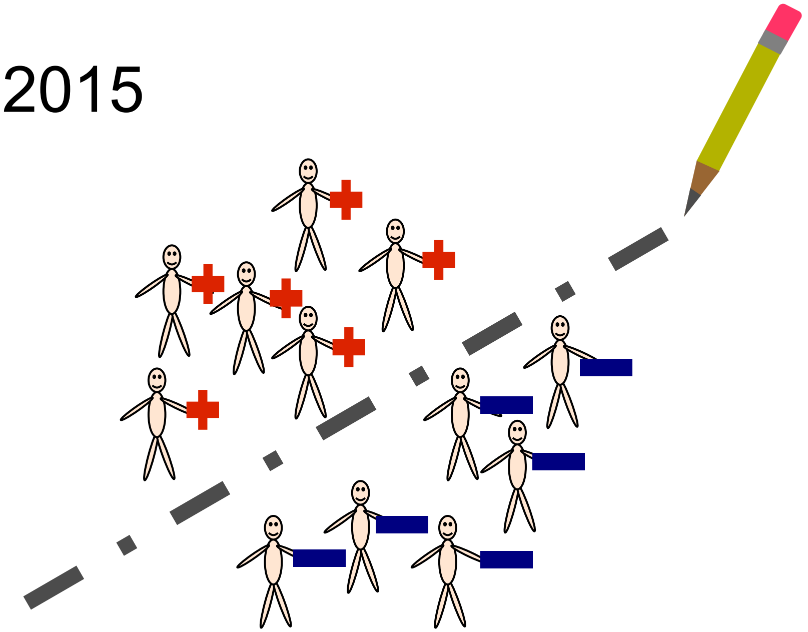
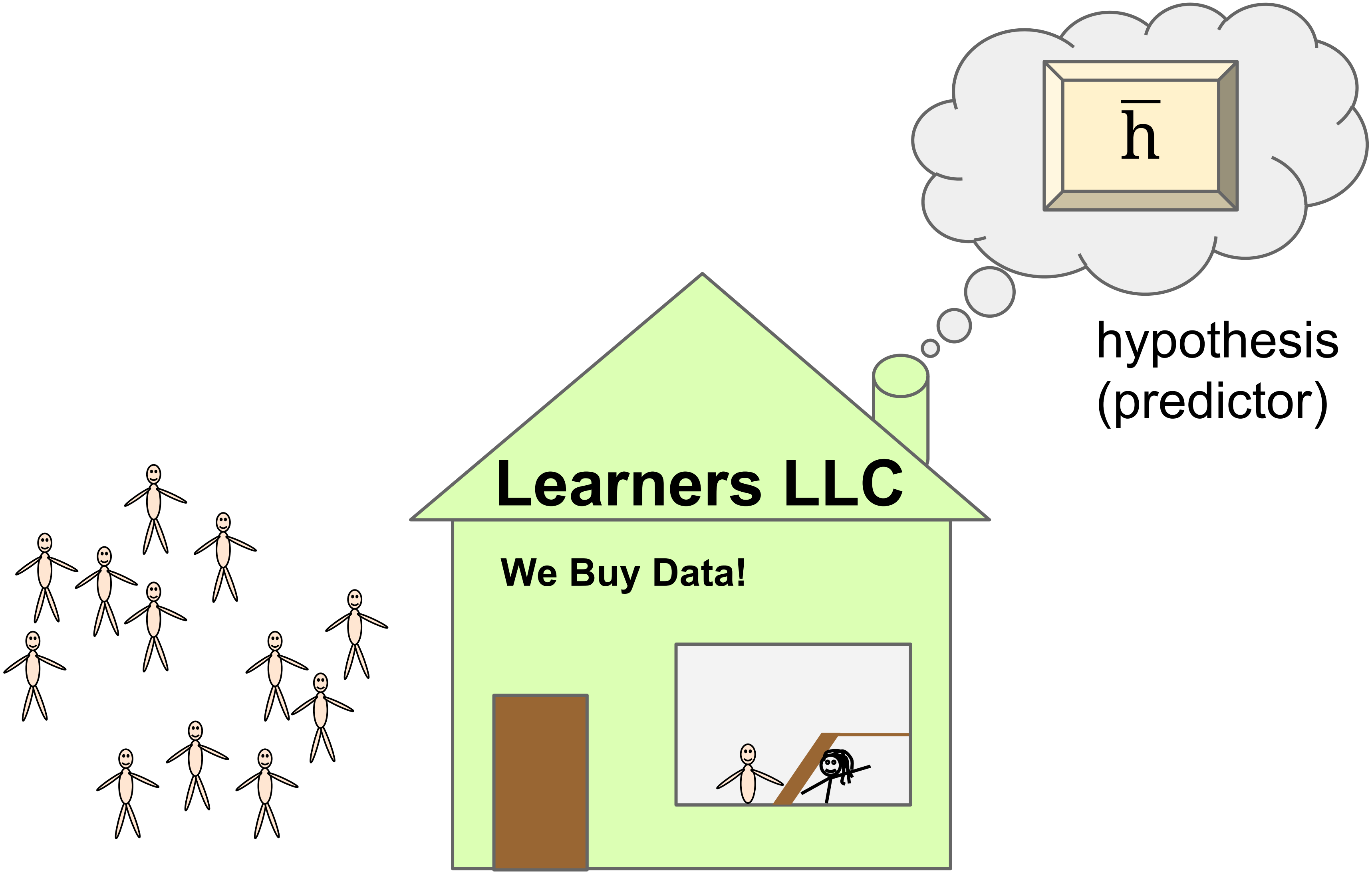
Low-Cost Learning via Active Data Procurement

