Multi-agent Performative Prediction

From Global Stability and Optimality to Chaos

Georgios Piliouras (Google DeepMind, SUTD)

Fang-Yi Yu, (George Mason University)

Pattern recognition/prediction

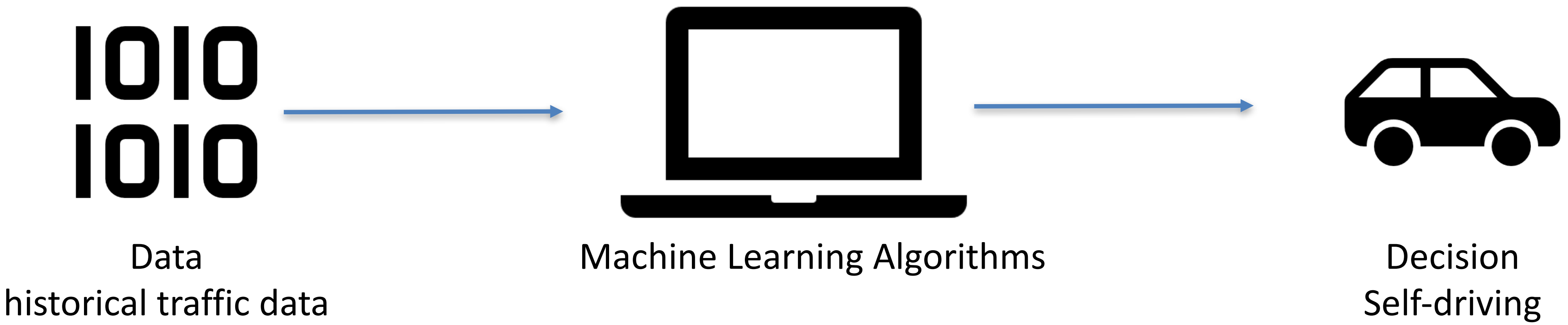
Data
historical traffic data

Machine Learning Algorithms

Predictive models
Estimated arrival time

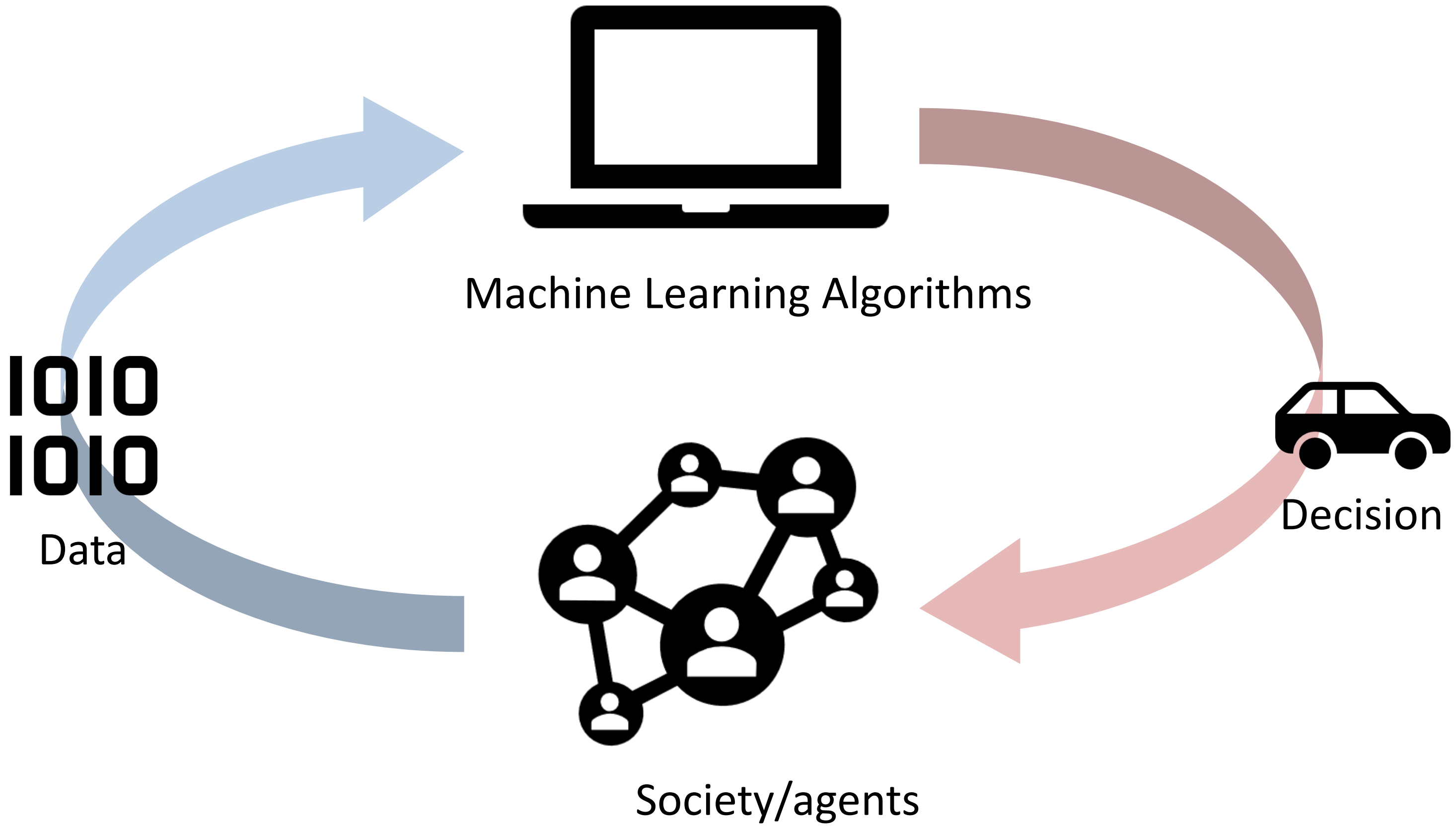
- E.g., Natural language processing, computer vision, traffic prediction
- Learning a model in isolation

Decision-making



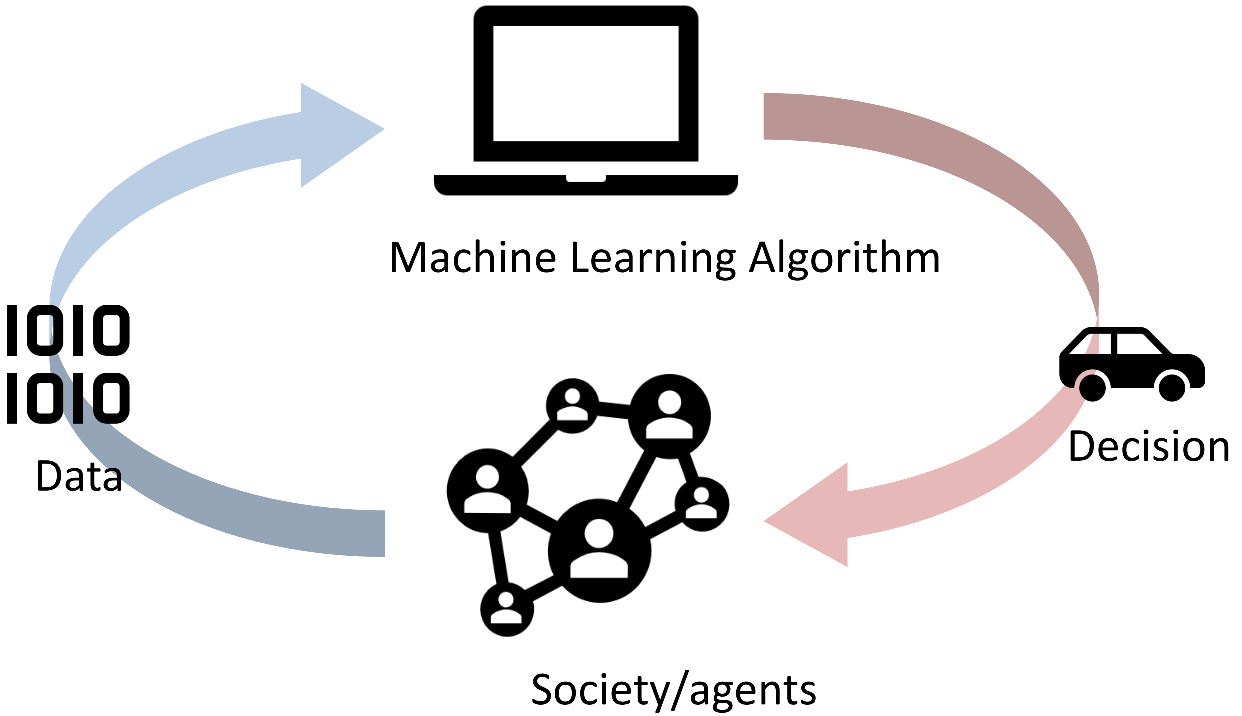
- E.g., Recommendation system, self-driving car, medical diagnosis
- Huge interconnected web of data, agents, decisions

Machine learning and society



Prediction is part of a broader system

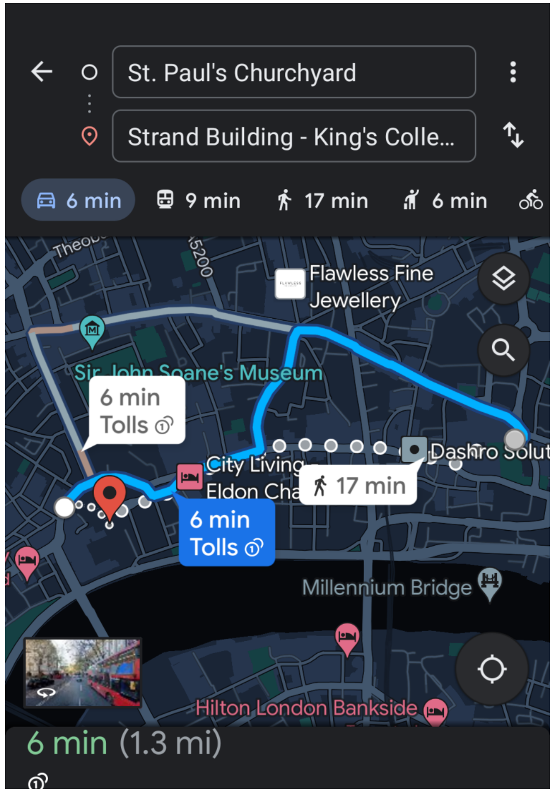
When predictions support decisions, we fundamentally change the distribution of future data. [JZMH20]



