac{\gamma}{\beta} = 1$

$$\ell((\mathrm{x},\mathrm{y});\theta) = (y - f_\theta(\mathrm{x}))^2 \text{ where } f_\theta(\mathrm{x}) = \theta x + 0.5$$

$\epsilon \leq 1$ → retraining converges to stable point $\theta_{\mathrm{PS}} = 0$

A direct calculation shows: $\mathrm{PR}(\theta) = 0.25 + (1 - 2\epsilon)\theta^2$

Non-convex for $\epsilon > 0.5$! For $\epsilon \in \left(\frac{\gamma}{2\beta}, \frac{\gamma}{\beta}\right)$ stable point maximizes $\mathrm{PR}(\theta)$

Optimizing the performative risk

Difficulties:

- no guarantee of convexity even is loss is convex

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

only convex in first argument

- no gradient access

$$\nabla \mathrm{PR}(\theta) = \mathrm{E}_{z \sim D(\theta)}\, [\nabla \ell(\mathrm{z}; \theta)] + \underbrace{\mathrm{E}_{z \sim D(\theta)}\, [\ell(\mathrm{z}; \theta) \nabla \log \mathrm{p}_{\theta}(\mathrm{z})]}_{\text{Distribution map is unknown!}}$$

Optimizing the performative risk

Difficulties:

- no guarantee of convexity even is loss is convex

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

only convex in first argument

- no gradient access

$$\nabla \mathrm{PR}(\theta) = \mathrm{E}_{z\sim D(\theta)}\ [\nabla \ell(\mathrm{z};\theta)] + \underbrace{\mathrm{E}_{z\sim D(\theta)}\ [\ell(\mathrm{z};\theta)\nabla \log \mathrm{p}_\theta(\mathrm{z})]}_{\text{Distribution map is unknown!}}$$

For $t = 1, \dots, T$:

- **Deploy model θ_t**
- **Collect data $\mathrm{z}_t{}^1, \dots, z_t^m \sim D(\theta_t)$**

Compute $\hat{\theta}_{\mathrm{PO}}$ based on $S = \{\theta_t, z_t^i\}_{t,i}$

We need to collect data from multiple deployments of $\theta_1, \dots, \theta_T$

Model-free approaches

- No explicit modeling of $D(\theta)$ required
- Based on bandits and other zeroth-order optimization methods
- Convergence relies on general regularity conditions (e.g. convexity, smooth distribution shifts, etc)
- Generally slow convergence

Model-based approaches

- Incorporate model of $D(\theta)$
- Based on economic models (e.g. utility-maximizing agents), other models from domain knowledge, etc
- Convergence relies on model correctness or degree of model misspecification
- Typically fast convergence

Model-free vs model-based: example

For $t = 1, \ldots, T$:

- Deploy model θ_t
- Collect data $z_t^1, \ldots, z_t^m \sim D(\theta_t)$

Compute $\hat{\theta}_{\mathrm{PO}}$ based on $S = \{\theta_t, z_t^i\}_{t,i}$

$f_\theta(x) = x^T\theta \qquad \ell((x,y);\theta) = (y - f_\theta(x))^2$

Model-free:

$$\widehat{\mathrm{PR}}(\theta_t) = \frac{1}{m}\sum_{i=1}^{m} \ell(z_t^i;\theta_t)$$

$$\hat{\theta}_{\mathrm{PO}} = \mathrm{argmin}_{\theta\in\{\theta_1,\ldots,\theta_T\}}\, \widehat{\mathrm{PR}}(\theta)$$

Model-based:

We model the data-generating process:

- agents manipulate features to maximize the prediction:

$$x = \mathrm{argmax}_x\ \gamma \cdot x^T\theta - \frac{1}{2}||x - x_0||^2$$

- agents can manipulate features, not label

utility-cost tradeoff

This is a distribution map model $D_\gamma(\theta)$

Fit $\hat{\gamma}$ using S and let $\hat{\theta}_{\mathrm{PO}} = \mathrm{argmin}_\theta\, \mathrm{Risk}(\theta, D_{\hat{\gamma}}(\theta))$

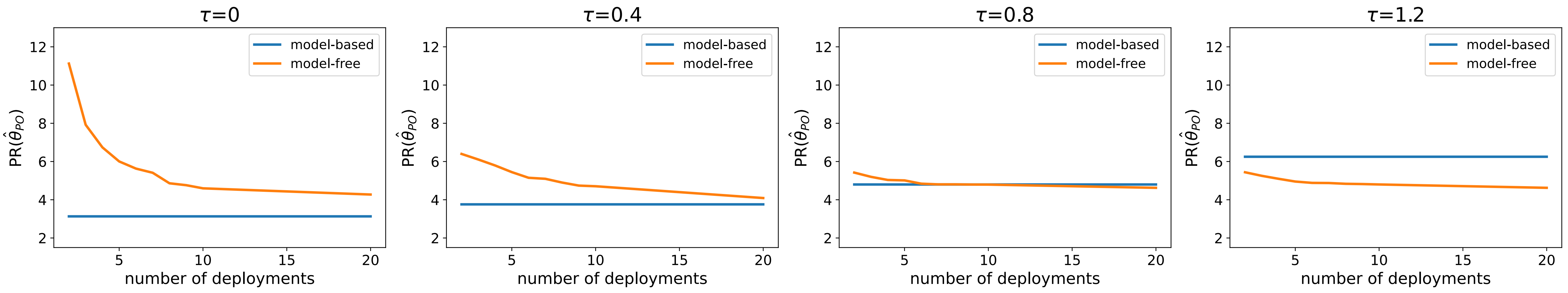
What is the tradeoff?

Model-free vs model-based: example

Suppose model is τ-misspecified: in addition to features, labels change too

$$y = y_0 + \tau \cdot x^T \theta$$

For example, if feature manipulations have a causal effect on true label



General tradeoff

Theorem [LZ24] (informal):

$$\mathrm{PR}(\hat{\theta}_{\mathrm{PO}}) - \mathrm{PR}(\theta_{\mathrm{PO}}) \leq \text{misspecification error} + \text{statistical error}$$

error due to modeling of the distribution map

error due to having finite deployments and finite data

General tradeoff

Theorem [LZ24] (informal):

$$\mathrm{PR}\left(\hat{\theta}_{\mathrm{PO}}\right) - \mathrm{PR}(\theta_{\mathrm{PO}}) \leq \text{misspecification error} + \text{statistical error}$$

error due to modeling of the distribution map

error due to having finite deployments and finite data

Model-free: $\mathrm{PR}\left(\hat{\theta}_{\mathrm{PO}}\right) - \mathrm{PR}(\theta_{\mathrm{PO}}) \leq \underbrace{\text{misspecification error}}_{0} + \underbrace{\text{statistical error}}_{\text{large}}$

Model-based: $\mathrm{PR}\left(\hat{\theta}_{\mathrm{PO}}\right) - \mathrm{PR}(\theta_{\mathrm{PO}}) \leq$ misspecification error + statistical error

depends on domain knowledge, complexity of true map, etc

small, often $\tilde{O}\left(\frac{1}{\sqrt{T}}\right)$

Model-free performative optimization

Convexity of the performative risk

<u>Theorem</u> [MPZ21]:

If the loss is γ-strongly convex and β-smooth in the data and the distribution map is ϵ-Lipschitz and sufficiently regular*, then $\mathrm{PR}(\theta)$ is guaranteed to be convex if and only if $\epsilon < \frac{\gamma}{2\beta}$.

*e.g. distributions obtained by translation and rescaling:

$z \sim D(\theta) \Leftrightarrow z = \Sigma(\theta)\, z_0 + \mu(\theta)$ for linear $\Sigma(\theta), \mu(\theta)$

(see [MPZ21] for details)

Convexity of the performative risk

Theorem [MPZ21]:

If the loss is γ-strongly convex and β-smooth in the data and the distribution map is ϵ-Lipschitz and sufficiently regular*, then $\mathrm{PR}(\theta)$ is guaranteed to be convex if and only if $\epsilon < \frac{\gamma}{2\beta}$.

*e.g. distributions obtained by translation and rescaling:

$z \sim D(\theta) \Leftrightarrow z = \Sigma(\theta)\, z_0 + \mu(\theta)$ for linear $\Sigma(\theta), \mu(\theta)$ (see [MPZ21] for details)

If $\mathrm{PR}(\theta)$ is convex, we can use derivative-free convex optimization [FKM04]

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{d}{\delta} \mathrm{PR}(\theta_t + \delta u_t) u_t, \qquad u_t \sim \mathrm{Unif}(S^{d-1}), \qquad \delta, \eta > 0$$

Converges to θ_{PO} at rate $\mathrm{O}(\sqrt{d} t^{-1/4})$

only queries PR, not its gradient

Beyond convexity?

Optimization with no gradients and no convexity = continuum-arm bandit problem?

- "pull arm" θ_t and observe bandit feedback $\widehat{\mathrm{PR}}(\theta_t)$ with $\mathrm{E}\left[\widehat{\mathrm{PR}}(\theta_t)\right] = \mathrm{PR}(\theta_t)$
- assuming only Lipschitzness of PR we can apply Lipschitz bandits [KSU08]

Beyond convexity?

Optimization with no gradients and no convexity = continuum-arm bandit problem?

- "pull arm" θ_t and observe bandit feedback $\widehat{\mathrm{PR}}(\theta_t)$ with $\mathrm{E}\left[\widehat{\mathrm{PR}}(\theta_t)\right] = \mathrm{PR}(\theta_t)$
- assuming only Lipschitzness of PR we can apply Lipschitz bandits [KSU08]

Performative feedback is more informative than bandit feedback!

At every time step we deploy a model θ_t and observe m samples of the induced distribution $D(\theta_t)$

→ faster convergence rates by constructing fine-grained confidence bounds

[JZM22]

Tighter confidence bounds

After deploying θ_t we observe $D(\theta_t)$ (ignoring finite-sample considerations for now)
What do we learn about the performative risk of an unexplored θ_{new}?

$$\mathrm{PR}(\theta_{\text{new}}) - \mathrm{PR}(\theta_t) = \mathrm{Risk}(\theta_{\text{new}}, D(\theta_{\text{new}})) - \mathrm{Risk}(\theta_{\text{new}}, D(\theta_t))$$

uncertainty due to distribution shift

$$+ \mathrm{Risk}(\theta_{\text{new}}, D(\theta_t)) - \mathrm{Risk}(\theta_t, D(\theta_t))$$

uncertainty due to changing predictive model

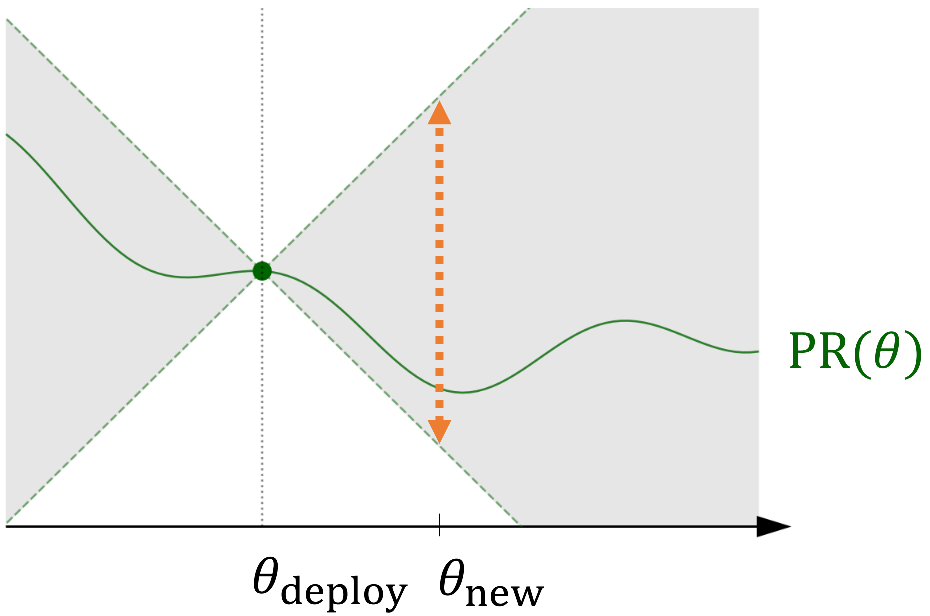
Second term is known because we know the loss and $D(\theta_t)$!

