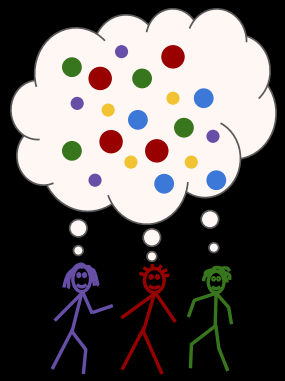
Market-Based Mechanisms for Acquiring and Aggregating Data



Bo Waggoner Microsoft NYC TTIC, Aug 2018

based on work with Jacob Abernethy (Georgia Tech)

and Rafael Frongillo (U. Colorado)

Question

How to **procure data** from **strategic sources**?

Question

How to **procure data** from **strategic sources**?

Interesting because: quantifies **value of information**

Question

How to **procure data** from **strategic sources**?

Interesting because: quantifies **value of information**

Not addressed today: **crowdsourcing** approaches...

Instead: how to **pay strategic agents for data**

One line of work

[Abernethy, Chen, Ho, **W** EC 2015][1]:

[1] cf [Roth, Schoenebeck EC 2011]; [Chen, Immorlica, Lucier, Syrgkanis, Ziani EC 2018]

One line of work

[Abernethy, Chen, Ho, **W** EC 2015][1]:

- convex loss function, hypothesis $w \in \mathbb{R}^d$
- agents hold i.i.d. data and **cost to reveal**
- agents sequentially offered random price menu
  prices $\propto$ **value** *of data*
- prove generalization error $O\left(\sqrt{\frac{\gamma}{\text{Budget}}}\right)$

[1] cf [Roth, Schoenebeck EC 2011]; [Chen, Immorlica, Lucier, Syrgkanis, Ziani EC 2018]

One line of work

[Abernethy, Chen, Ho, **W** EC 2015][1]:

- convex loss function, hypothesis $w \in \mathbb{R}^d$
- agents hold i.i.d. data and **cost to reveal**
- agents sequentially offered random price menu
  prices $\propto$ **value** *of data*
- prove generalization error $O\left(\sqrt{\frac{\gamma}{\text{Budget}}}\right)$

Agents bid strategically, cannot modify or falsify data

future work!

[1] cf [Roth, Schoenebeck EC 2011]; [Chen, Immorlica, Lucier, Syrgkanis, Ziani EC 2018]

Today: (prediction) market design

First cut: incentivize **experts** to aggregate knowledge.

[Abernethy, Frongillo NIPS 2011]:

Today: (prediction) market design

First cut: incentivize **experts** to aggregate knowledge.

[Abernethy, Frongillo NIPS 2011]:

- Choose a **loss function** $\ell(\text{hypothesis}, \text{data})$
- Choose initial hypothesis h^0
- Experts $1, \ldots, T$ arrive and update $h^{t-1} \to h^t$
- Reveal a holdout/test data set

Today: (prediction) market design

First cut: incentivize **experts** to aggregate knowledge.

[Abernethy, Frongillo NIPS 2011]:

- Choose a **loss function** $\ell(\text{hypothesis}, \text{data})$
- Choose initial hypothesis h^0
- Experts $1, \ldots, T$ arrive and update $h^{t-1} \to h^t$
- Reveal a holdout/test data set
- Payoff to t is

$$\ell(h^{t-1}, \text{data}) \; - \; \ell(h^t, \text{data})$$

Today: (prediction) market design

First cut: incentivize **experts** to aggregate knowledge.