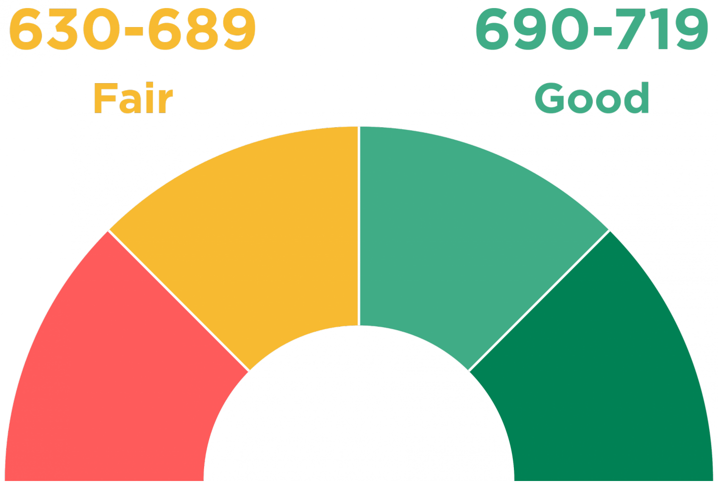
Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. 2020. Performative prediction

Performative predictions are everywhere

Traffic prediction



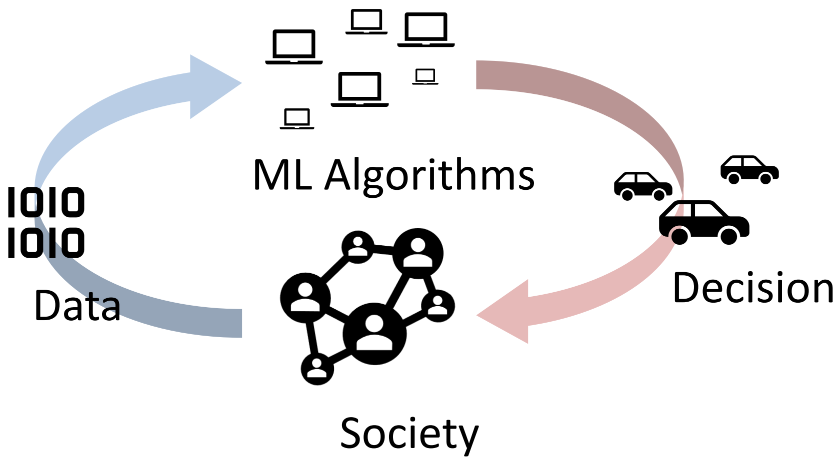
Credit score



Policy patrol with predictions

Contributions

- Formalize the concept of multi agent performative predictions and a novel solution concept, multi agent performative stability
- A threshold result on a common learning algorithm
  - Convergence for small learning rate
  - Li-Yorke chaos when too much influence



Outline

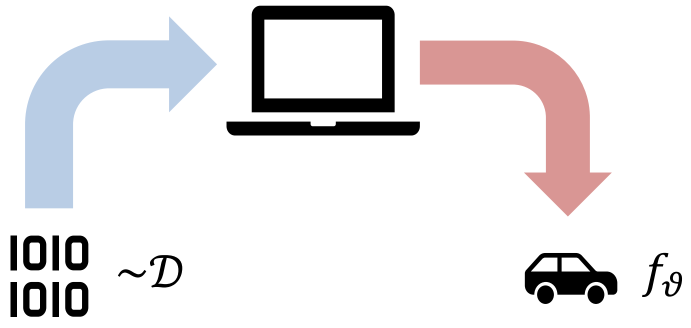
- Data from strategic agents
- Multi agent performative predictions
  - Model
  - Contributions
  - Simulation
- Future directions

Performative prediction

Supervised learning

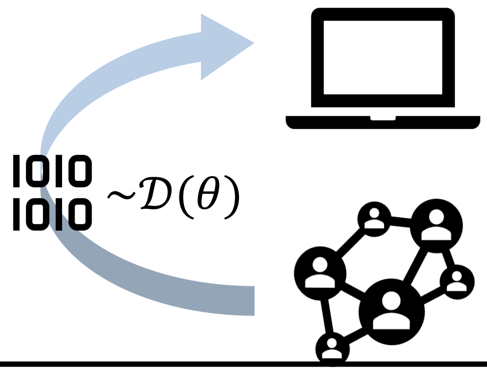
- Data $(x, y) \in \mathcal{X} \times \mathcal{Y}$ sampled from a fixed distribution $\mathcal{D}$
- Expected loss of a model f_ϑ with $\vartheta \in \Theta$

$$\ell(\vartheta; \mathcal{D}) \coloneqq \mathbb{E}_{(x,y)\sim\mathcal{D}}[\ell(f_\vartheta(x), y)]$$



Performative prediction

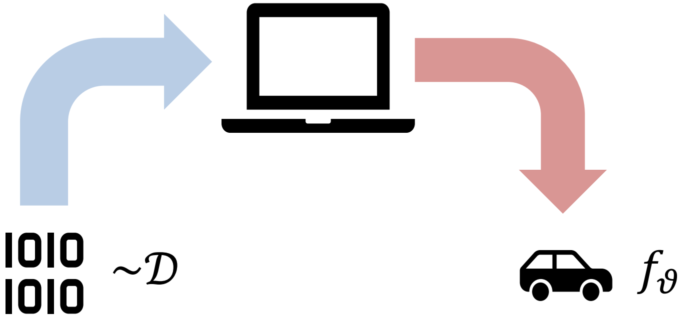
- Given $\theta \in \Theta$, data (x, y) is sampled from a distribution map $\mathcal{D}(\theta)$



Performative prediction

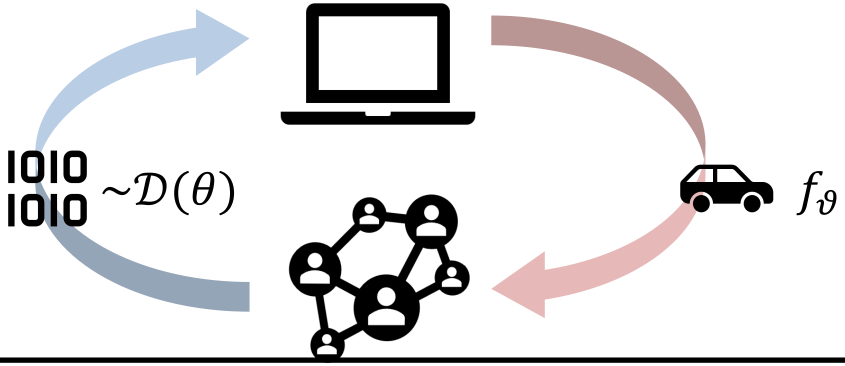
Supervised learning

- Data $(x, y) \in \mathcal{X} \times \mathcal{Y}$ sampled from a fixed distribution $\mathcal{D}$
- Expected loss of a model f_ϑ with $\vartheta \in \Theta$
  $\ell(\vartheta; \mathcal{D}) := \mathbb{E}_{(x,y)\sim\mathcal{D}}[\ell(f_\vartheta(x), y)]$



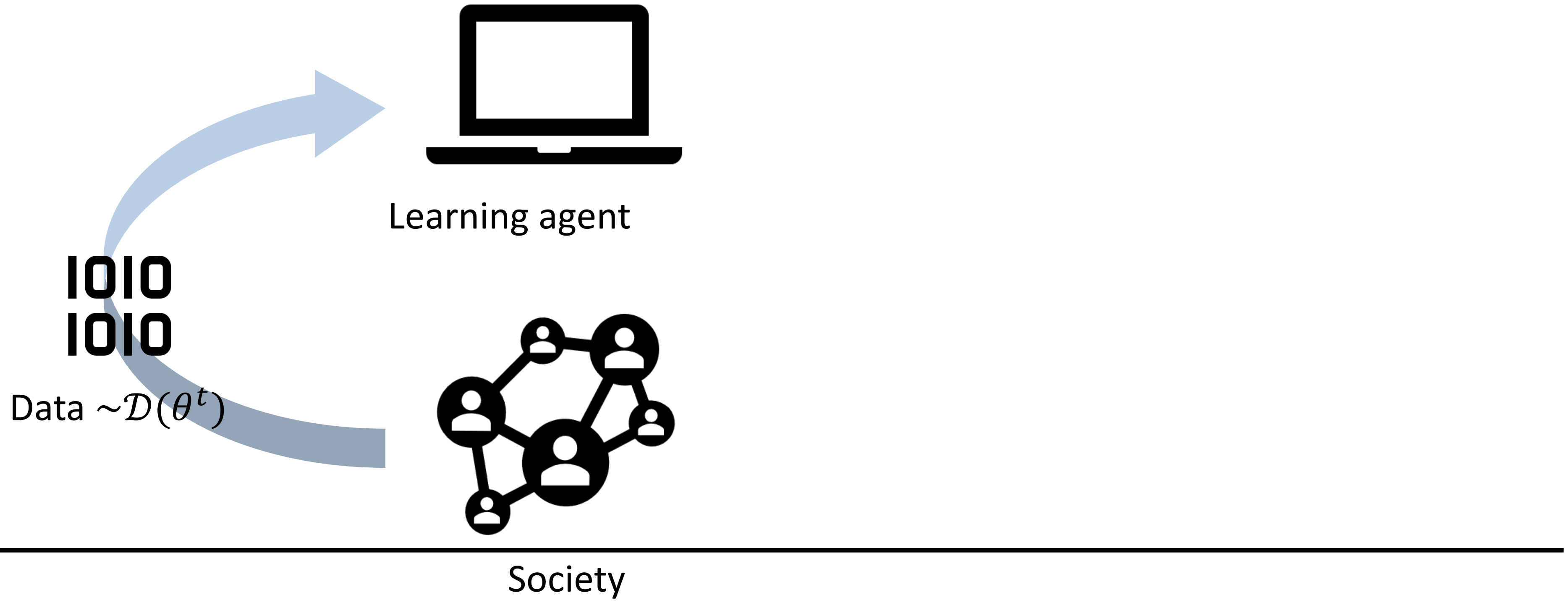
Performative prediction

- Given $\theta \in \Theta$, data (x, y) is sampled from a distribution map $\mathcal{D}(\theta)$
- Expected loss of a predictive model $\vartheta \in \Theta$ on deployed model $\theta \in \Theta$
  $\ell(\vartheta; \mathcal{D}(\theta)) := \mathbb{E}_{(x,y)\sim\mathcal{D}(\theta)}[\ell(f_\vartheta(x), y)]$