→ We only pay for uncertainty due to distribution shift

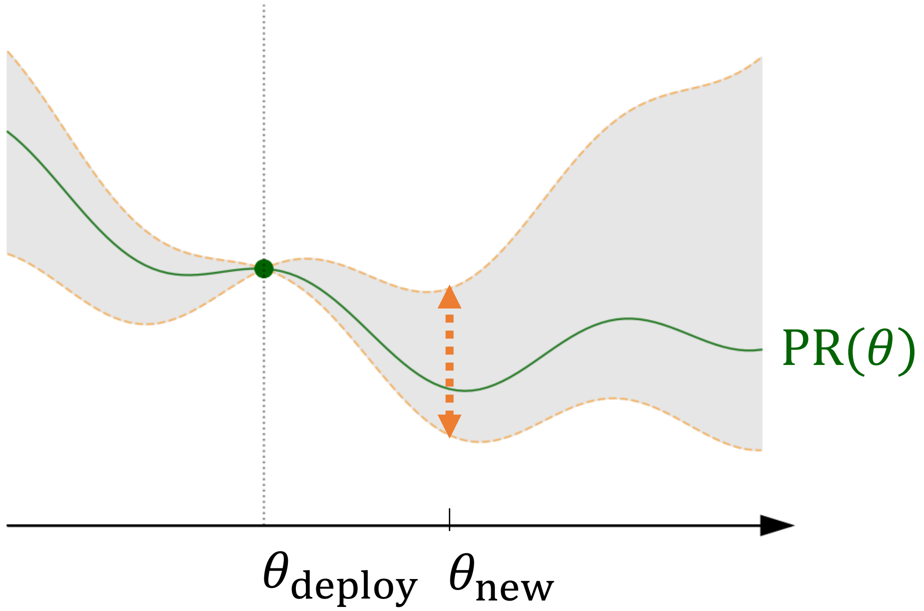
Tighter confidence bounds

[JZM22]

Confidence bound with bandit feedback
(Lipschitz $\mathrm{PR}(\theta)$)



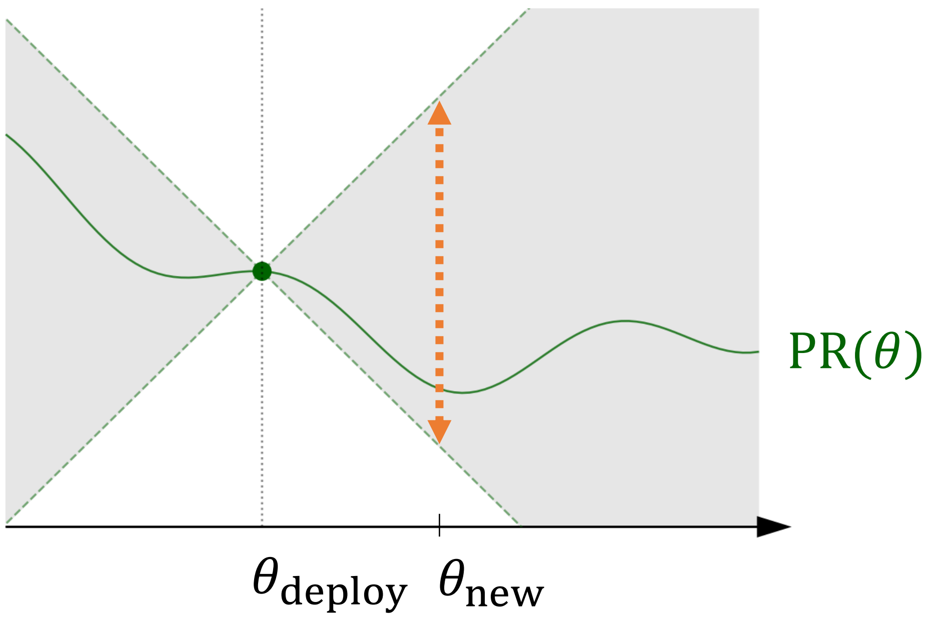
Confidence bound with performative feedback
(Lipschitz $\mathrm{Risk}(\theta, D(\phi))$ in ϕ)



[JZM22]

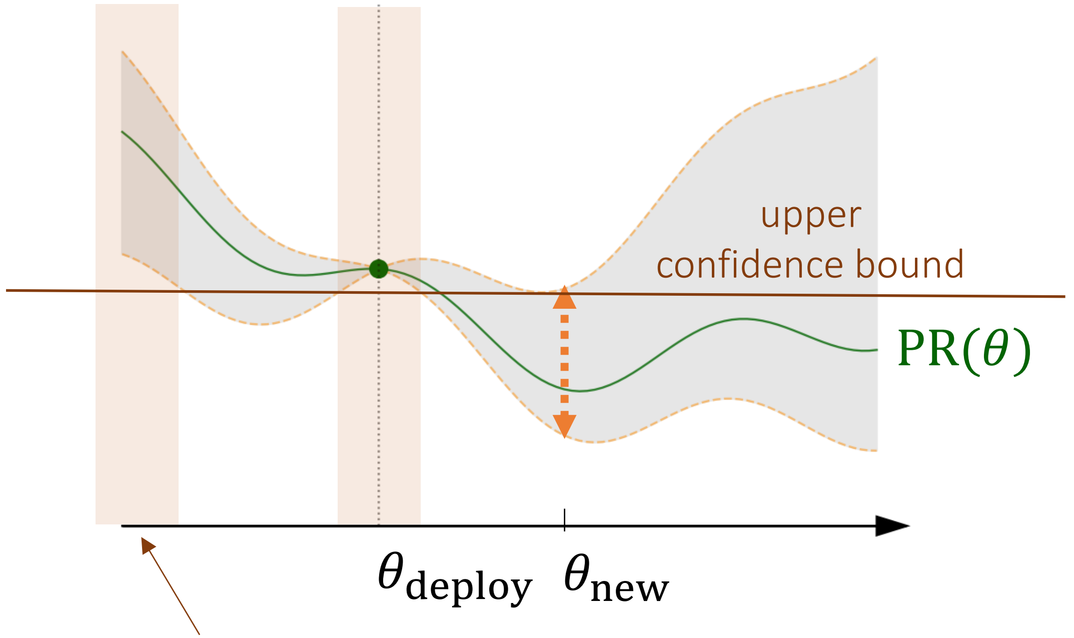
Tighter confidence bounds

Confidence bound with bandit feedback
(Lipschitz $\mathrm{PR}(\theta)$)



Use **successive elimination**
to deal with finite-sample uncertainty [EMM06]

Confidence bound with performative feedback
(Lipschitz $\mathrm{Risk}(\theta, D(\phi))$ in ϕ)



the algorithm can discard regions of the parameter space
that have never been explored!

Performative regret bound

Theorem [JZM22]:
Assume the distribution map $D(\theta)$ is ϵ-sensitive and the loss $\ell(z;\theta)$ is L_z-Lipschitz in z. Then, the **performative confidence bounds** algorithm that after T deployments achieves a regret

$$\mathrm{Reg}(T) = \sum_{t=1}^{T} \mathrm{E}\left[\mathrm{PR}(\theta_t)\right] - \mathrm{PR}(\theta_{\mathrm{PO}}) = \tilde{O}\left(\sqrt{T} + T^{\frac{d+1}{d+2}}(L_z\epsilon)^{\frac{d}{d+2}}\right)$$

where d denotes the "zooming dimension" of the problem.

Baseline: Lipschitz bandits [KSU08]: $\mathrm{Reg}(T) = \tilde{O}\left(T^{\frac{d'+1}{d'+2}}\, L^{\frac{d'}{d'+2}}\right)$

L Lipschitz constant PR
$d' \geq d$ zooming dimension

Benefits of performative confidence bounds:

- regret bound scales with ϵ (no distribution shift → fast rate)
- as $\epsilon \to 0$ bound scales as $\tilde{O}(\sqrt{T})$ (no dimension dependence)
- no assumption on loss as a function of θ

Model-based performative optimization

Model-based approaches in a nutshell

Basic idea: learn a model of $D(\theta)$ and plug it into performative risk

$\widehat{D}(\theta)$ - fitted model of $D(\theta)$ based on collected data

Then we can solve

$$\hat{\theta}_{\mathrm{PO}} = \mathrm{argmin}_{\theta}\, \mathrm{Risk}(\theta, \widehat{D}(\theta))$$

$\widehat{D}(\theta)$ identified correctly → we can find the optimal solution $\hat{\theta}_{\mathrm{PO}}$ offline for any loss function!

related to **omniprediction** [GKRSW22, KP23]

Microfoundations

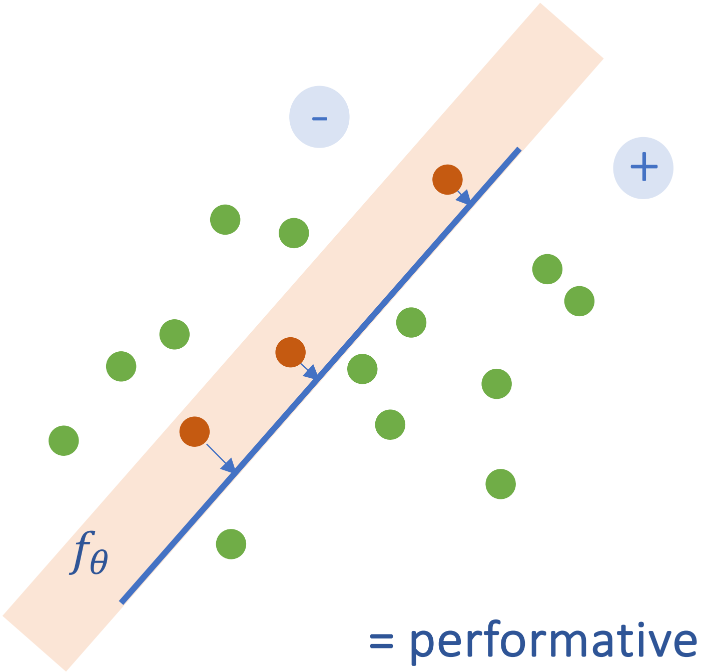
Modeling $D(\theta)$ ("macro level") in terms of the behavior of individual agents in the population ("micro level")