EC 2015



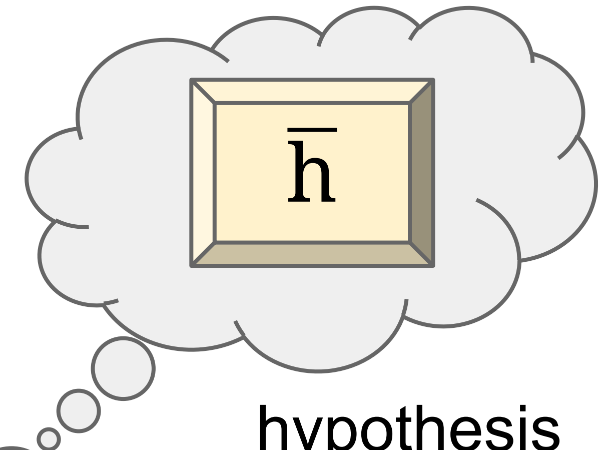
Jacob Abernethy
Yiling Chen
Chien-Ju Ho
Bo Waggoner

General problem: buy data for learning



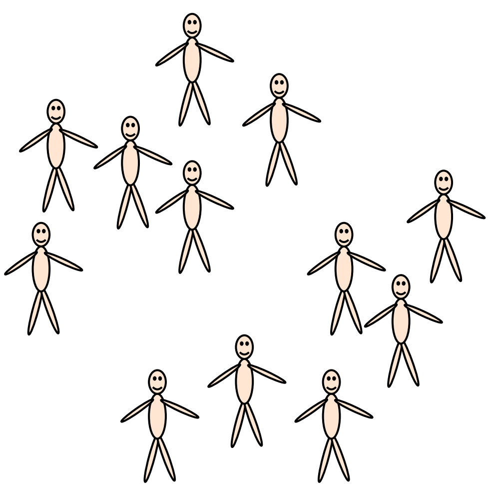
General problem: buy data for learning

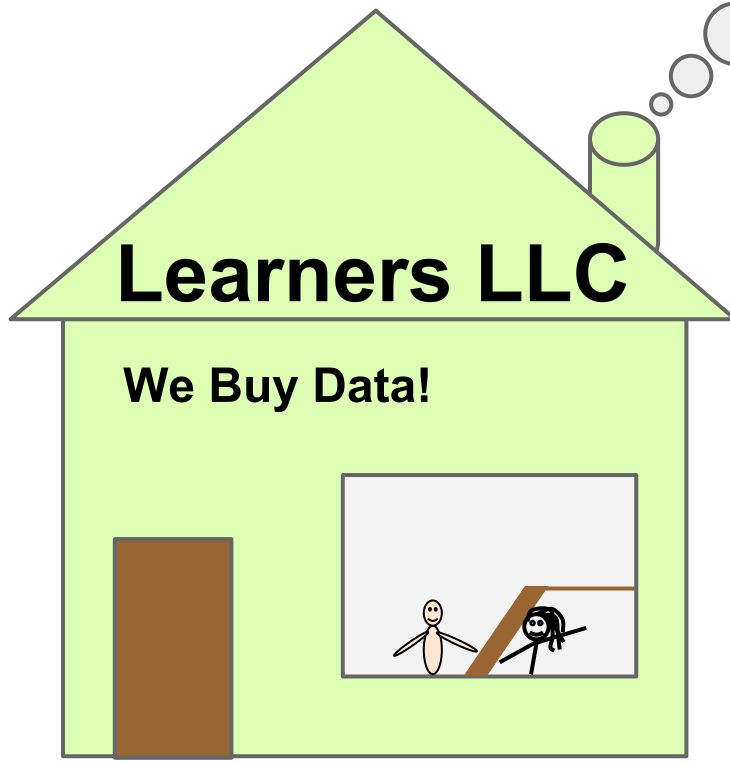
… learn to predict disease



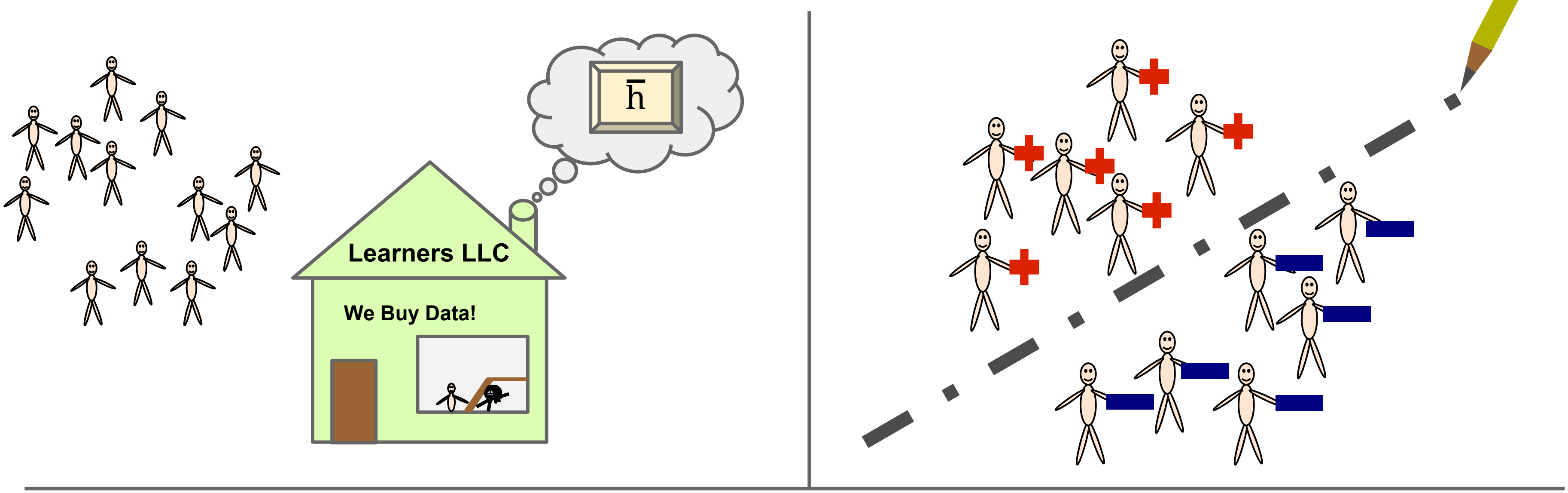
hypothesis (predictor)

Example: each person has medical data...