[Abernethy, Frongillo NIPS 2011]:

- Choose a **loss function** $\ell(\text{hypothesis}, \text{data})$
- Choose initial hypothesis h^0
- Experts $1, \ldots, T$ arrive and update $h^{t-1} \to h^t$
- Reveal a holdout/test data set
- Payoff to t is

$$\ell(h^{t-1}, \text{data}) \; - \; \ell(h^t, \text{data})$$

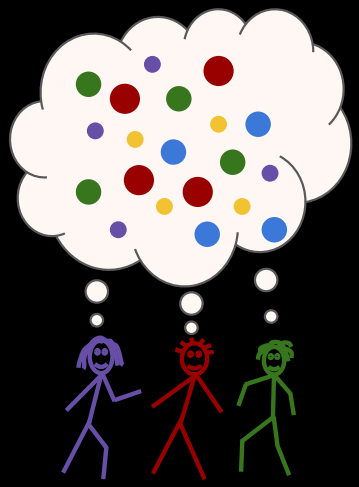
Designer pays: $\ell(h^0, \text{data}) - \ell(h^T, \text{data})$

Extensions: markets for data

[W, Frongillo, Abernethy NIPS 2015]:

- Conditional (generalized regression) markets
- Kernel-ization
- Differentially private in data/modifications