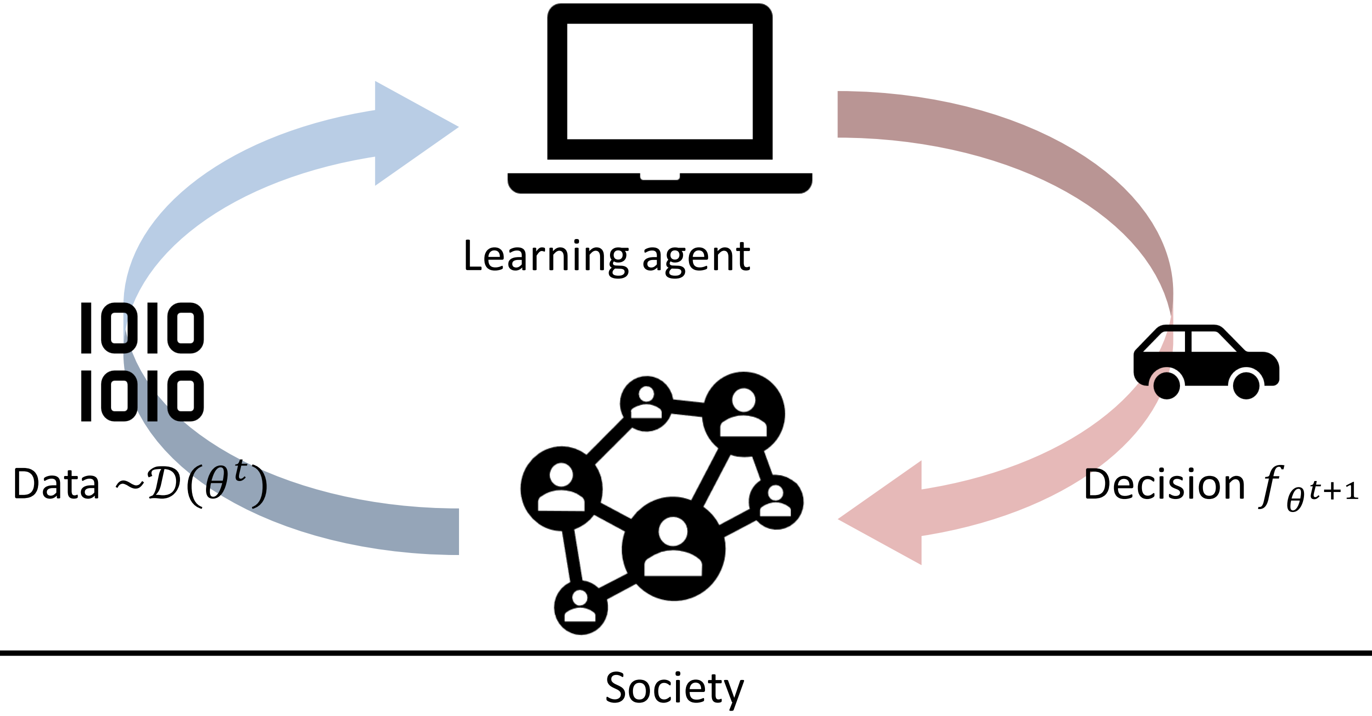
Dynamics of learning (retraining)

Since the distribution map $\mathcal{D}(\cdot)$ is unknown, the learning agent may iteratively access samples of $\mathcal{D}(\theta^t)$



Dynamics of learning

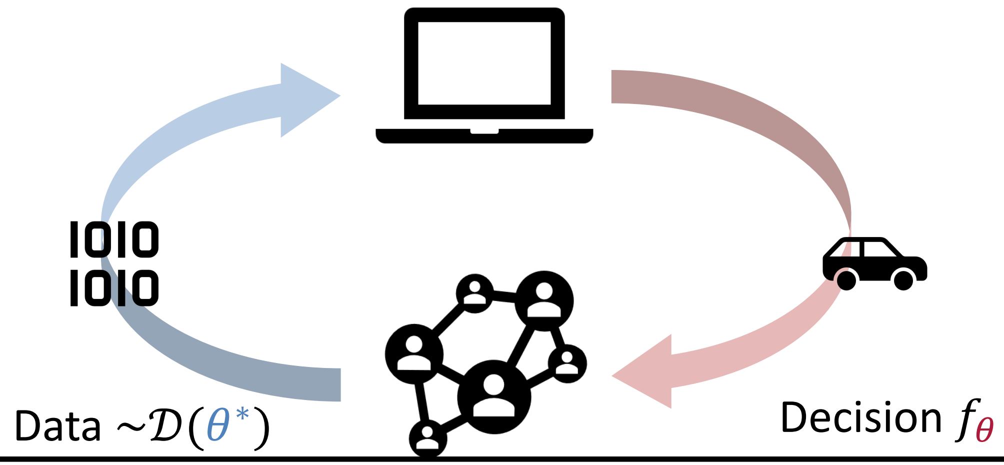
Since the distribution map $\mathcal{D}(\cdot)$ is unknown, the learning agent may iteratively access samples of $\mathcal{D}(\theta^t)$ and improve the model to θ^{t+1}



Dynamics of learning

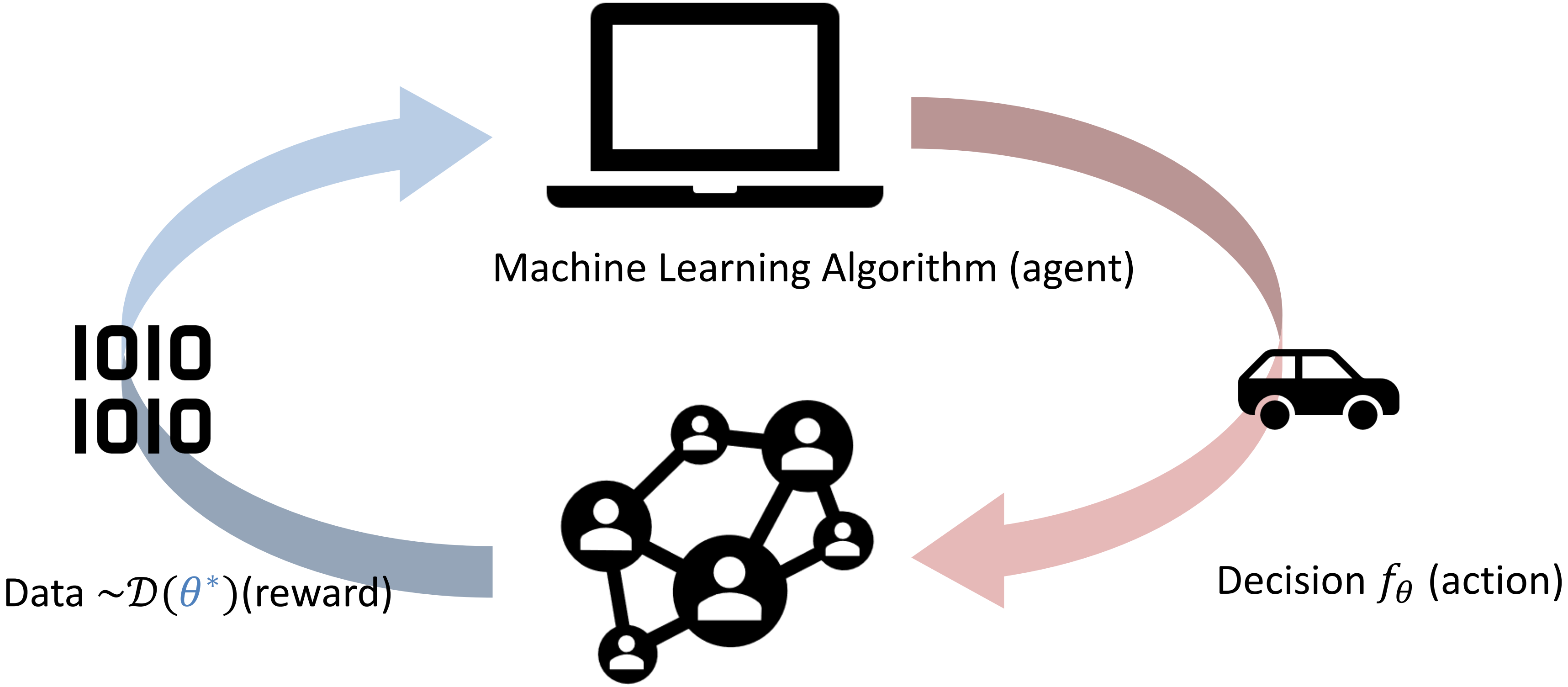
- The learning agent iteratively access samples of $\mathcal{D}(\theta^t)$ and improve the model to θ^{t+1}
- Fixed point: the model is optimal for distribution that it induces => performative stable point θ^*

$$\ell\big(\theta^*; \mathcal{D}(\theta^*)\big) = \min_{\theta} \ell\big(\theta; \mathcal{D}(\theta^*)\big)$$