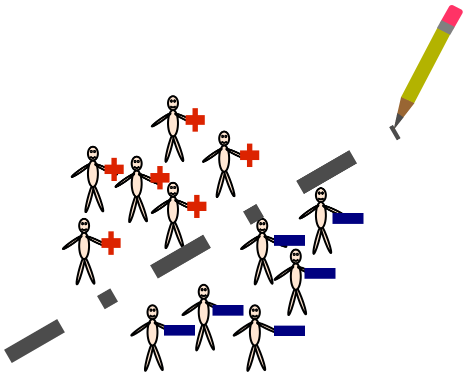


Example task: classification



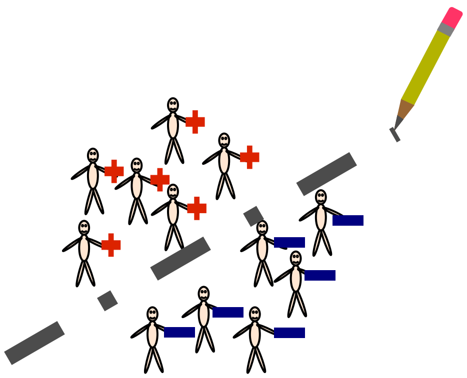
- **Data point**: pair (x, label) where label is + or −
- **Hypothesis**: hyperplane separating the two types
- **Loss**: 0 if $h(x)$ = correct label, 1 if incorrect label
- **Goal**: pick $\bar{h}$ with low expected loss on new data point

General Goal:

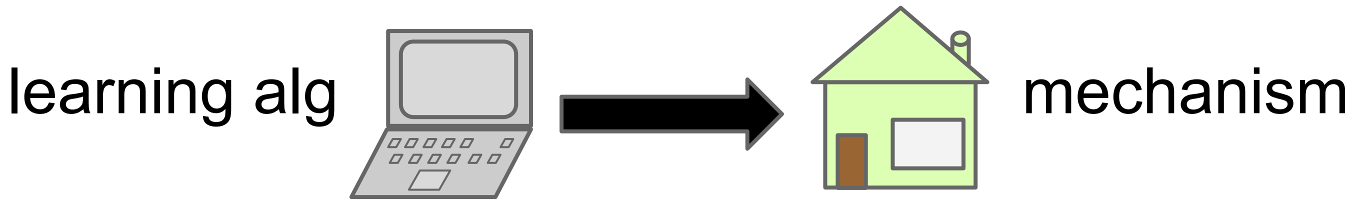


Learn a good hypothesis
by purchasing data from the crowd

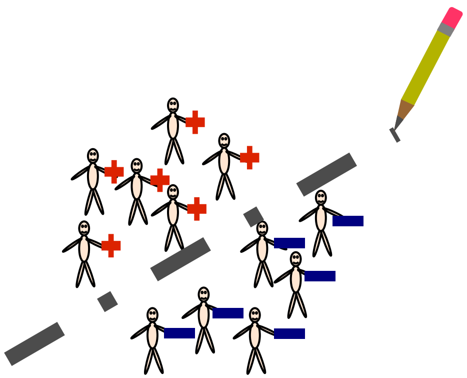
This paper:



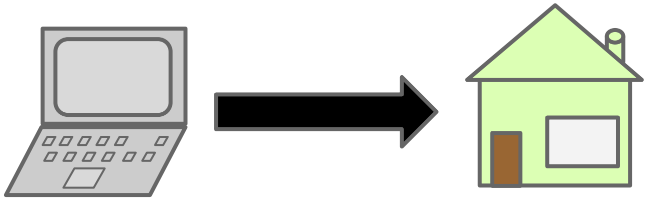
1. price data actively based on value
2. machine-learning style bounds
3. transform learning algs to mechanisms



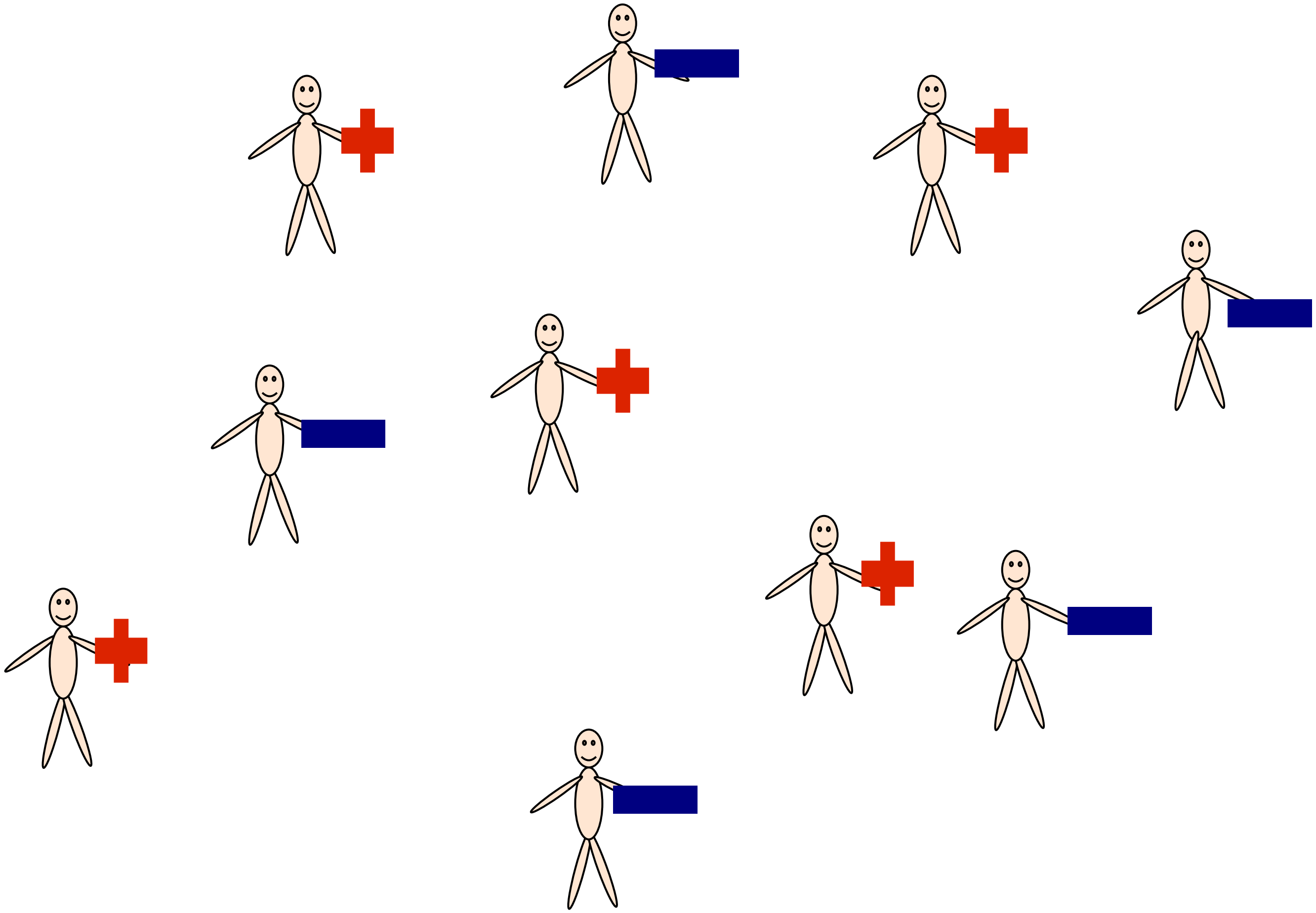
This paper:



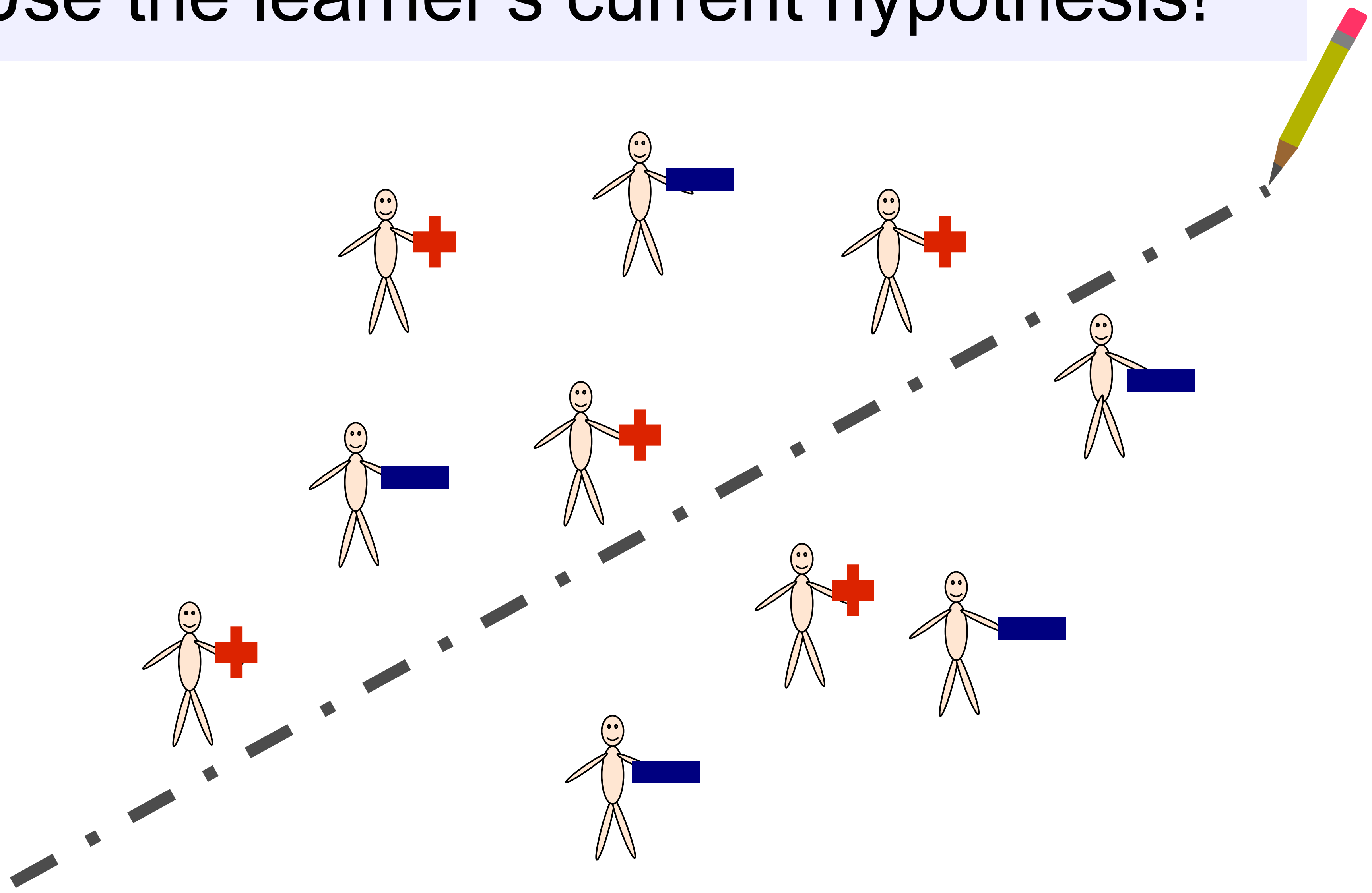
1. price data actively based on value
2. machine-learning style bounds
3. transform learning algs to mechanisms