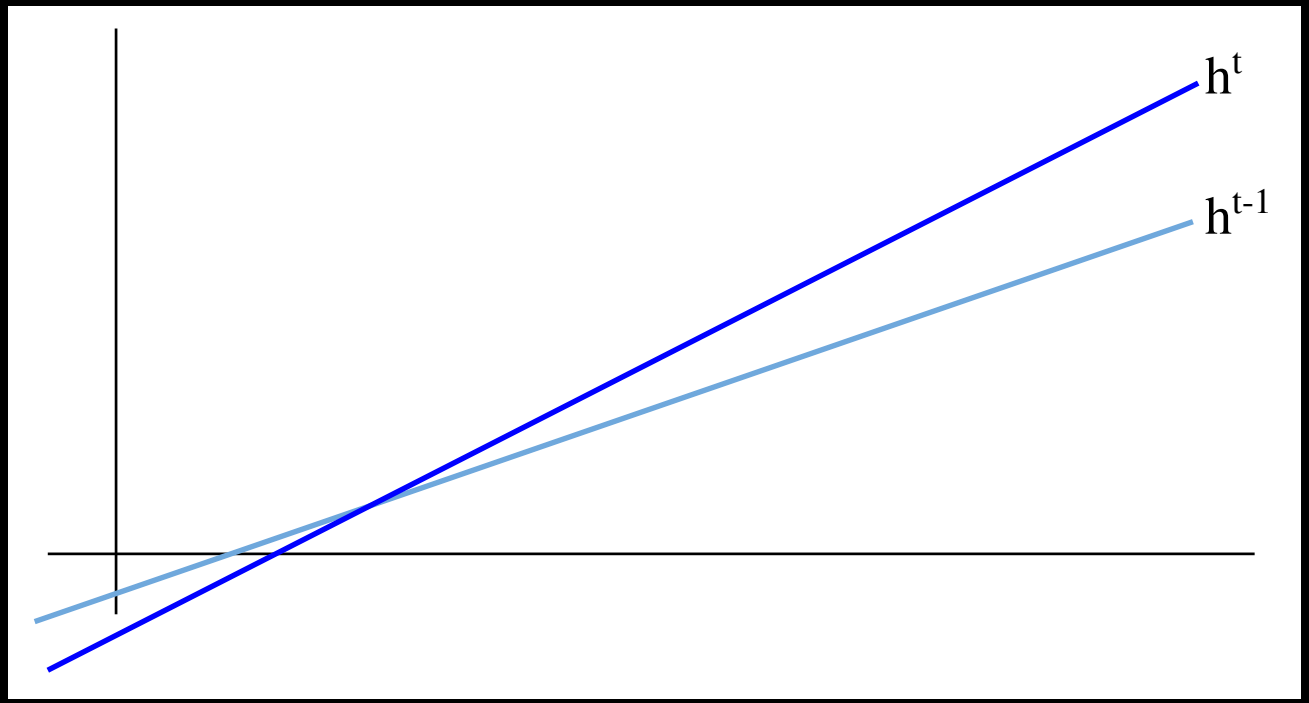
  not covered today



Example: linear regression

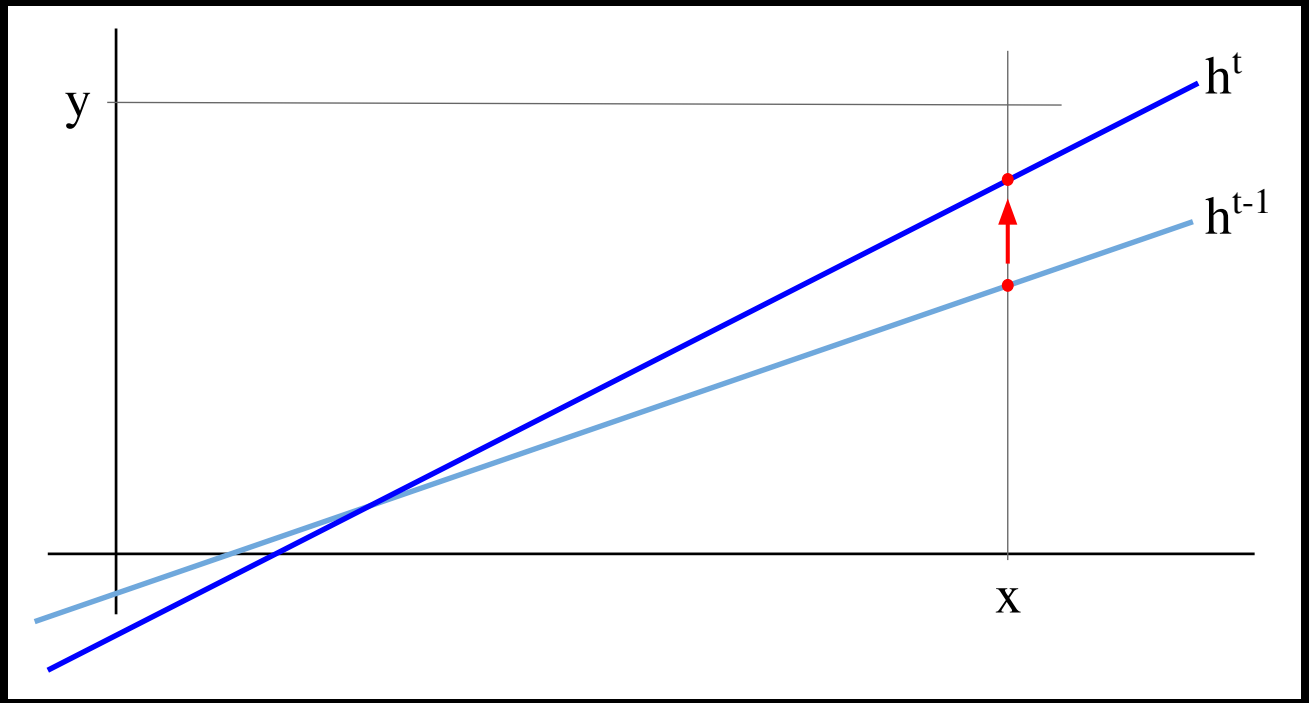
Market hypothesis: $h \in \mathbb{R}^d$

$\hat{y} = h \cdot x$



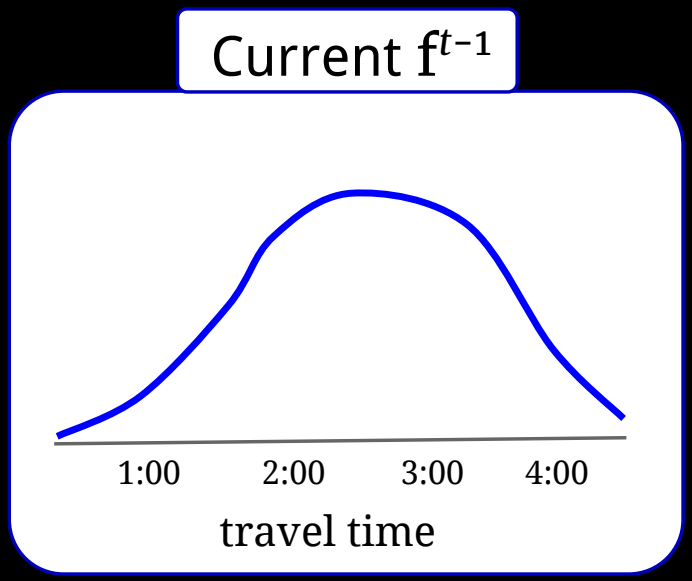