Performative prediction as a game

- The predictive agent is playing a game against himself



Related work

- Convergence of learning dynamics
  - Naïve retraining with smoothness/convexity assumption
    - Repeated empirical risk minimization [JZMH20, DCRMF23]
    - Stochastic gradient descent [MPZH21, DX 20]
  - Learning with partial knowledge of $\mathcal{D}(\cdot)$
    - Stochastic gradient descent with known $\mathcal{D}(\cdot)$ [IYZ21]
    - Regret minimization with performative feedback [JZM22]
    - Plug-in estimator with a proxy $\mathcal{D}'(\cdot)$ [LZ23]

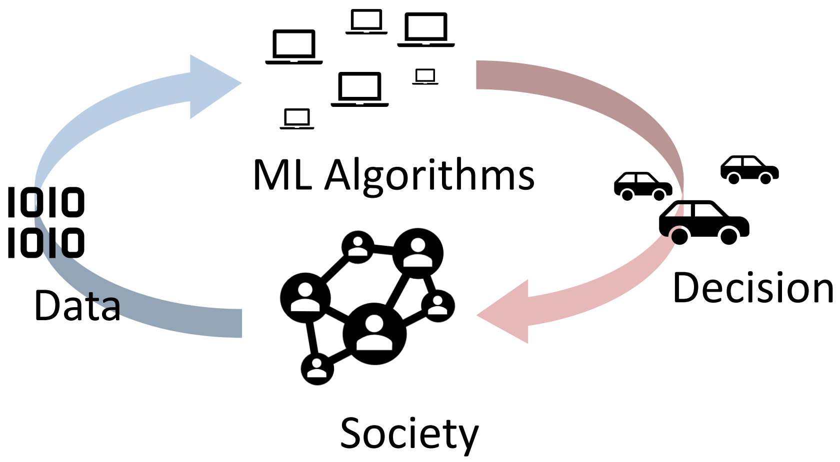
Related work

- Convergence of learning dynamics
  - **Naïve retraining with smoothness/convexity assumption**
    - Repeated empirical risk minimization [JZMH20, DCRMF23]
    - Stochastic gradient descent [MPZH21, DX 20]
  - Learning with partial knowledge of $\mathcal{D}(\cdot)$
    - Stochastic gradient descent with known $\mathcal{D}(\cdot)$ [IYZ21]
    - Regret minimization with performative feedback [JZM22]
    - Plug-in estimator with a proxy $\mathcal{D}'(\cdot)$ [LZ23]

When does learning fail?

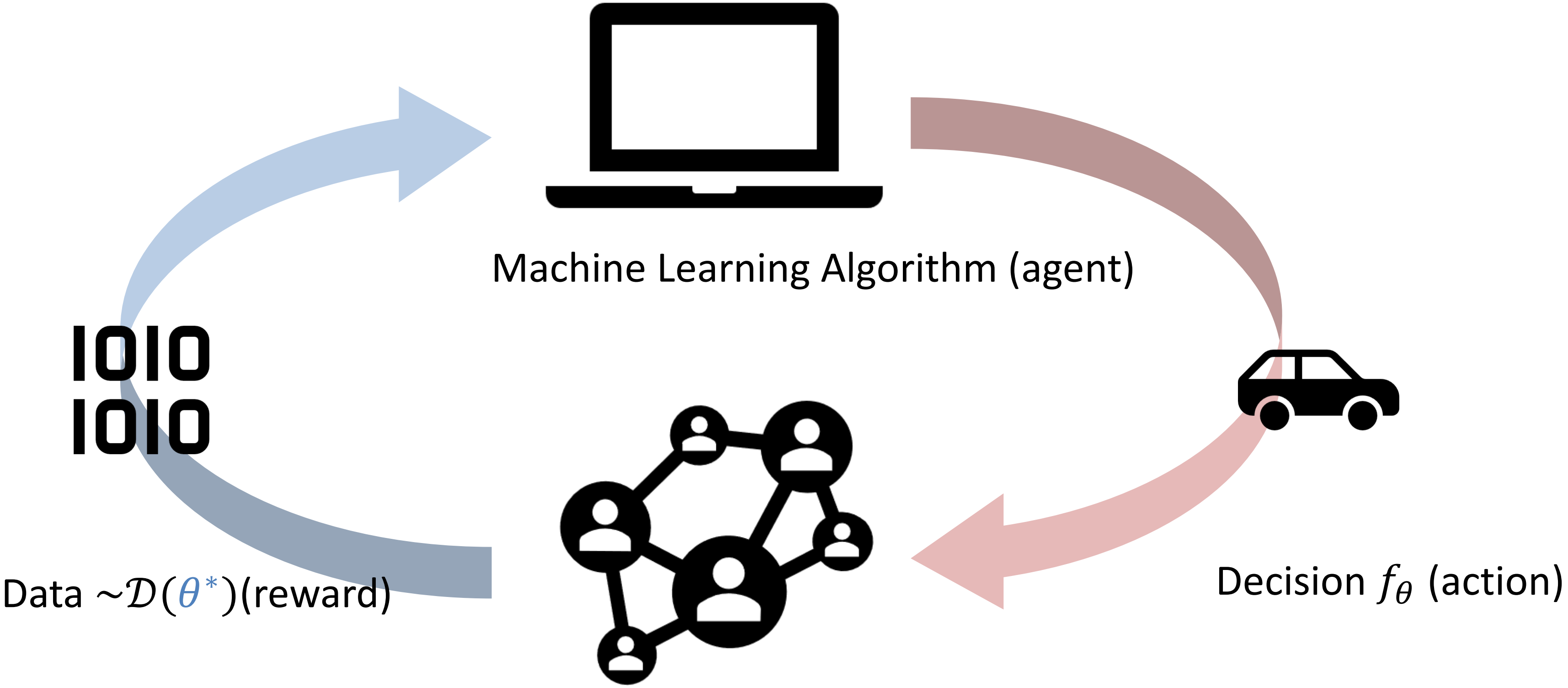
Contributions

- Formalize the concept of multi agent performative predictions and a novel solution concept, multi agent performative stability
- Main result: a threshold result on online learning algorithms
  - Convergence for small learning rate
  - Li-Yorke chaos when too much influence



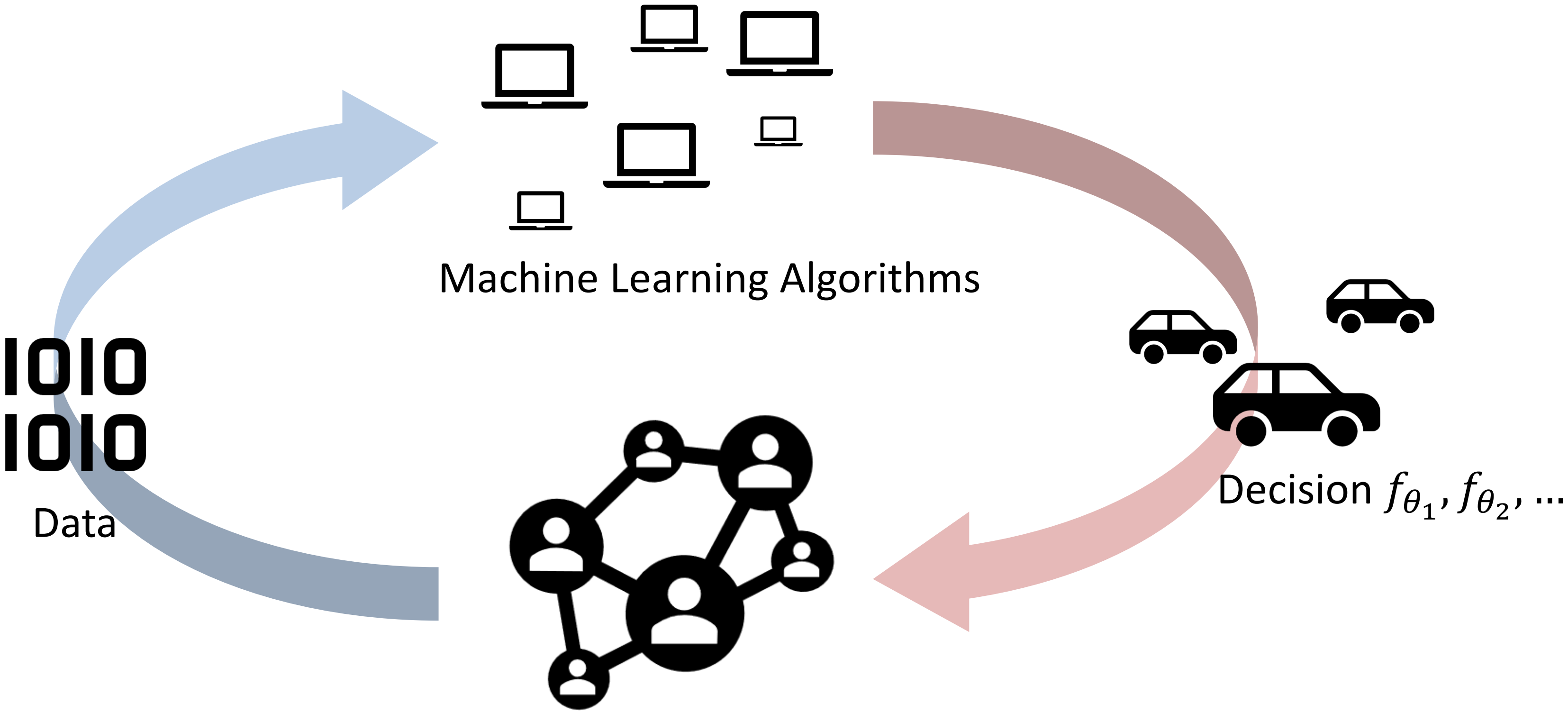
Performative prediction as a game

- The predictive agent is playing a game against himself



Multi-agent performative prediction

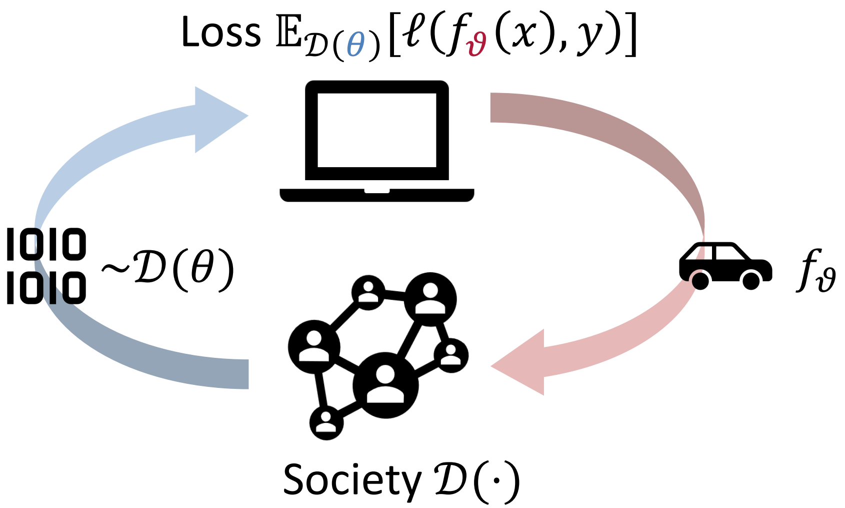
- Multiple predictive agents making decisions that collectively influence the distribution of future data.