Microfoundations

Modeling $D(\theta)$ ("macro level") in terms of the behavior of individual agents in the population ("micro level")

Example: strategic classification [HMPW16]

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



Rational-agent model

$$x(\theta) = \operatorname*{argmax}_{x} \; \gamma \, f_\theta(x) - \mathrm{cost}(x_0, x)$$

gain of positive clasification

cost of feature manipulation

$D_\gamma(\theta)$ is "best response map" over $(x(\theta), y)$

strategic clasification
= performative prediction + microfoundation model

Location families

"Macro level" model:

$$z \sim D(\theta) \Leftrightarrow z_0 + \mu_*^T \theta, \qquad z_0 \sim D_0$$

unknown
$\mu_* \in R^{d \times m}$

"base" distribution,
$z_0 \in R^m$

Theorem [JZM22]:
There exists an algorithm that after T deployments achieves a regret

$$\mathrm{Reg}(T) = \sum_{t=1}^{T} \mathrm{E}\left[\mathrm{PR}(\theta_t)\right] - \mathrm{PR}(\theta_{\mathrm{PO}}) = \tilde{O}\left(\sqrt{T}\max\{d, \sqrt{dm}\}\right).$$

Bandit approach: $\mathrm{Reg}(T) = \tilde{O}\left(\sqrt{T} + T^{\frac{d+1}{d+2}}(L_z\epsilon)^{\frac{d}{d+2}}\right)$

fast rate regardless of the strength
of performative effects

Location families

"Macro level" model:

$$z \sim D(\theta) \Leftrightarrow z_0 + \mu_*^T \theta, \qquad z_0 \sim D_0$$

unknown
$\mu_* \in R^{d \times m}$

"base" distribution,
$z_0 \in R^m$

Satisfied in strategic classification model:

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

$$f_\theta(x) = \theta^T x$$

$$\operatorname{cost}(x_0, x) = \frac{1}{2}(\mathrm{x} - \mathrm{x}_0)^{\mathrm{T}} \Lambda (x - x_0)$$

<u>Theorem</u> [JZM22]:
There exists an algorithm that after T deployments achieves a regret

$$\operatorname{Reg}(T) = \sum_{t=1}^{T} \mathrm{E}\,[\mathrm{PR}(\theta_t)] - \mathrm{PR}(\theta_{\mathrm{PO}}) = \tilde{O}\left(\sqrt{T} \max\{d, \sqrt{dm}\}\right).$$

Bandit approach: $\operatorname{Reg}(T) = \tilde{O}\left(\sqrt{T} + T^{\frac{d+1}{d+2}} (L_z \epsilon)^{\frac{d}{d+2}}\right)$

fast rate regardless of the strength of performative effects

Performative modeling through causal inference

Modeling $D(\theta)$ is fundamentally a causal inference problem

$D(\theta)$ is the "effect" of deploying model θ

$$\boxed{\text{learning } D(\theta) \quad \Leftrightarrow \quad \text{causal identification}}$$

Performative modeling through causal inference

Modeling $D(\theta)$ is fundamentally a causal inference problem

$D(\theta)$ is the "effect" of deploying model θ

learning $D(\theta)$ $\Leftrightarrow$ causal identification

Causal identification impossible if θ is fixed!

Example: if Zillow's housing pricing algorithm is fixed, we can't tell $D(\theta)$ and D_{static} apart

Randomizing θ allows identification

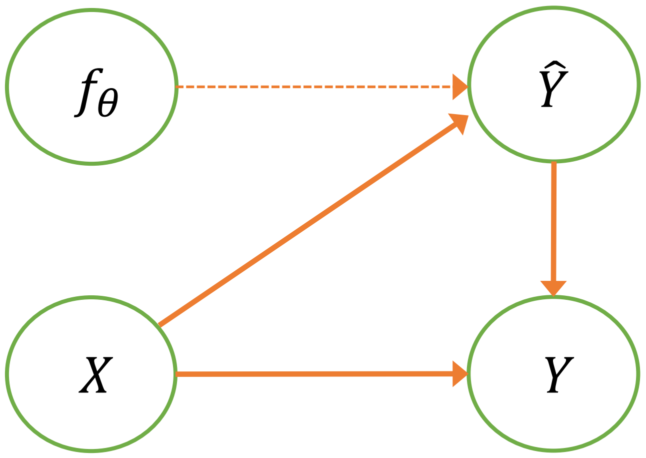
Performative modeling through causal inference

Modeling $D(\theta)$ is fundamentally a causal inference problem

$D(\theta)$ is the “effect” of deploying model θ

learning $D(\theta)$ $\Leftrightarrow$ causal identification

Special case: performative effects mediated by model predictions [MDW22 , KP23]



<u>Key challenge for causal identification</u>

violation of “positivity”: $f_\theta(X)$ often deterministic!

Identification achieved by

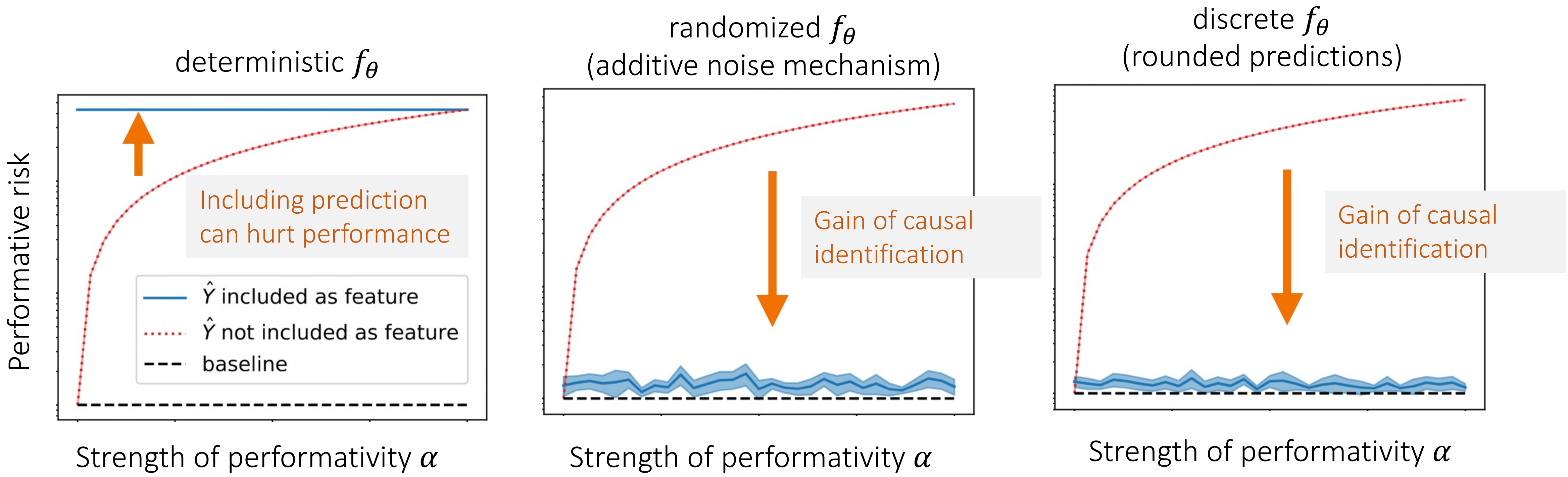
- randomizing predictions
- discrete predictions
- overparameterized predictions

see [MDW22]

[MDW22]

Performative modeling through causal inference

Semi-synthetic experiment: predict income on US census data
Performative effects simulated on top of real census data



Performative optimality: recap

- Performative stability can be far from performatively optimal
- Finding performative optima requires exploring different models $\theta_1, \ldots, \theta_T$
- Performative optimization can be done via model-based and model-free approaches
- Model-free approaches make fewer assumptions and converge slowly; model-based approaches make stronger assumptions and converge fast, but they can suffer from modeling biases

Next: Extensions

Extensions of the framework and connections

Performative prediction framework: recap

After deploying θ the learner experiences performative risk

$$\mathrm{PR}(\theta) \coloneqq \mathrm{Risk}\big(\theta, D(\theta)\big) = E_{Z\sim D(\theta)}\ell(z;\theta)$$

Performative optimality:

$$\theta_{\mathrm{PO}} = \mathrm{argmin}_{\theta}\ \mathrm{PR}(\theta) = \mathrm{argmin}_{\theta}\ \mathrm{Risk}(\theta, D(\theta))$$

Performative stability:

$$\theta_{\mathrm{PS}} = \mathrm{argmin}_{\theta}\ \mathrm{Risk}(\theta, D(\theta_{\mathrm{PS}}))$$

[BHK22, RRDF22, LW22, IZY22]

Stateful distribution shifts

After deployment, the environment does not respond immediately

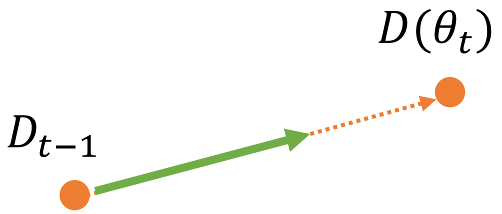
It **remembers past deployments** and **gradually** approaches $D(\theta)$

Distribution at time t:

$$D_t = (1-\delta) \cdot D_{t-1} + \delta \cdot D(\theta_t)$$

how fast environment forgets past deployments

model deployed at time t



Model θ deployed over multiple steps
→ distribution close to $D(\theta)$

[NFDFR23, PY23, WYW23]

Multiplayer performative prediction

Performativity arises in the context of n competing decision-makers

Risk of decision-maker $\boldsymbol{i}$ depends on all decisions:

$$\mathrm{PR_i}(\boldsymbol{\theta}) = \mathrm{E}_{z \sim D_i(\boldsymbol{\theta})}\ [\ell(z; \theta_i)], \qquad \text{where } \boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$$

Example: multiple navigation apps predict travel time; people respond by considering multiple predictions

Main solution concept: Nash equilibrium $\boldsymbol{\theta^*} = (\theta_1^*, \dots, \theta_n^*)$

$$\theta_i^* \in \mathrm{argmin}_{\theta_i}\ \mathrm{PR_i}(\theta_1^*, \dots, \theta_i, \dots, \theta_n^*)$$

Considerations in fairness

How do we choose ℓ? In a performative context, ℓ shouldn't just measure predictive accuracy! We want to optimize some notion of welfare in equilibrium

Neglecting performative feedback can amplify unfairness and polarization over time, even if starting from a fair model [LDRSH18, HSNL18, JXLZ24]

The performative risk captures welfare through the dependence on $D(\theta)$

For example, we can choose: $\ell((x,y);\theta) = \underbrace{\ell^{\mathrm{SL}}((x,y);\theta)}_{\text{supervised learning loss}} - \frac{\lambda}{2}\underbrace{y^2}_{\text{promotes large values of the label}}$

Loss can even depend only on data! e.g. $\ell((x,y);\theta) = y$ is a valid loss, because $\mathrm{PR}(\theta) = E_{D(\theta)}[y]$

Shift in perspective:
Focus on those impacted by performativity

Performative risk

Two levers to achieve small risk

$$\mathrm{Risk}(\theta, D(\theta)) = \underbrace{\mathrm{E}_{(x,y)\sim D(\theta)}}_{\text{Steer data}} \underbrace{[\,\mathrm{loss}((x,y);\theta)\,]}_{\text{Fit patterns given data}}$$

Steer data

Fit patterns given data

Finding optimal points = minimize $\mathrm{PR}(\theta) := \mathrm{Risk}\big(\theta, D(\theta)\big)$

Steering as a major concern for competition

EU vs Google

“[T]he General Court [of the European Union] finds that, by favouring its own comparison shopping service on its general results pages […] by means of ranking algorithms, Google departed from competition on the merits.”