learning alg  mechanism

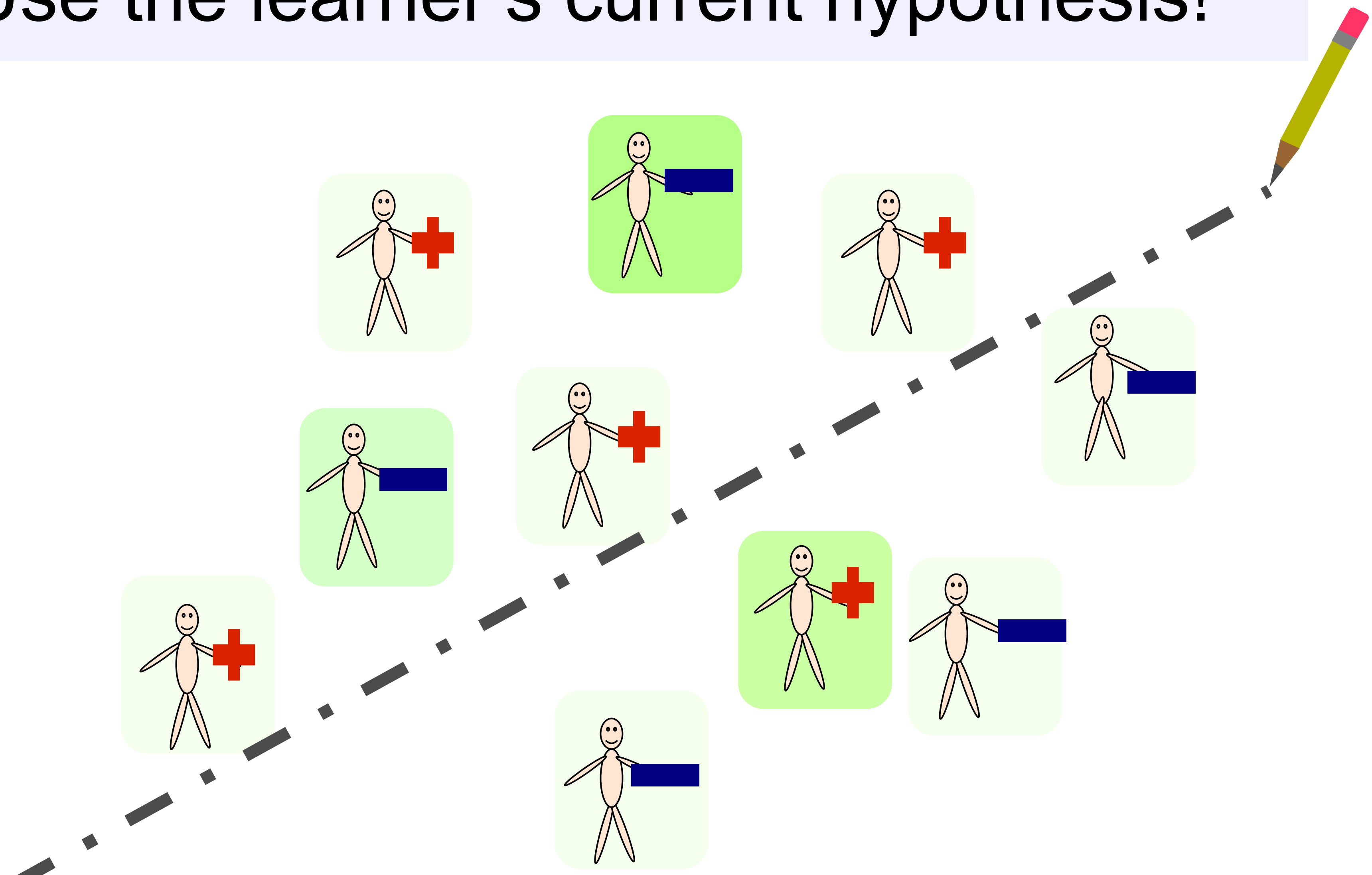
How to assess value/price of data?