Multi-agent performative prediction

Performative prediction

- Deploy $\theta \in \Theta$
- Distribution map $\mathcal{D}: \Theta \to \Delta_{\mathcal{X}\times\mathcal{Y}}$
- Predict $\vartheta \in \Theta$

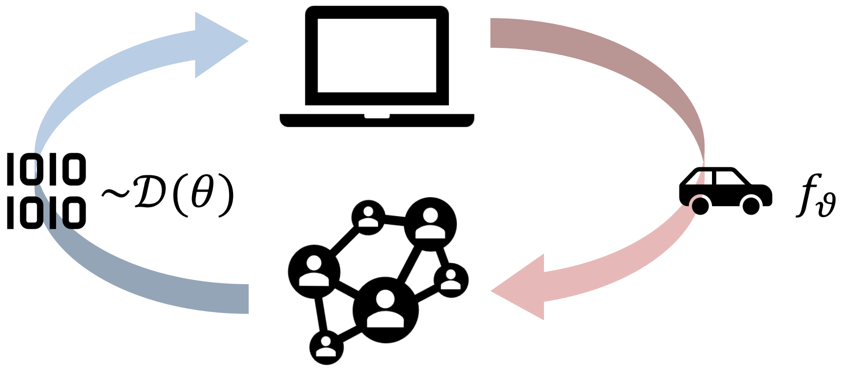


Multi-agent performative prediction

Performative prediction

- Deploy $\theta \in \Theta$
- Distribution map $\mathcal{D}: \Theta \to \Delta_{\mathcal{X}\times\mathcal{Y}}$
- Predict $\vartheta \in \Theta$

Loss $\mathbb{E}_{\mathcal{D}(\theta)}[\ell(f_\vartheta(x), y)]$

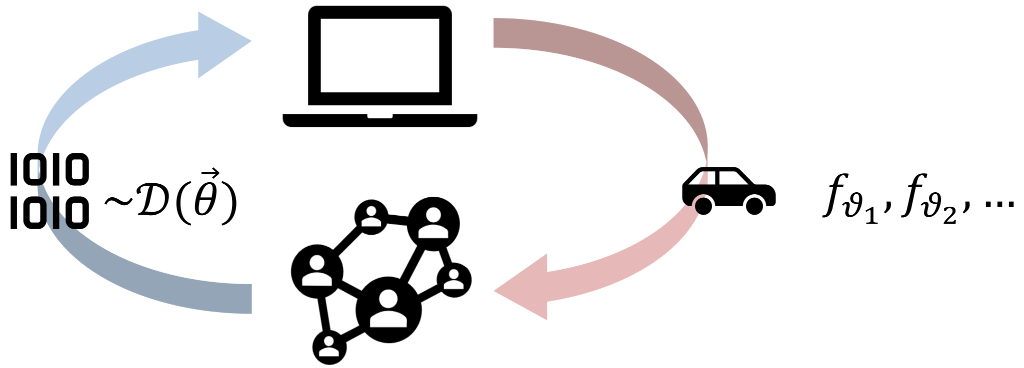


Society $\mathcal{D}(\cdot)$

Multi-agent performative prediction

- Deploy $\vec{\theta} = (\theta_1, \dots, \theta_n) \in \Theta^n$
- Distribution map $\mathcal{D}: \Theta^n \to \Delta_{\mathcal{X}\times\mathcal{Y}}$
- Predict $\vartheta \in \Theta$

Agent i's loss $\mathbb{E}_{\mathcal{D}(\vec{\theta})}[\ell_i(f_{\vartheta_i}(x), y)]$



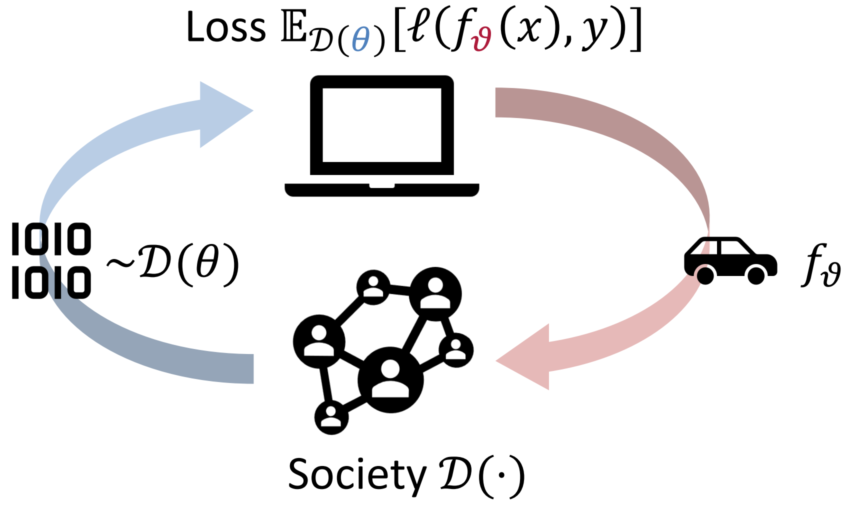
Society $\mathcal{D}(\cdot)$

Multi-agent performative prediction

Performative prediction

- Performative stable, $\theta^* \in \Theta$

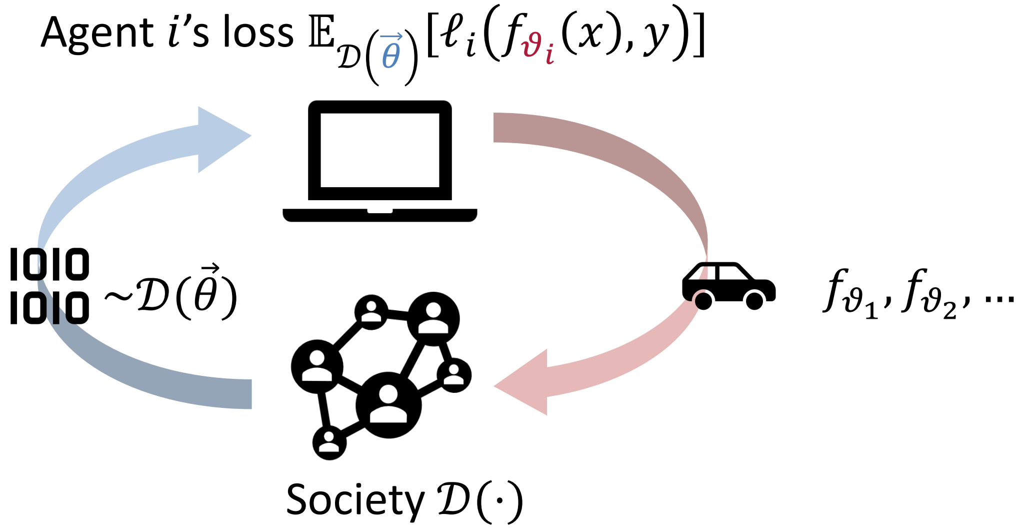
  $\ell(\theta^*; \mathcal{D}(\theta^*)) = \min_{\theta} \ell(\theta; \mathcal{D}(\theta^*))$



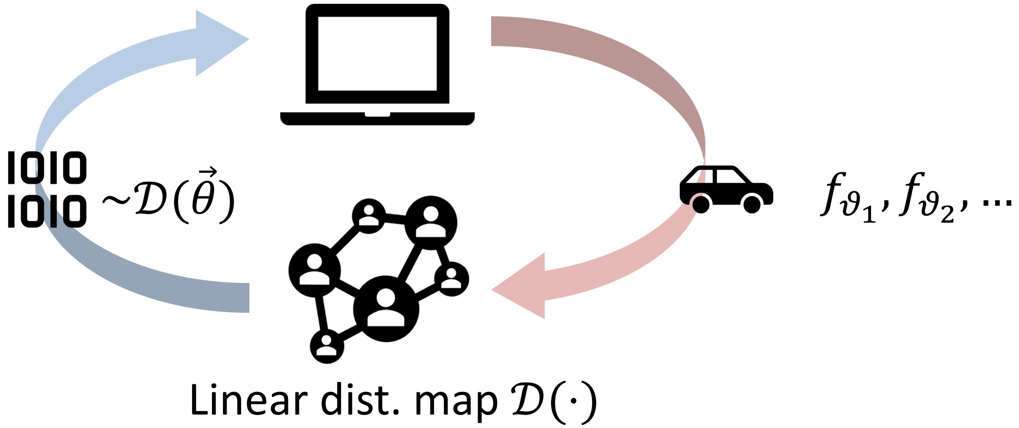
Multi-agent performative prediction

- Multi-agent performative stable, $\vec{\theta}^* \in \Theta^n$: for all $i \in [n]$,

  $\ell_i\left(\theta_i^*; \mathcal{D}(\vec{\theta}^*)\right) = \min_{\theta_i} \ell_i\left(\theta_i; \mathcal{D}(\vec{\theta}^*)\right)$