European Commission

Traditionally market power enabled a firm to set prices,
in digital markets power enables firms to steer users and drive consumption

[HJM22]

Performative power

Quantifying the strength of performativity as a notion of power

Performative power: The ability to impact individual outcomes through algorithmic actions, on average across a population of users

$$\mathrm{P} := \sup_{\text{action } f \in \mathrm{F}} \frac{1}{|U|} \sum_{u \in U} \mathrm{E}[\mathrm{dist}(Z_u(f_0), Z_u(f))]$$

choice of algorithmic action

current outcome for user u

counterfactual outcome for user u under f

[HJM22]

Performative power

Quantifying the strength of performativity as a notion of power

Performative power: The ability to impact individual outcomes through algorithmic actions, on average across a population of users

Average treatment effect

$$P := \sup_{\text{action } f \in F} \boxed{\frac{1}{|U|} \sum_{u \in U} \mathrm{E}[\mathrm{dist}(Z_u(f_0), Z_u(f))]}$$

choice of algorithmic action

current outcome for user u

counterfactual outcome for user u under f

A causal inference problem

Through performativity we can relate the abstract concept of power to a causal inference problem.

How much would the average outcome change if the firm were to deploy a different model?

click on a website/consumption of a service

A causal inference problem

Through performativity we can relate the abstract concept of power to a causal inference problem.

How much would the average outcome change if the firm were to deploy a different model?

click on a website/consumption of a service

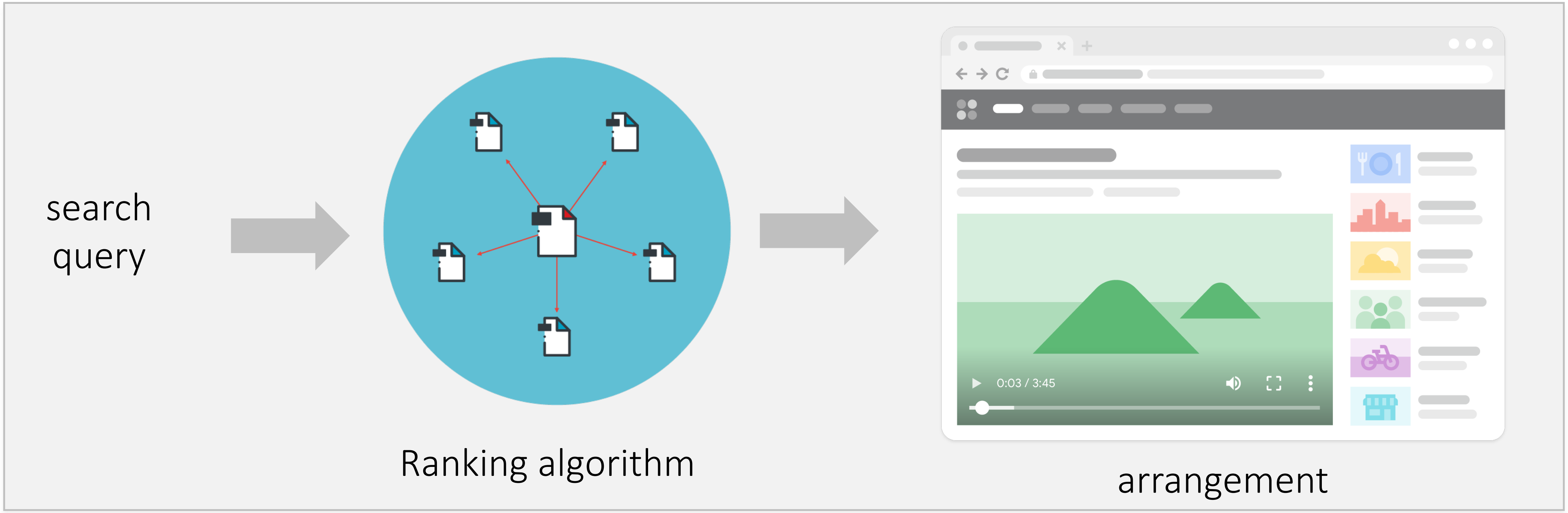
The mechanism behind performativity is complex. It depends on social and economic context, behavioral biases, and many design decisions on how predictions are displayed.

Experimental designs offer a promising avenue!

Performativity gap

How do we measure performative power if we don't have control over the algorithm?

Predictions are displayed through content arrangements!



Performativity gap

How do we measure performative power if we don't have control over the algorithm?

Predictions are displayed through content arrangements!

Performativity gap : $\delta_i(a) = \mathrm{CTR}_i(a) \; - \mathrm{CTR_i}(a_0)$

"Change in click through rate of an item under two different arrangements"

Assume independence across interactions, then performative power across the population of platform participants is bounded as

$$\mathrm{P} \geq \max_{a \in \mathcal{A}} \; \delta_i(a)$$

$\mathcal{A}$ are possible arrangements resulting from $f \; \in F$

Quasi experimental designs

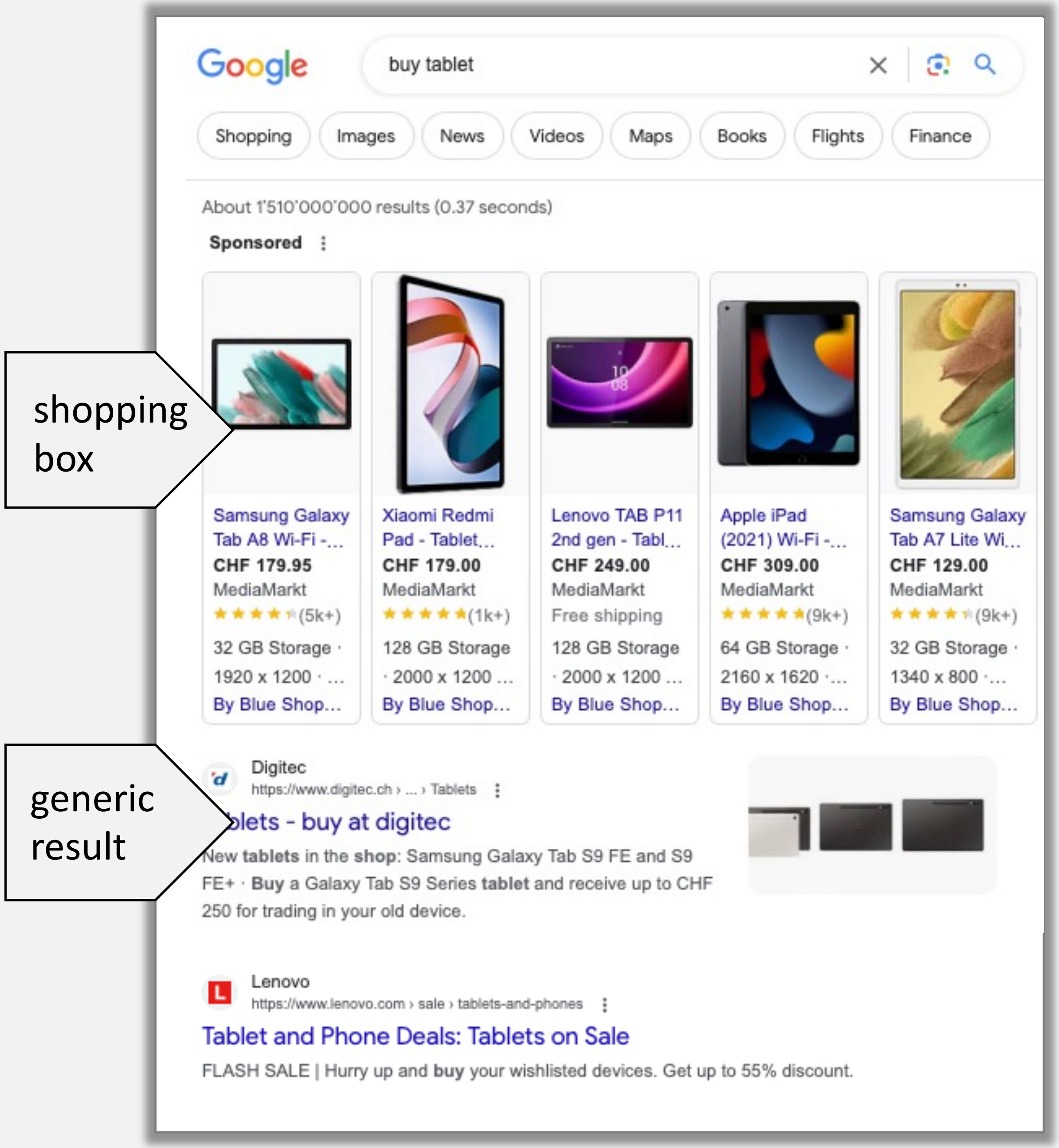
Consider alternative arrangements where items around the decision boundary are swapped.