Example: linear regression

Market hypothesis: $h \in \mathbb{R}^d$ $\hat{y} = h \cdot x$

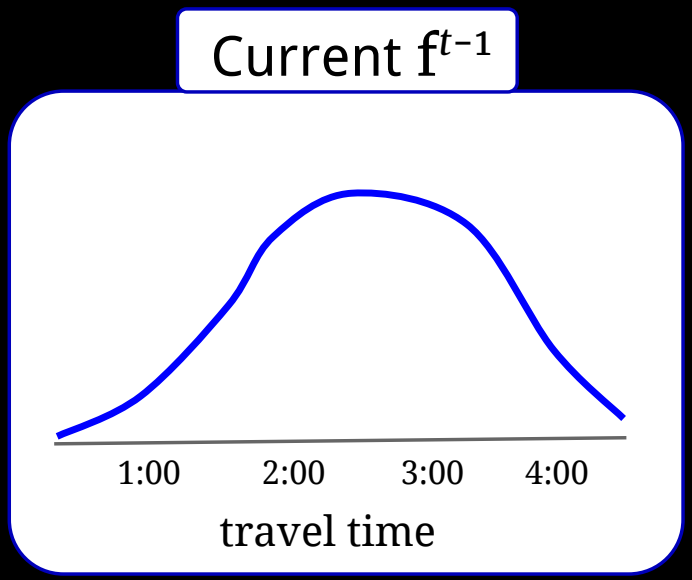


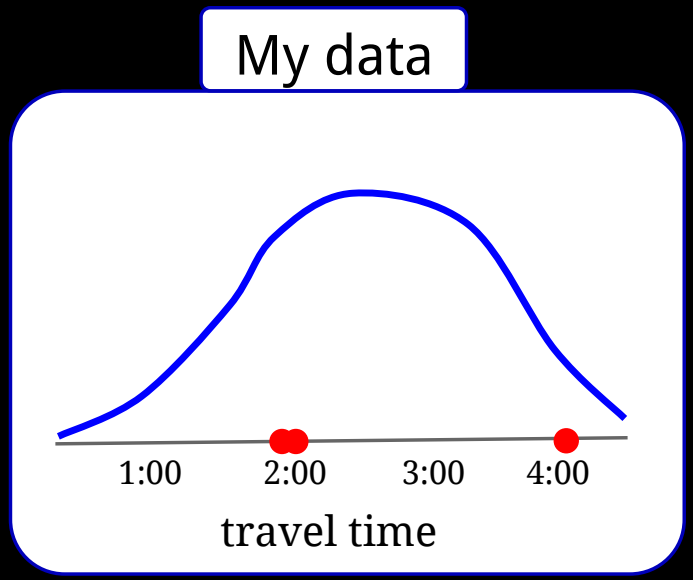
Pay $\approx h^{t-1} \cdot x$ per "share"; get payoff y