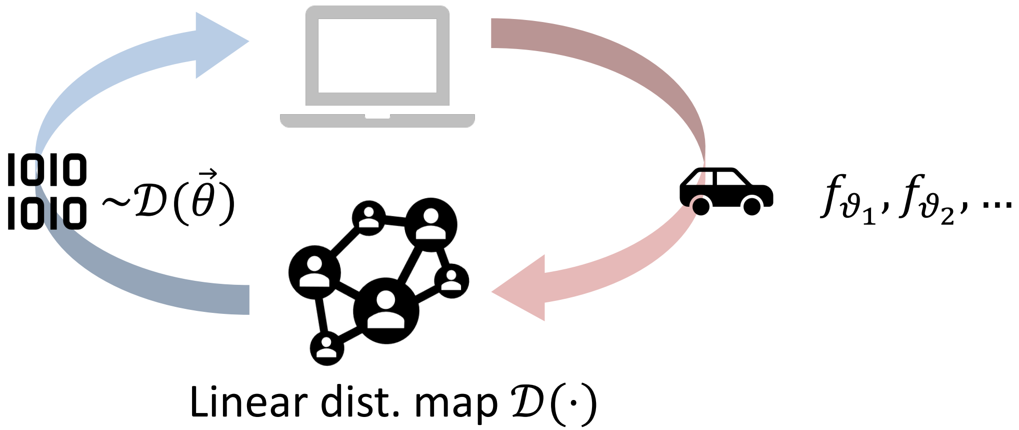
A toy learning dynamics $\vec{\theta}^1, \vec{\theta}^2, \dots, \vec{\theta}^t, \dots$



A toy learning dynamics $\vec{\theta}^1, \vec{\theta}^2, \ldots, \vec{\theta}^t, \ldots$

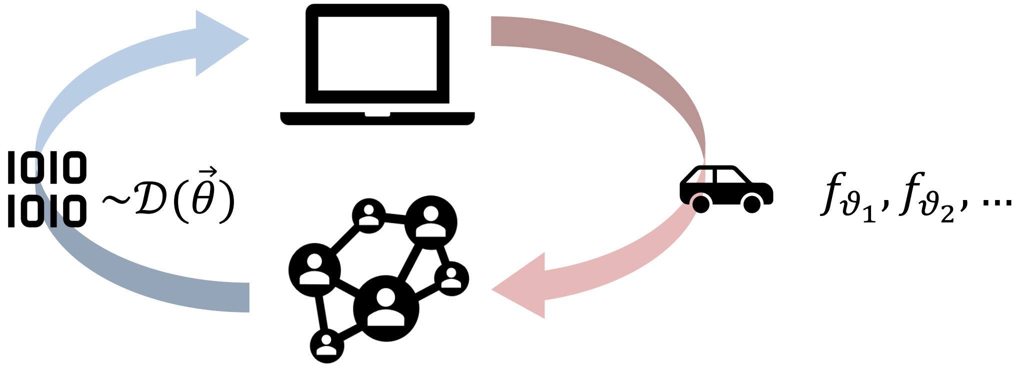
- Linear distribution mapping, $\mathcal{D}(\vec{\theta})$: given parameters, $\mathcal{D}_X$, θ_0, and influence parameter λ

$$x \sim \mathcal{D}_X, and\ y = x^\top \left(\theta_0 - \lambda \sum_{i \in [n]} \theta_i \right) + noise$$



Learning dynamics $\vec{\theta}^0, \vec{\theta}^1, \ldots, \vec{\theta}^t, \ldots$

- Linear distribution mapping, $\mathcal{D}(\vec{\theta})$: given $\mathcal{D}_X$, θ_0, and λ, $x \sim \mathcal{D}_X$ and $y = x^\top(\theta_0 - \lambda \sum_i \theta_i) + noise$
- Exponentiated gradient descent w/ learning rate η and initial conditions $\vec{\theta}^0$[KW97]

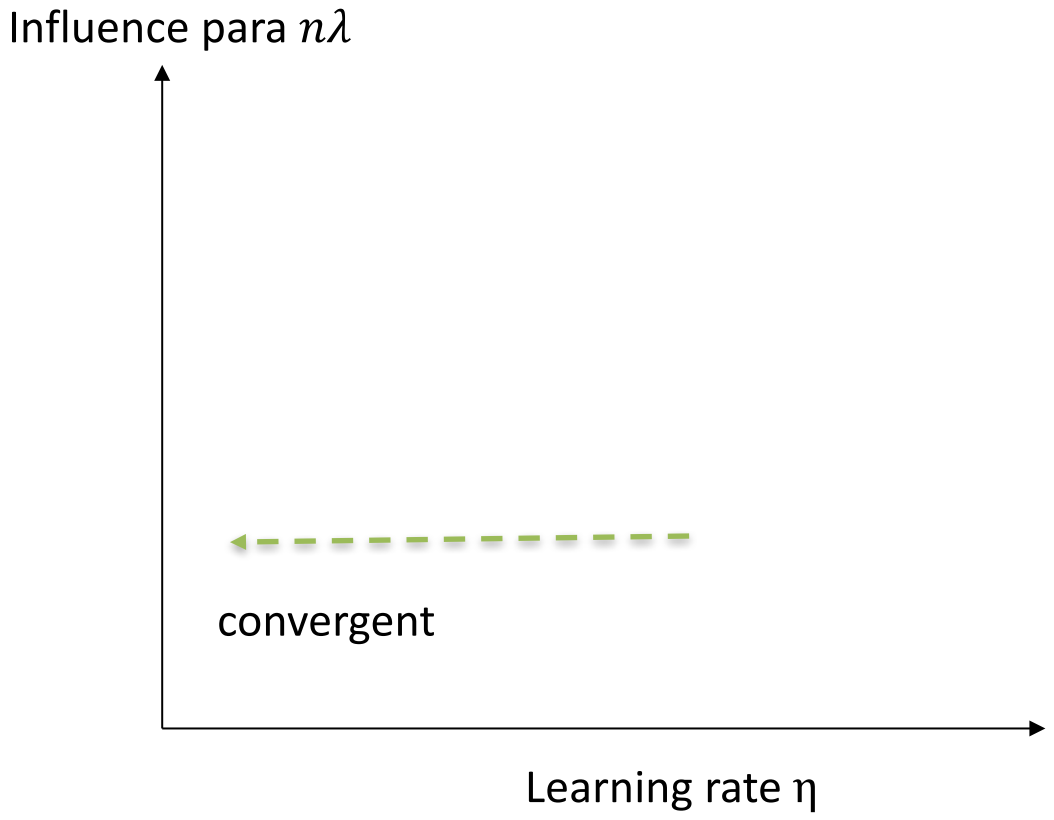


Jyrki Kivinen and Manfred K Warmuth. Exponentiated gradient versus gradient descent for linear predictors. information and computation, 132(1):1–63, 1997.

A threshold result on MAPP

Consider n learning agents with influence parameter λ using exponentiated GD with learning rate η.

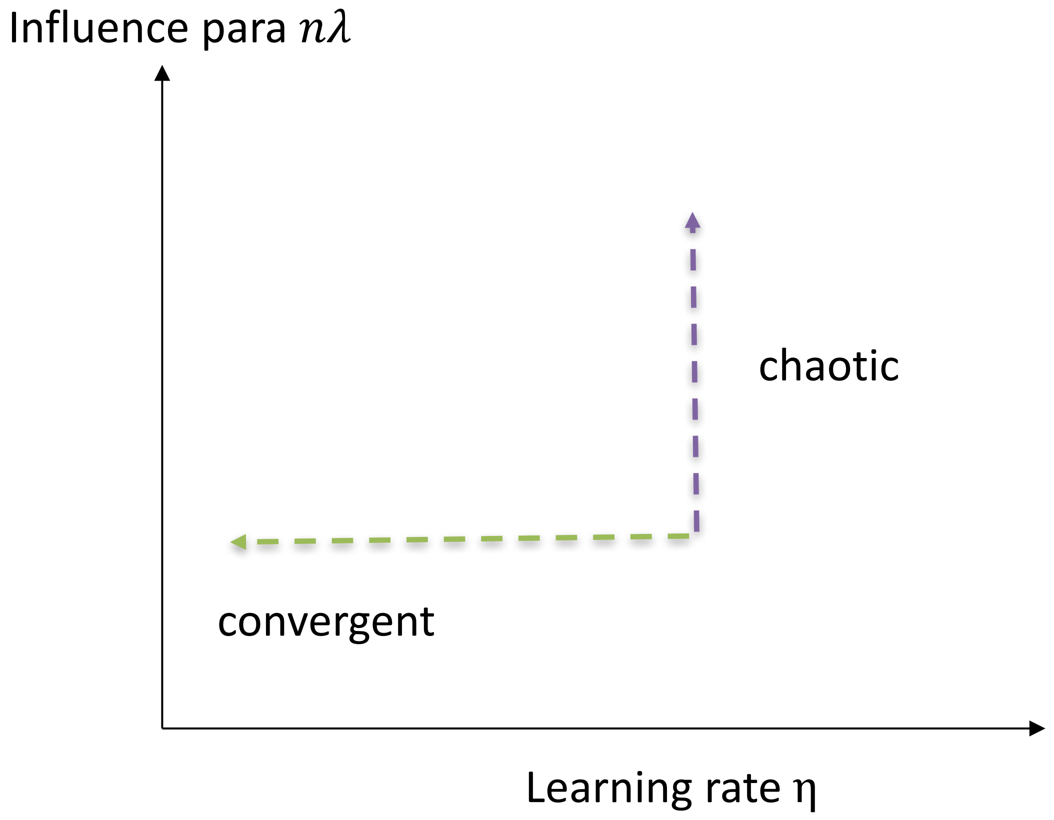
1. Fixing n and λ, if η is small enough, $\lim_{t\to\infty} \vec{\theta}^t$ is performative stable.



A threshold result on MAPP

Consider n learning agents with influence parameter λ using exponentiated GD with learning rate η.