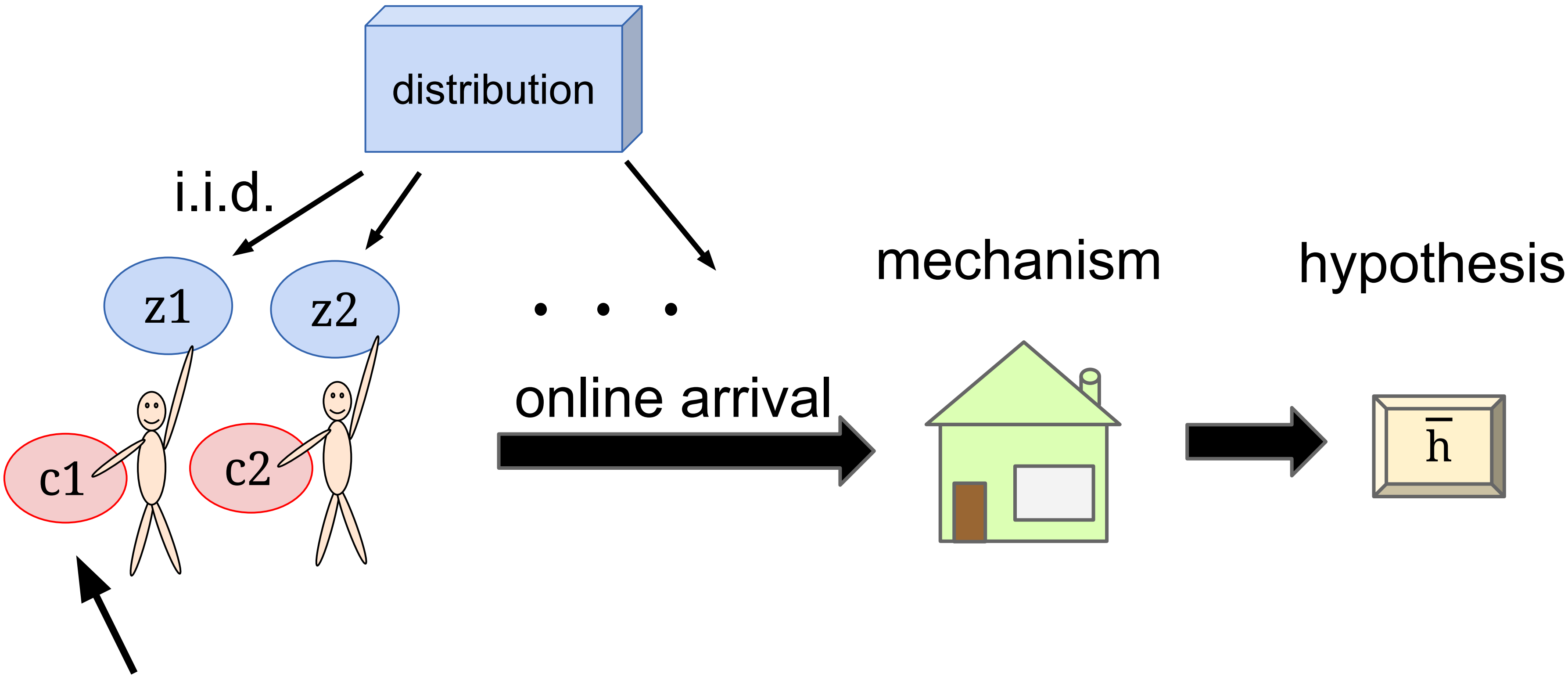
Use the learner's current hypothesis!



Use the learner's current hypothesis!