Example: kernel-ization

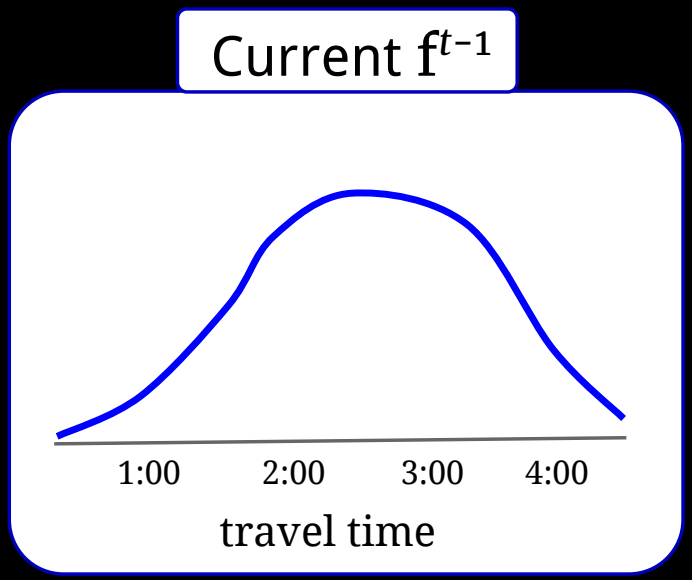


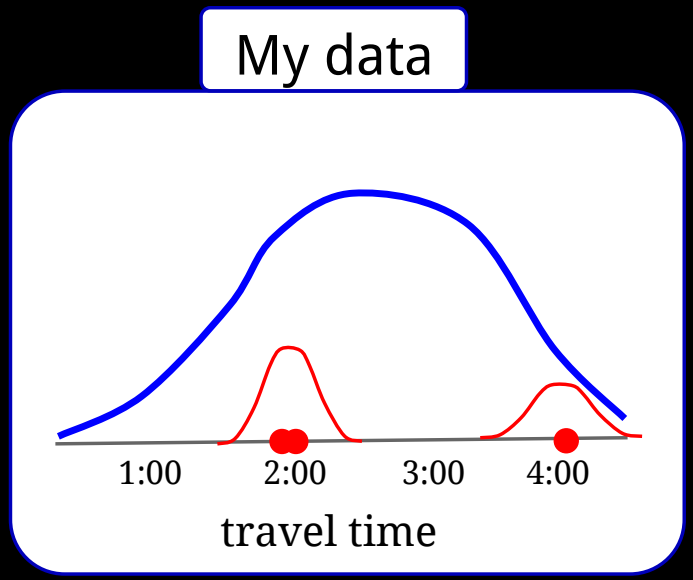
Example: kernel-ization





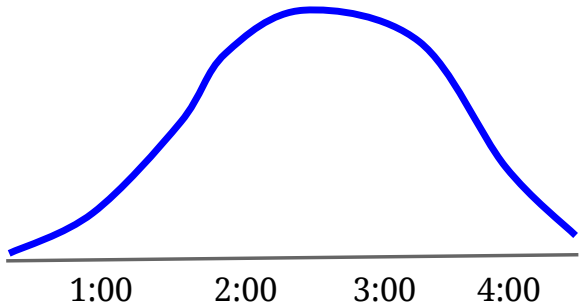
Example: kernel-ization



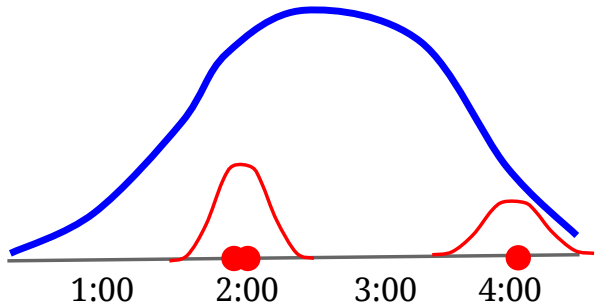


Example: kernel-ization

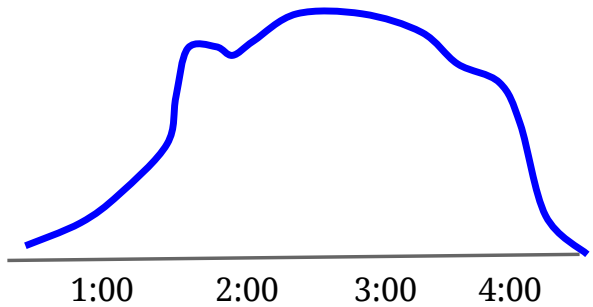
Current f^{t-1}



My data

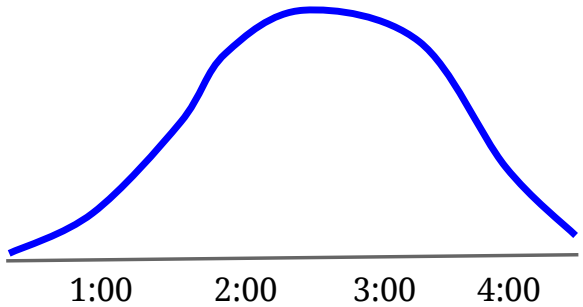


Updated f^{t}

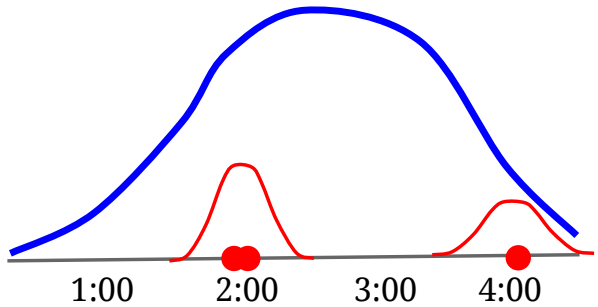


Example: kernel-ization

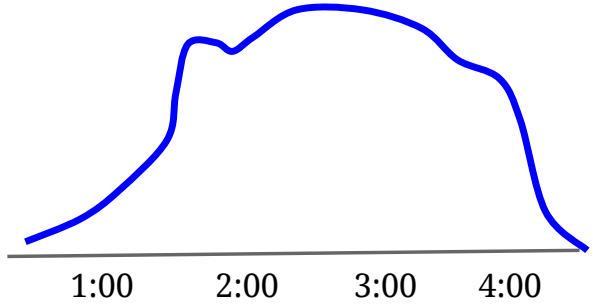
Current f^{t-1}



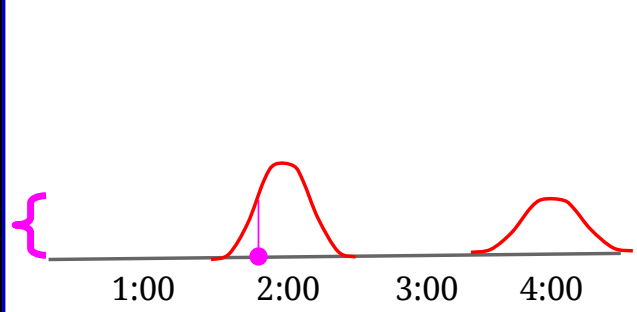
My data



Updated f^{t}



My reward



When do you get a "market"?

Theorem (Frongillo, W ITCS 2018)

A learning mechanism can be written as a full "market" if and only if: the loss function is a **Bregman divergence**

Full market: *can resell previously-purchased contracts.*

When do you get a "market"?