- *Anderson, Magruder* (2012) “An extra half-star rating [on Yelp] causes restaurants to sell out 19 percentage points (49%) more frequently”

- *Narayanan, Kalyanam* (2015) “Being ranked 2 instead of 1 in Google Ads reduces CTR by 21%”

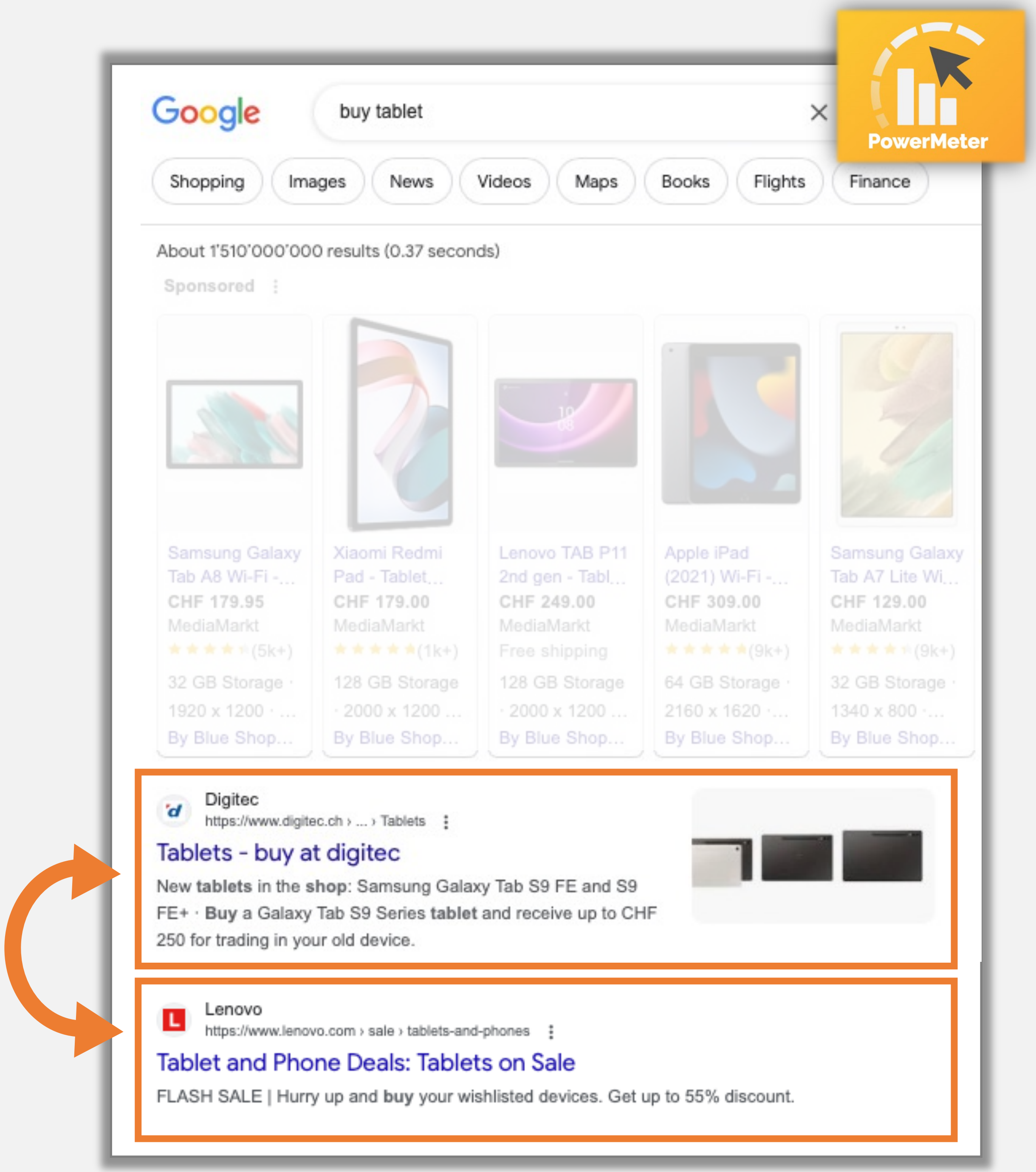
These numbers speak to the performative power of the platform over its participants

Since performativity is context specific, we need to reassess it in each individual case



What is the causal effect of Google's ranking algorithm on user clicks?

- Experiment as gold standard: change the algorithm and inspect effect
- Algorithm is proprietary and complex
- Intervene at the level of display to emulate algorithmic updates



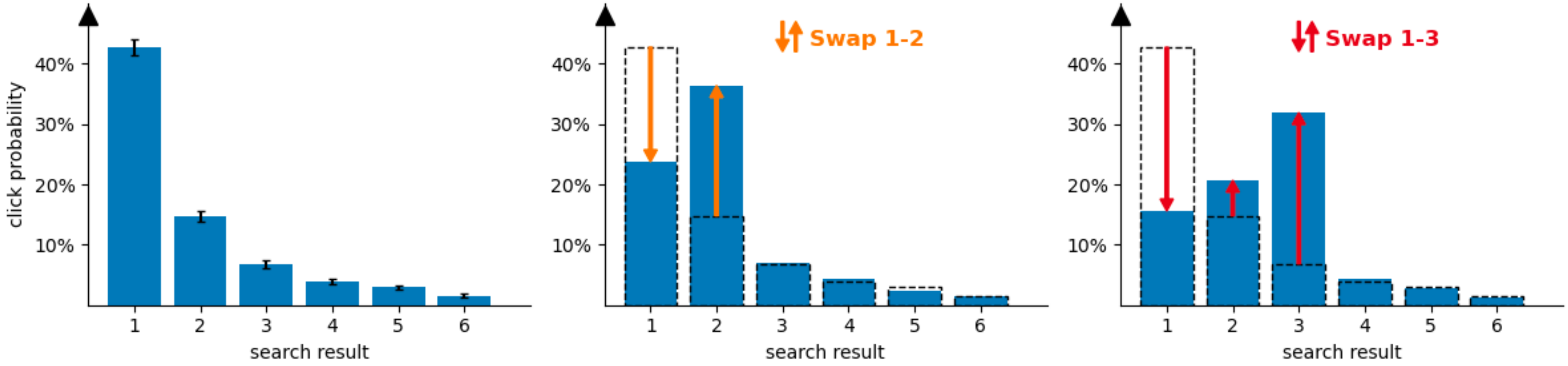
Browser extension

What is the causal effect of Google's ranking algorithm on user clicks?

- Experiment as gold standard: change the algorithm and inspect effect
- Algorithm is proprietary and complex
- Intervene at the level of display to emulate algorithmic updates

Performativity gap in online search

>70’000 search queries of 85 users collected over 3 months.
Randomized display for each query



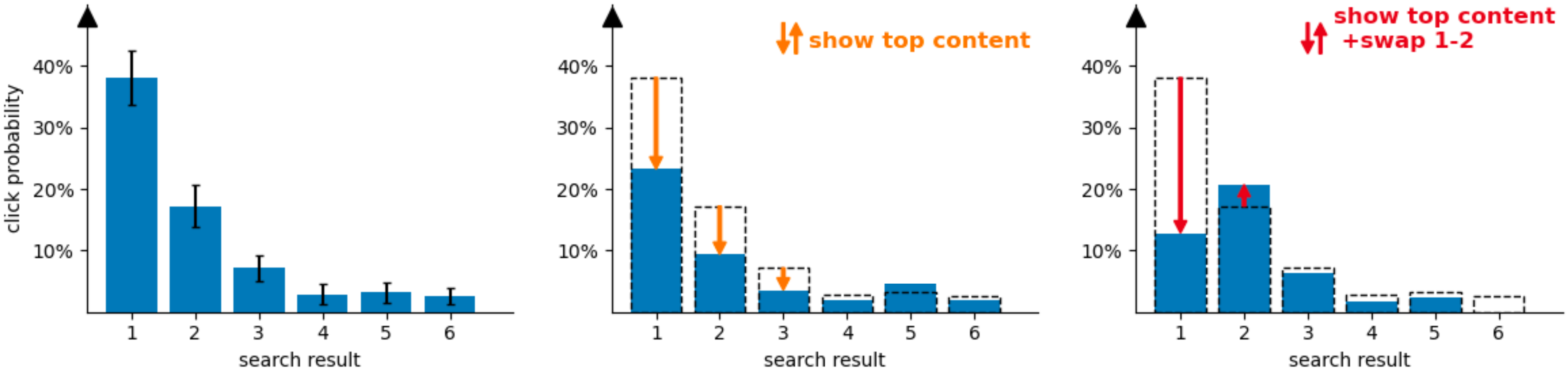
$\delta_1 = 0.44\ \mathrm{CTR}_1$ $\delta_1 = 0.63\ \mathrm{CTR}_1$

Power of Google search over the incoming traffic to a website ranked in first position corresponds to more than 44% of base traffic.

[MCH24]

Performativity gap in online search

Focus on queries with boxes naturally present.



$\delta_1 = 0.44 \; \mathrm{CTR}_1$

$\delta_1 = 0.66 \; \mathrm{CTR}_1$

The effect of adding top content and downranking an element is larger than the effect of any of the two individual conducts

As its name suggests…

it is a search *engine* not a camera

Applications

- In antitrust investigations we care about performative power over a population of consumers in a specific relevant market.

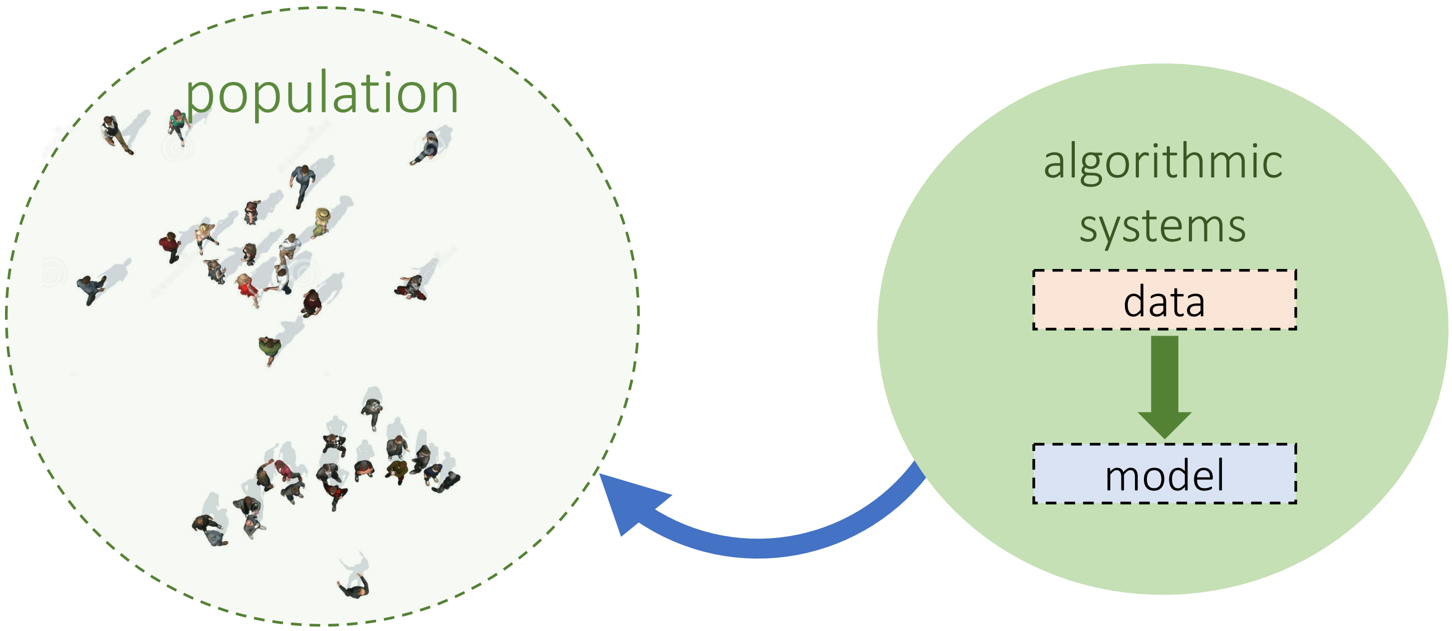
  Google Shopping case these are consumers in CSS market.
  A large fraction of them use Google search for navigating to the services.