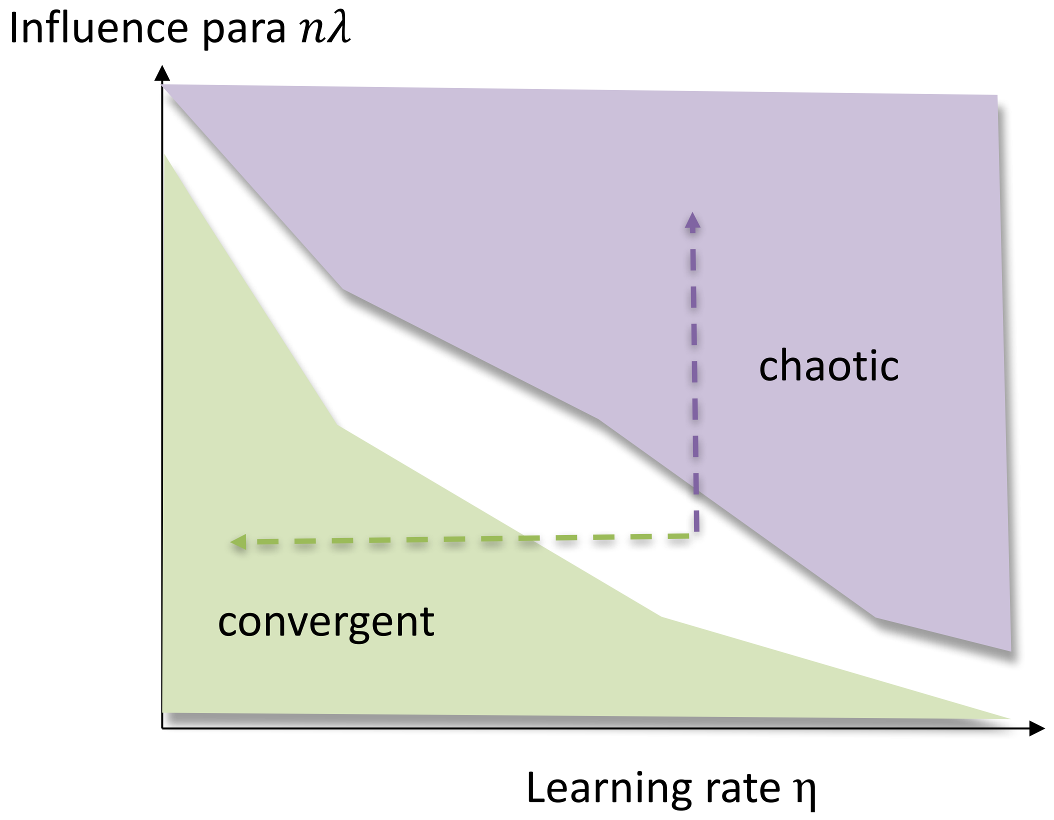
1. Fixing n and λ, if η is small enough, $\lim_{t\to\infty} \vec{\theta}^t$ is performative stable.
2. Fixing η, if $n\lambda$ is large enough, $\vec{\theta}^t$ is Li-Yorke chaotic.



A threshold result on MAPP

Consider n learning agents with influence parameter λ using exponentiated GD with learning rate η.

1. Fixing n and λ, if η is small enough, $\lim_{t\to\infty} \vec{\theta}^t$ is performative stable.
2. Fixing η, if $n\lambda$ is large enough, $\vec{\theta}^t$ is Li-Yorke chaotic.



Proof idea

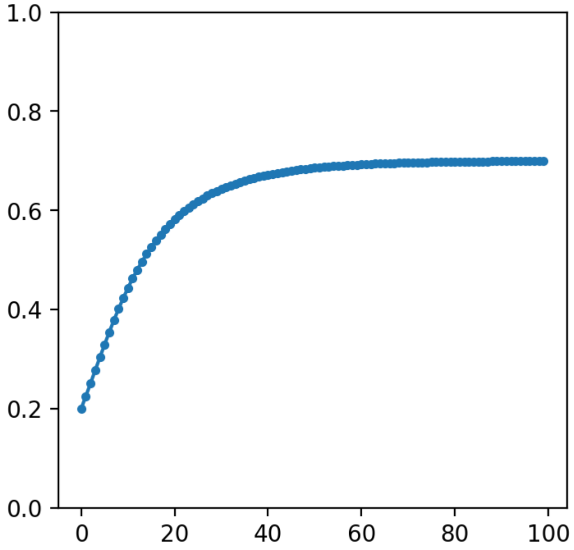
Connection of Hedge learning on congestion game with linear cost

- Convergence
  - Replicator dynamics ($\eta \to 0$) on congestion game [KPT09]
  - Linear MWU with constant η on congestion game [PPP 17]
- Li Yorke chaos
  - Large learning rate in Hedge [CFMP 20]

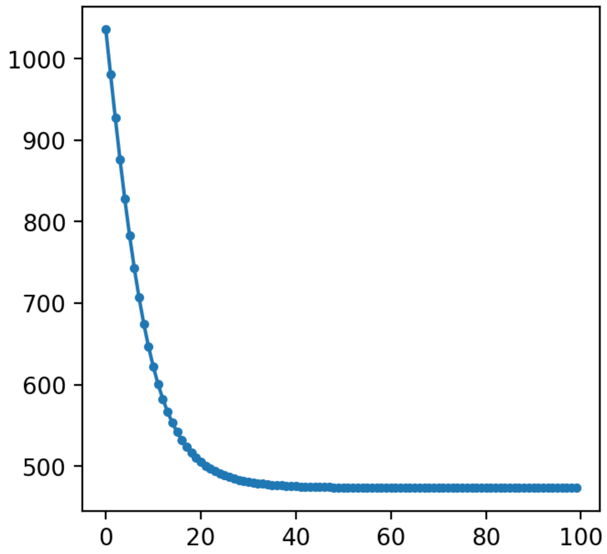
Robert Kleinberg, Georgios Piliouras, and Éva Tardos. 2009. Multiplicative updates outperform generic no-regret learning in congestion games.
Gerasimos Palaiopanos, Ioannis Panageas, and Georgios Piliouras. 2017. Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos.
Thiparat Chotibut, Fryderyk Falniowski, Michał Misiurewicz, and Georgios Piliouras. 2020. The route to chaos in routing games: When is Price of Anarchy too optimistic?

Simulation

Small learning rate (η)

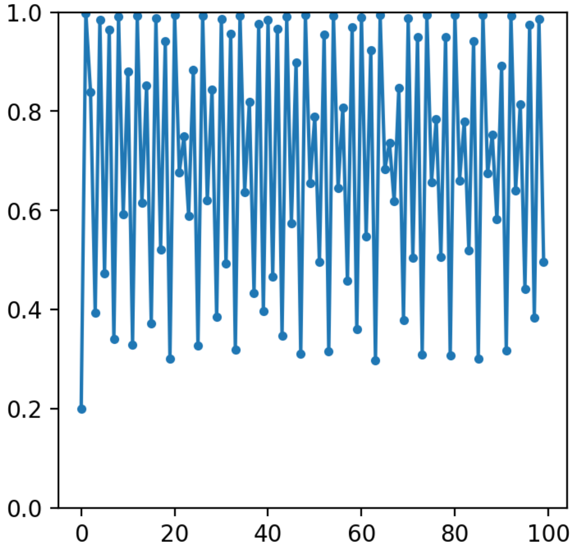


Model θ^t

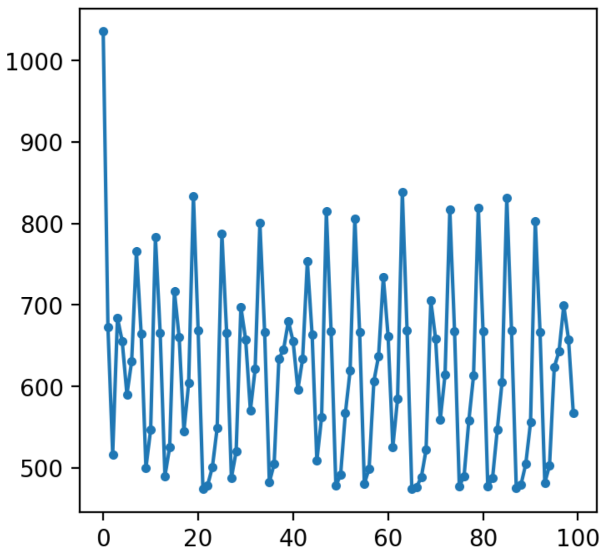


Loss $\ell\left(\theta^t; \mathcal{D}(\theta^t)\right)$

Large influence ($n\lambda$)