Theorem (Frongillo, W ITCS 2018)

A learning mechanism can be written as a full "market" if and only if: the loss function is a **Bregman divergence**

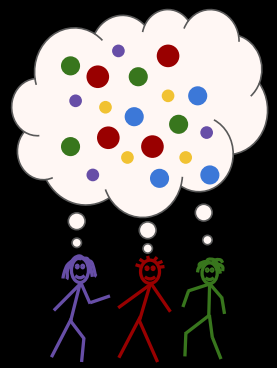
equivalently

the learning problem is to **predict the mean**.

Full market: *can resell previously-purchased contracts.*

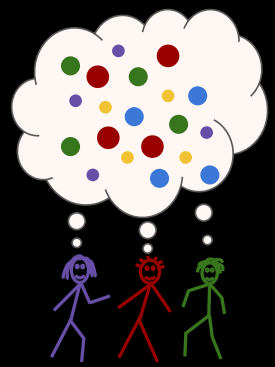
Key points

- Understanding **data procurement** is interesting
- Market mechanism:
  - **aligns incentives**
  - interface with experts and data-providers
  - theory of elicitation $\rightarrow$ implications for design



Key points

- Understanding **data procurement** is interesting
- Market mechanism:
  - **aligns incentives**
  - interface with experts and data-providers
  - theory of elicitation $\rightarrow$ implications for design



Tons of open directions for data procurement!

Thanks!