- When mandating remedies we can use measures of performative power to monitor effectiveness over population of interest.

- In consumer protection and fairness we care about power over subpopulations.

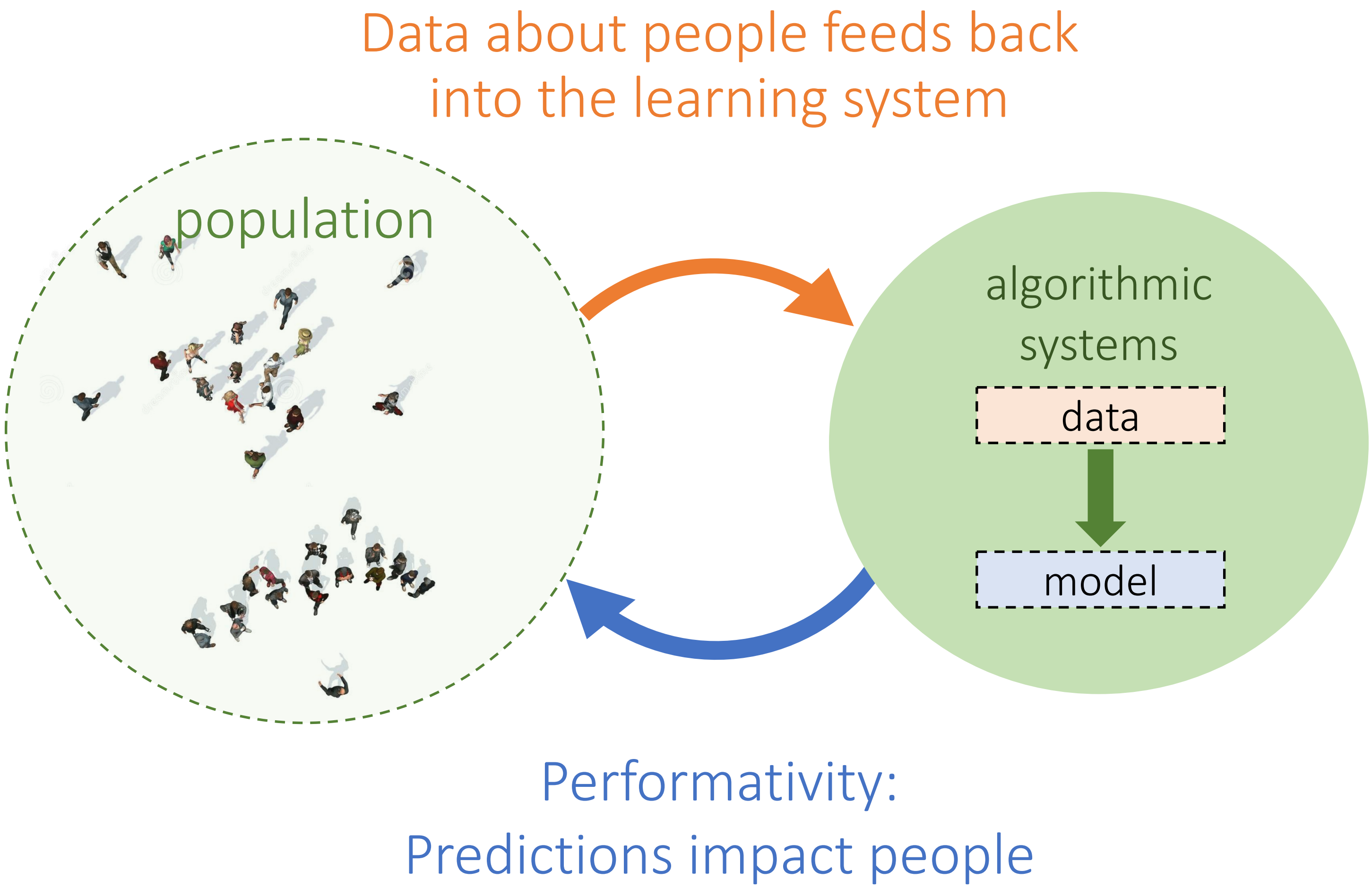
Power surfaces in performativity
Performative power can be instantiated flexibly

Let’s zoom out



Performativity:
Predictions impact people

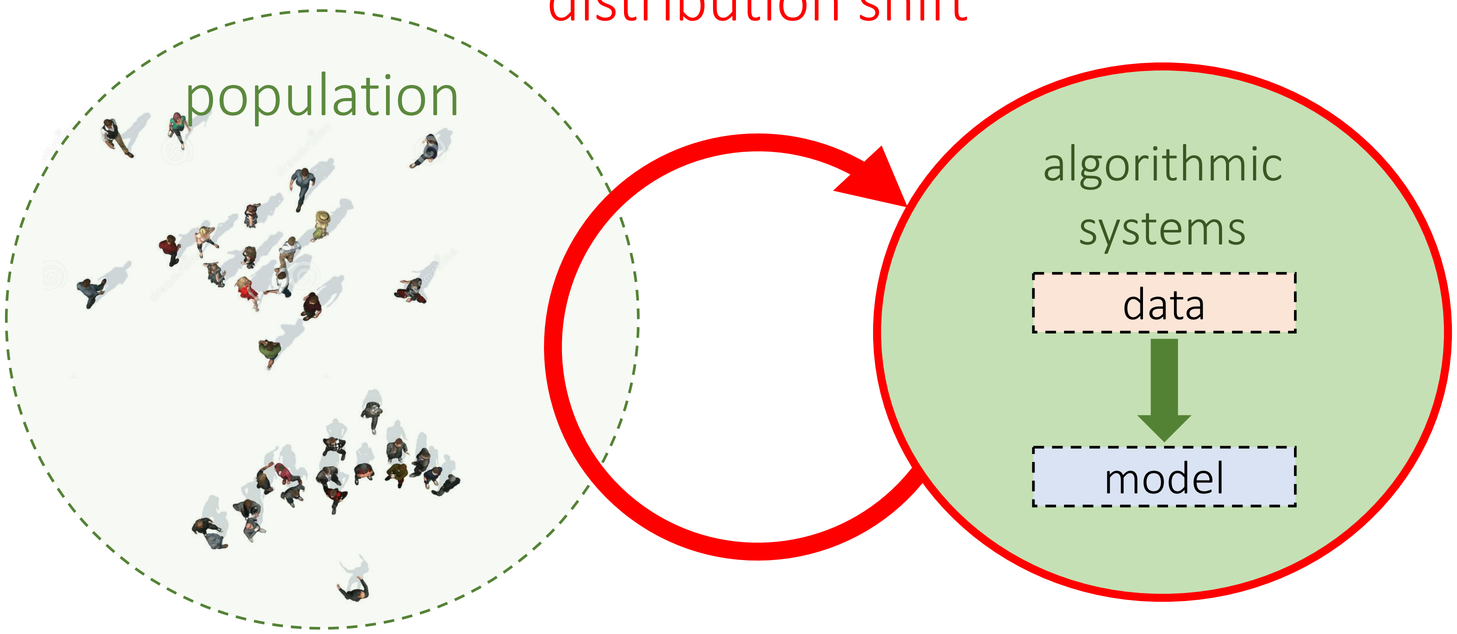
Let’s zoom out



Let’s zoom out

Finding performative optima:

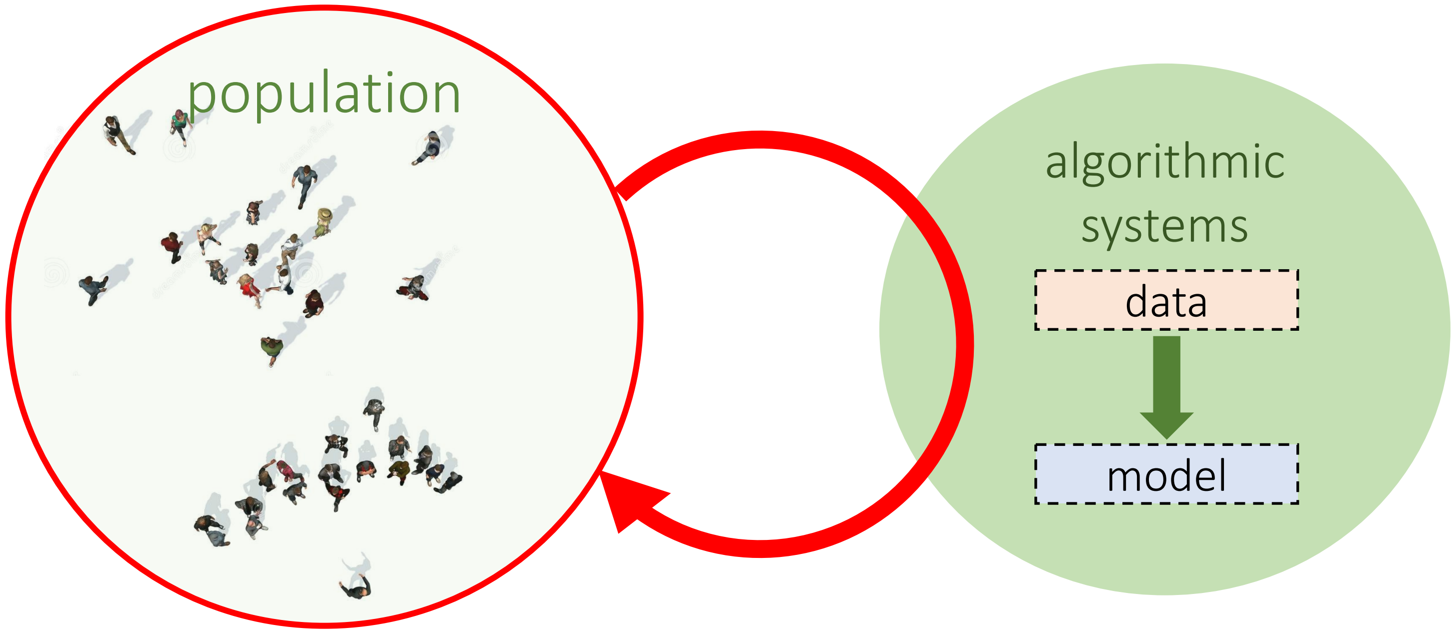
Firm anticipates
distribution shift



Let’s zoom out

Reversing the order of play:

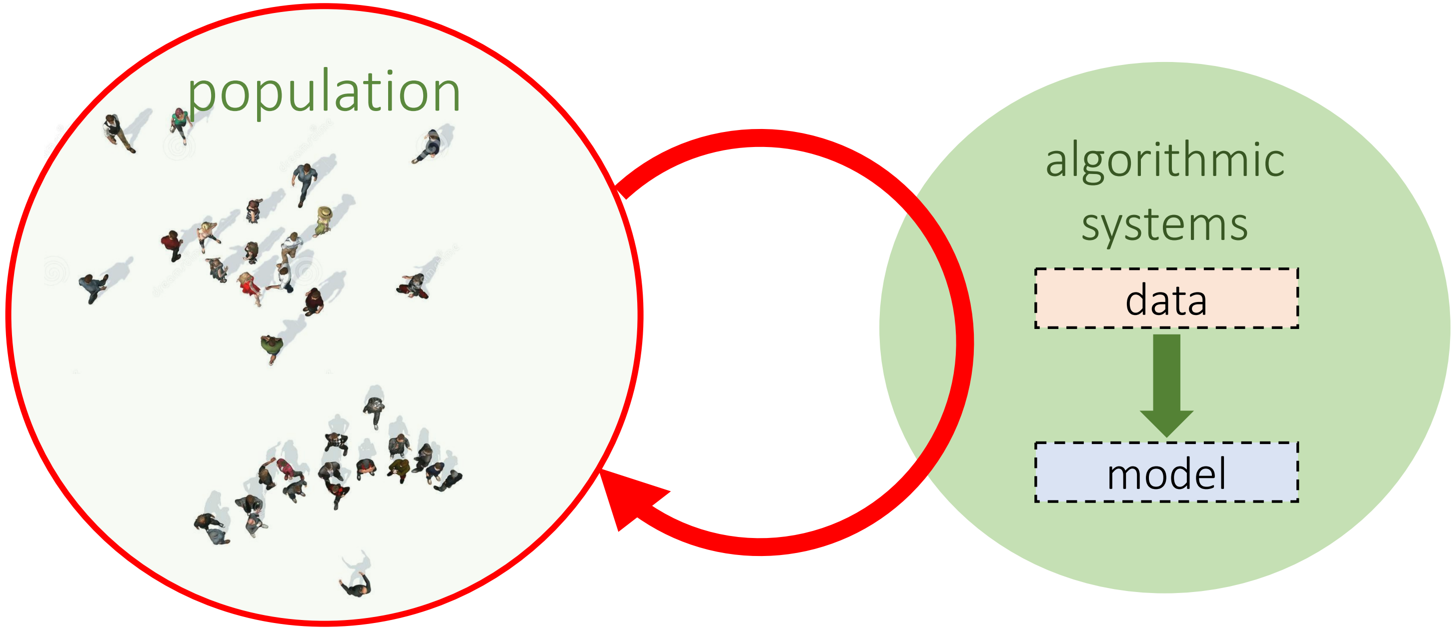
Individuals anticipate the use of their data for learning



Let's zoom out

Reversing the order of play:

Individuals anticipate the use of their data for learning



Performativity is the reason they care

Algorithmic resistance

Uber & Lyft Drivers Reportedly Rigging App to Create Surge Pricing

"And we all know, rule number one, we don't talk about 'Surge Club.'"

INDEPENDENT

NEWS VOICES SPORT CULTURE INDY/LIFE INDYBEST VIDEO DAILY EDITION

News > Business > Business News

Uber drivers work together to create price surge and charge customers more, researchers find

Some drivers are deli
when they log back i

Ben Chapman | @b_c_chap

DIGITAL TRENDS

Trending: Vape Ban Disney+ Review Early Black Friday Deals

Uber drivers reportedly triggering

Restricted access | Research article | First published online September 6, 2017

Thrown under the bus and outrunning it! The logic of Didi and taxi drivers' labour and activism in the on-demand economy

Julie Yujie Chen Vie

Volume 20, Issue 8

TECH • ARTIFICIAL INTELLIGENCE

Gig Workers Behind AI Face 'Unfair Working Conditions,' Oxford Report Finds

6 MINUTE READ

Algorithmic resistance

Uber & Lyft Drivers Reportedly Rigging App to Create Surge Pricing

"And we all know, rule number one, we don't talk about 'Surge Club.'"

INDEPENDENT

NEWS VOICES SPORT CULTURE INDY/LIFE INDYBEST VIDEO DAILY EDITION

News > Business > Business News

Uber drivers work together to create price surge and charge customers more, researchers find

Some drivers are deli… when they log back i…

Ben Chapman | @b_c_chap…

DIGITAL TRENDS

Trending: Vape Ban Disney+ Review Early Black Friday Deals

Uber drivers reportedly triggering higher fares through Surge Club

Restricted access | Research article | First published online September 6, 2017

Thrown under the bus and outrunning it! The logic of Didi and taxi drivers' labour and activism in the on-demand economy

Julie Yujie Chen Vie…

Volume 20, Issue 8 |

TECH • ARTIFICIAL INTELLIGENCE

Gig Workers Behind AI Face 'Unfair Working Conditions,' Oxford Report Finds

6 MINUTE READ

THE GUARDIAN UNIVERSITY OF CALIFORNIA, SAN DIEGO

Unbalanced Power Dynamics in the Music Industry

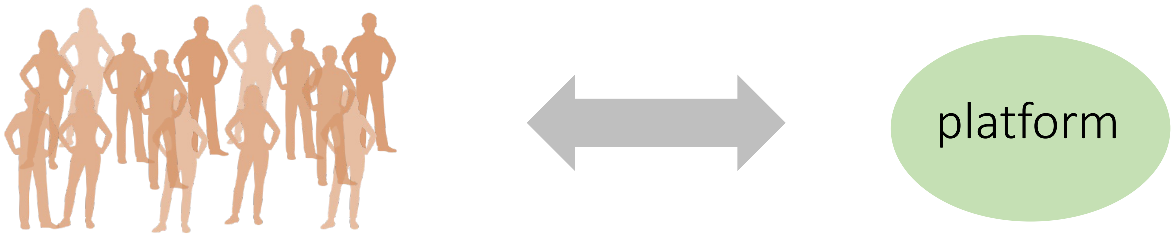
NEWS

Union of Musicians and Allied Workers Launches "Justice at Spotify" Campaign

JUSTICE AT SPOTIFY

Coordinated efforts

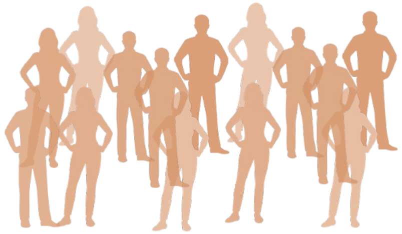
Coordination is effective on the side of the users to have influence on the learning algorithm.



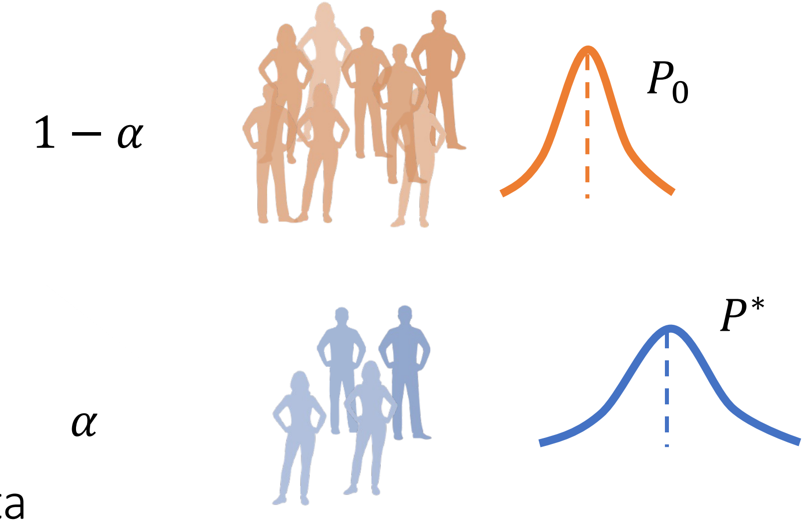
Data leverage:
- data strikes, concious data contribution [VH21, VHS19, VLTCH21]
- algorithmic collective action [HMMZ23]
- collective infrastructure, e.g., GigSense [IFS24]

A single datapoint has little influence, but systematic patterns will be picked up on

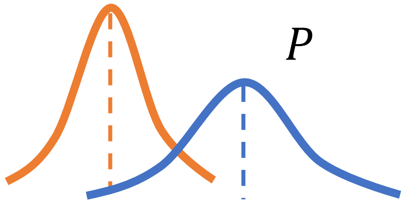
Model of algorithmic collective action



Individuals' initial data
$(x, y) \sim P_0$



α-fraction of the population joins the collective and implements a strategy to change data



Platform observes mixture distribution
$P = (1 - \alpha)P_0 + \alpha P^*$

platform

Platform trains ML model f on P

Collective goal:
Favorable property of f

Anticipate retraining

<u>Theorem</u> [HMMZ23]: For controlling the output of a gradient-based learner it is sufficient to have a collective of fraction α repeatedly modifying their data as long as

$$\alpha \geq O\left(\mathrm{E}_{z \sim P_0} ||\nabla \ell(\theta^*; z)|| \right)$$

$\alpha \propto$ suboptimality of the targeted solution θ^*

Anticipate retraining

<u>Theorem</u> [HMMZ23]: For controlling the output of a gradient-based learner it is sufficient to have a collective of fraction α repeatedly modifying their data as long as

$$\alpha \geq O\left(\mathrm{E}_{z\sim P_0} ||\nabla \ell(\theta^*; z)||\right)$$

$\alpha \propto$ suboptimality of the targeted solution θ^*

"provoke target classification at test time"

$$f\big(g(x)\big) = y^*$$

<u>Theorem</u> [HMMZ23]: For planting a signal against an ϵ-optimal risk minimizing learner with success p^* it is sufficient to have a collective of fraction α with

$$\alpha \geq \frac{\xi}{1 - p^* + \xi}$$

where ξ = $P_0\{g(x): x \in X\}$

$\alpha \propto$ uniqueness of the signal to be planted

Small collectives can be effective

By strategically correlating a single character in the CV with a skill at training time, gig workers can plant a trigger to be exploited at test time.

- Against a Bert classifier a collective size of 0.1% is sufficient [HMMZ23].