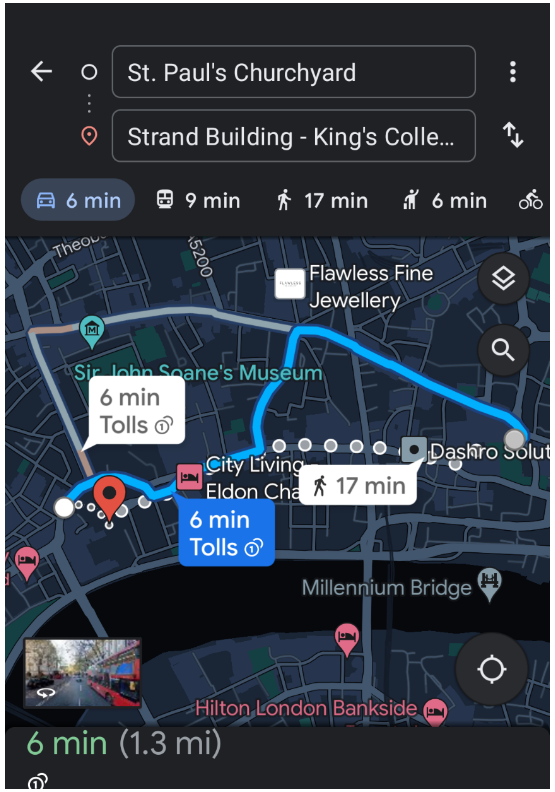
Model θ^t



Loss $\ell\left(\theta^t; \mathcal{D}(\theta^t)\right)$

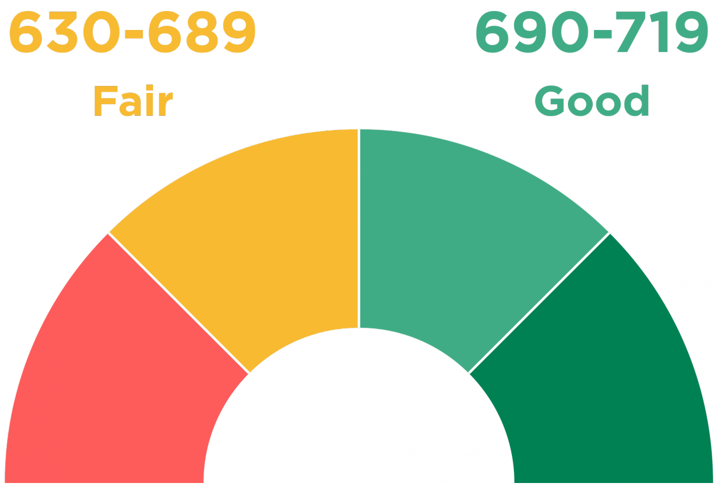
Connections

Traffic prediction



More drivers follow=> large λ=> chaos

Credit score



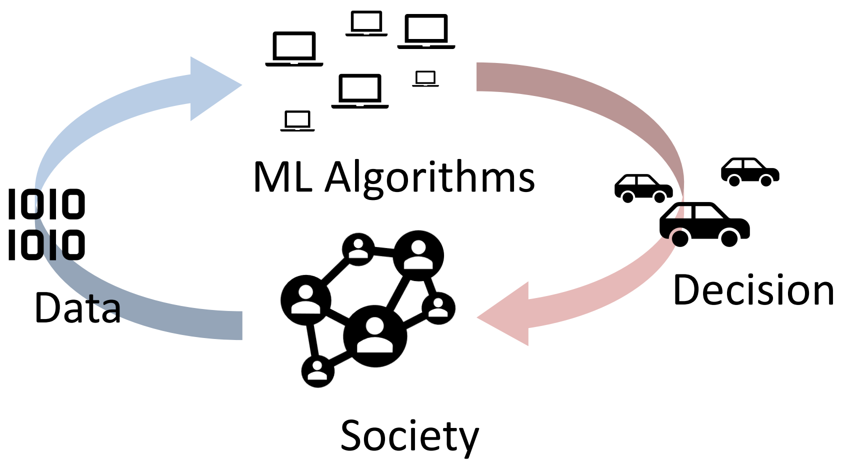
More banks=> large n=> chaos

Policy patrol with predictions

More dependency => large λ=> chaos

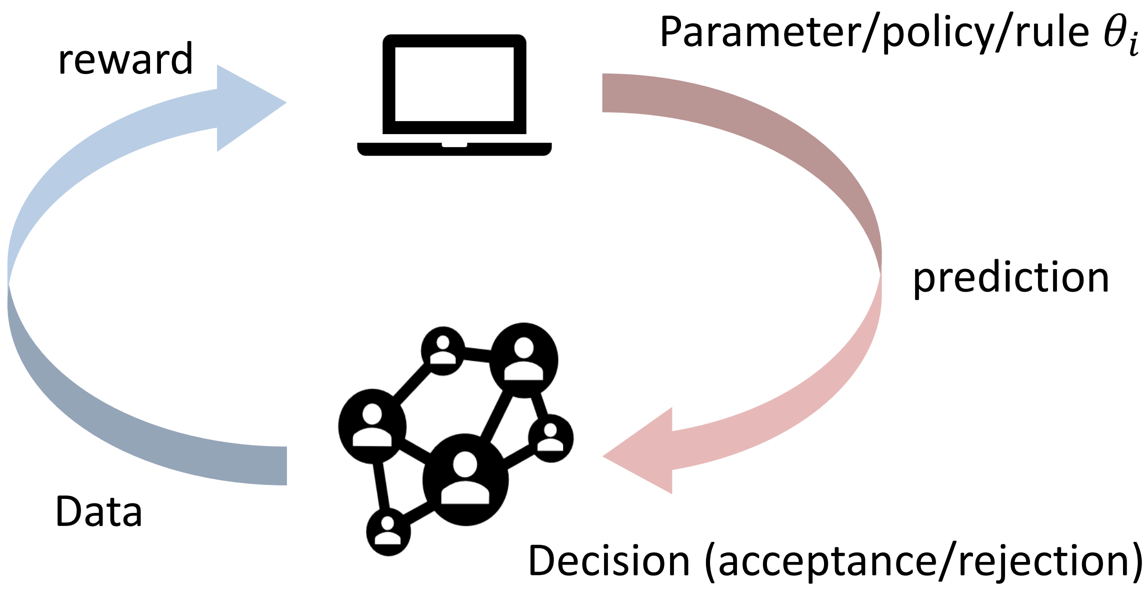
Open questions

- Multi agent performative prediction
  - How can we learn and avoid chaos?
    - Algorithm
    - Mechanism
  - Competition between learning agents
- When are predictions performative?
  - Strategic classification
  - Price competition
  - Recommendation system

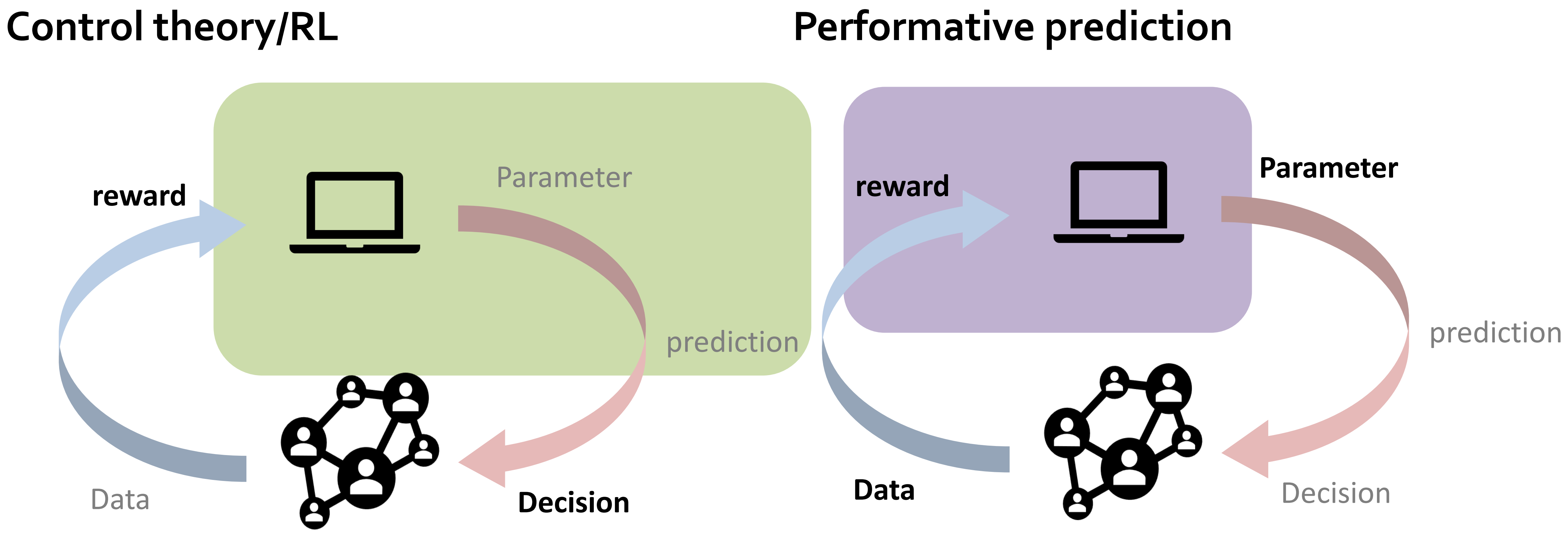


Some thoughts

- Which is performative?
  - Decision
  - Prediction
  - Policy
- Examples of θ_i
  - Navigation app's recommendation algorithms
  - College's acceptance rule
  - Hedge fund's automatic bidder

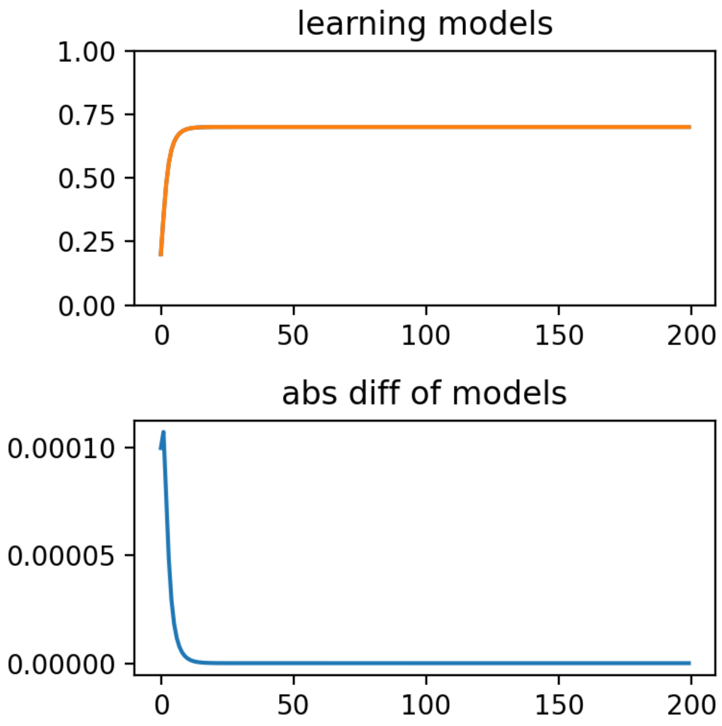


Some thoughts



Simulation

Small learning rate (η)



Large influence ($n\lambda$)