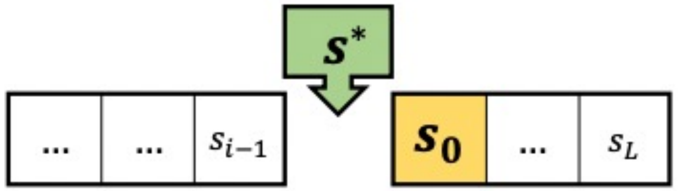
  The more accurate the learner, the more effective the strategy

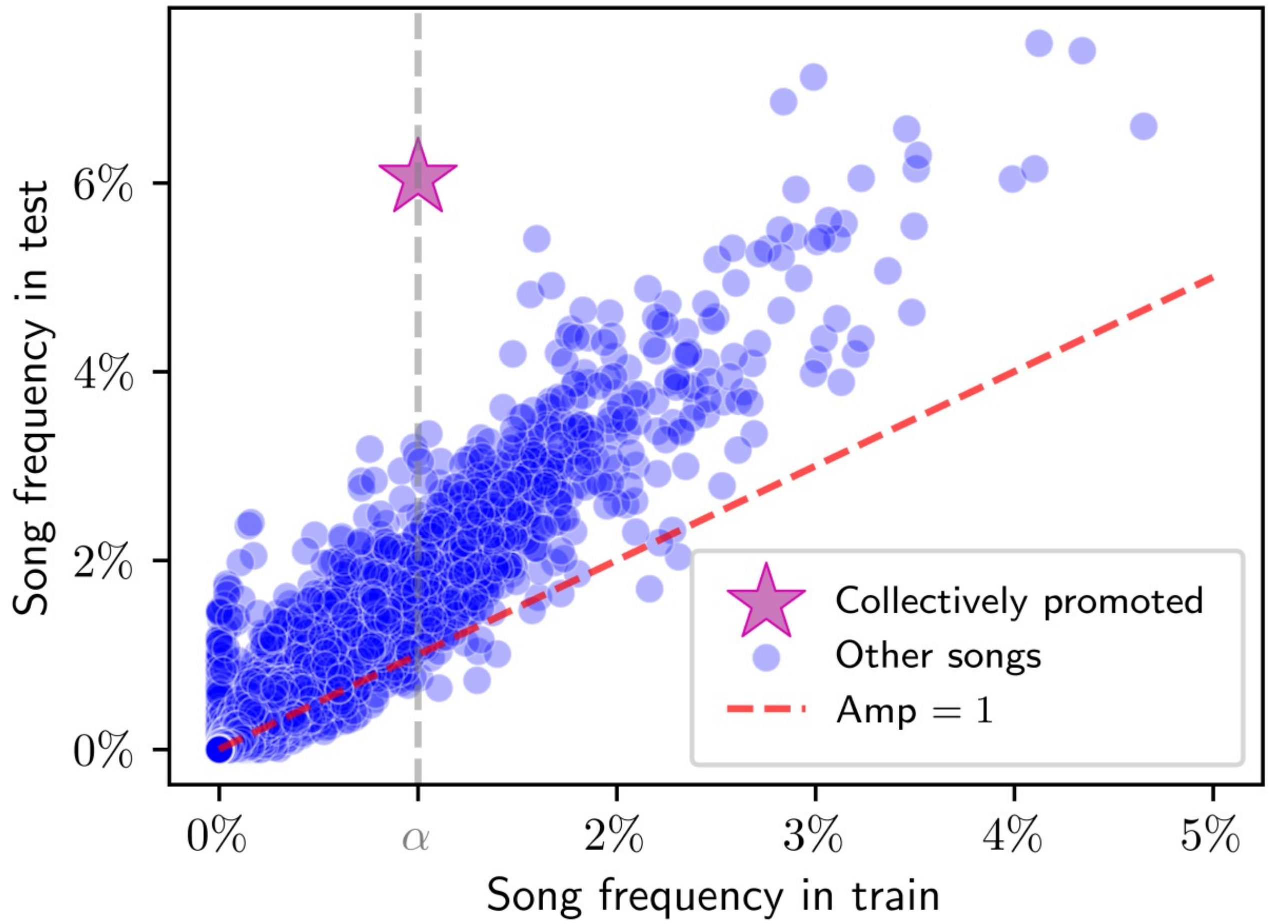
Various data poisoning example demonstrate the feasibility of impacting learner with few datapoints, see [TCLY22] for an overview.

[BM24]

Small changes can be sufficient

By reordering playlists and strategically choosing the position of a target song, fans can have disproportionate impact on transformer-based recommendations at test time.





utility preserving actions can be effective

Incentives to participate

Utility of firm and participants often not aligned

Collective action gives power to participants

Incentives for participation:

- collective action comes with overheads and constraints
- typically not self-incentivized (see Olson 1956)
- firms might want to protect against it, punish participation, or move away from statistical learning

[HM23]

Incentives to participate

Utility of firm and participants often not aligned

Collective action gives power to participants

Incentives for participation:

- collective action comes with overheads and constraints
- typically not self-incentivized (see Olson 1956)
- firms might want to protect against it, punish participation, or move away from statistical learning

The performativity of predictions determines payoff of strategies and how much cost individuals are willing to incur

Discussion

Discussion

- When deployed in the real world AI predictions are part of a broader sociotechnical ecosystem
- Performativity is pervasive
- Changing predictions means changing outcomes
- Prediction is no longer a purely technical endeavor
- Solution concepts are context dependent

Open problems and challenges

- Major challenge for practical developments: data availability
- Performative optimization requires exploring models; how do we do so safely? Should we aim for a different solution concept?
- How should performativity in machine learning be regulated? What kinds of performative effects are acceptable?
- We saw that predictions impact people and people impact predictions; how do we model this jointly?

Thank you!

Questions?

References

[PZMH20] Perdomo, J., Zrnic, T., Mendler-Dünner, C., & Hardt, M. Performative prediction. *International Conference on Machine Learning (ICML)*, 2020

[MPZH20] Mendler-Dünner, C., Perdomo, J., Zrnic, T., & Hardt, M. Stochastic optimization for performative prediction. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020

[DX23] Drusvyatskiy, D., & Xiao, L. Stochastic optimization with decision-dependent distributions. *Mathematics of Operations Research*, 2023

[WBD21] Wood, K., Bianchin, G., & Dall'Anese, E. Online projected gradient descent for stochastic optimization with decision-dependent distributions. *IEEE Control Systems Letters*, 2021

[MMG23] Mofakhami, M., Mitliagkas, I., & Gidel, G. Performative prediction with neural networks. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2023

[LW24] Li, Q., & Wai, H. T. Stochastic optimization schemes for performative prediction with nonconvex loss. *ArXiv preprint arXiv:2405.17922, 2024*

[LZ24] Lin, L., & Zrnic, T. Plug-in performative optimization. *International Conference on Machine Learning (ICML)*, 2024

[MPZ21] Miller, J. P., Perdomo, J. C., & Zrnic, T. Outside the echo chamber: Optimizing the performative risk. *International Conference on Machine Learning (ICML)*, 2021

[KSU08] Kleinberg, R., Slivkins, A., & Upfal, E. Multi-armed bandits in metric spaces. *ACM Symposium on Theory of Computing (STOC)*, 2008

[EMM06] Even-Dar, E., Mannor, S., Mansour, Y., & Mahadevan, S. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research (JMLR),* 2006

[GKRSW22] Gopalan, P., Kalai, A. T., Reingold, O., Sharan, V., & Wieder, U. Omnipredictors. *Innovations in Theoretical Computer Science (ITCS), 2022*

[KP23] Kim, M. P., & Perdomo, J. C. Making decisions under outcome performativity. *Innovations in Theoretical Computer Science (ITCS)*, 2023

[HMPW16] Hardt, M., Megiddo, N., Papadimitriou, C., & Wootters, M. Strategic classification. *Innovations in Theoretical Computer Science (ITCS)*, 2016

[MDW22] Mendler-Dünner, C., Ding, F., & Wang, Y. Anticipating performativity by predicting from predictions. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022

[BHK22] Brown, G., Hod, S., & Kalemaj, I. Performative prediction in a stateful world. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022

[RRDF22] Ray, M., Ratliff, L. J., Drusvyatskiy, D., & Fazel, M. Decision-dependent risk minimization in geometrically decaying dynamic environments. *AAAI Conference on Artificial Intelligence (AAAI)*, 2022

[LW22] Li, Q., & Wai, H. T. State dependent performative prediction with stochastic approximation. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.

[IZY22] Izzo, Z., Zou, J., & Ying, L. How to learn when data gradually reacts to your model. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022

[NFDFR23] Narang, A., Faulkner, E., Drusvyatskiy, D., Fazel, M., & Ratliff, L. J. Multiplayer performative prediction: Learning in decision-dependent games. *Journal of Machine Learning Research (JMLR)*, 2023

[PY23] Piliouras, G., & Yu, F. Y. Multi-agent performative prediction: From global stability and optimality to chaos. *ACM Conference on Economics and Computation (EC)*, 2023

[WYW23] Wang, X., Yau, C. Y., & Wai, H. T. Network effects in performative prediction games. *International Conference on Machine Learning (ICML)*, 2023

[MTR23] Mandal, D., Triantafyllou, S., & Radanovic, G. Performative reinforcement learning. *International Conference on Machine Learning (ICML)*, 2023

[CHM24] Cheng, G., Hardt, M., Mendler-Dünner, C. Causal Inference out of Control: Identifying the Steerability of Consumption. *International Conference on Machine Learning (ICML)*, 2024

[RTMR22] Rank, B., Triantafyllou, S., Mandal, D., & Radanovic, G. Performative reinforcement learning in gradually shifting environments. *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2024

[LDRSH18] Liu, L. T., Dean, S., Rolf, E., Simchowitz, M., & Hardt, M. Delayed impact of fair machine learning. *International Conference on Machine Learning (ICML)*, 2018

[HSNL18] Hashimoto, T., Srivastava, M., Namkoong, H., & Liang, P. Fairness without demographics in repeated loss minimization. *International Conference on Machine Learning (ICML)*, 2018

[JXLZ24] Jin, K., Xie, T., Liu, Y., & Zhang, X. Addressing polarization and unfairness in performative prediction. *ArXiv preprint arXiv:2406.16756*, 2024

[HJM22] Hardt, M., Jagadeesan M., Mendler-Dünner C. Performative power. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[MCH24] Mendler-Dünner, C., Caravano, G., Hardt, M. An engine not a camera: measuring performative power of online search. *ArXiv preprint arXiv:2405.19073*, 2024

[VHS19] Vincent, N., Hecht, B., &Sen S. “Data Strikes”: Evaluating the Effectiveness of a New Form of Collective Action Against Technology Companies. *The World Wide Web Conference (WWW).* 2019

[VH21] Vincent, N., Hecht, B. Can “Conscious Data Contribution” Help Users to Exert “Data Leverage” Against Technology Companies? *ACM Human-Computer Interaction*. 2021

[VLTCH21] Vincent, N., Li, H., Tilly, N., Chancellor, S., Hecht, B. Data Leverage: A Framework for Empowering the Public in its Relationship with Technology Companies. *ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. 2021

[HMMZ23] Hardt, M., Mazumdar, E., Mendler-Dünner, C., Zrnic, T. Algorithmic collective action in machine learning. *International Conference on Machine Learning (ICML)*, 2023

[IFS24] Imteyaz, K., Flores-Saviaga, C., Savage, S. GigSense: An LLM-Infused Tool for Workers’ Collective Intelligence. *Arxiv preprint arXiv:2405.02528*, 2024

[TCLY22] Tian, Z., Cui, L., Liang, J., Yu, S.. A comprehensive survey on poisoning attacks and countermeasures in machine learning. *ACM Computing Surveys*, 55(8), 2022.

[BM24] Baumann, J., Mendler-Dünner, C. Algorithmic collective action in recommender systems.
European Workshop on Algorithmic Fairness (EWAF), 2024

[HM23] Hardt, M., & Mendler-Dünner, C. Performative prediction: past and future.
ArXiv preprint arXiv:2310.16608, 2023