Our model



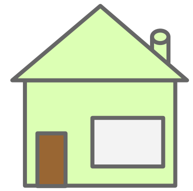
Cost of revealing data

- lies in [0,1]
- worst-case, arbitrarily correlated with the data

Agent-mechanism interaction

At each time $t = 1, \ldots, T$:

1. mechanism posts **menu**

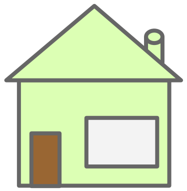


data:	65 −	30 +	65 +
price:	$0.22	$0.41	$0.88

Agent-mechanism interaction

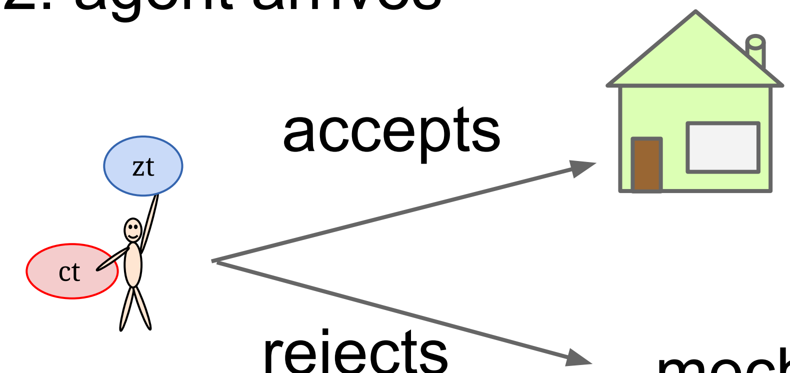
At each time t = 1, …, T:

1. mechanism posts **menu**



data:	65 −	30 +	65 +
price:	\$0.22	\$0.41	\$0.88

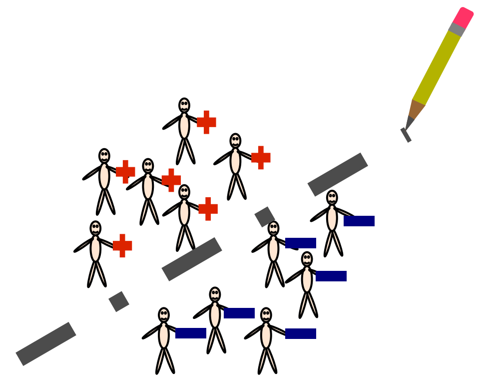
2. agent arrives



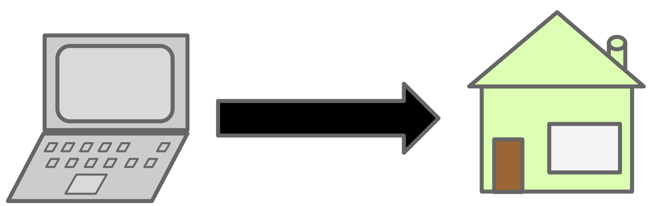
mechanism learns (zt, ct)
and pays price(zt)

mechanism sees rejection
and pays nothing

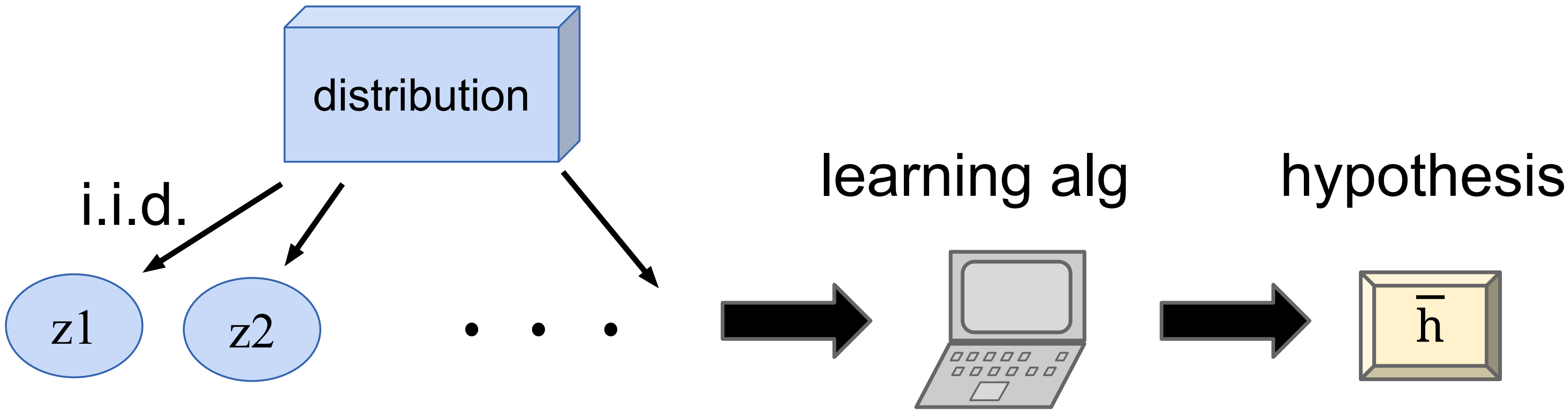
This paper:



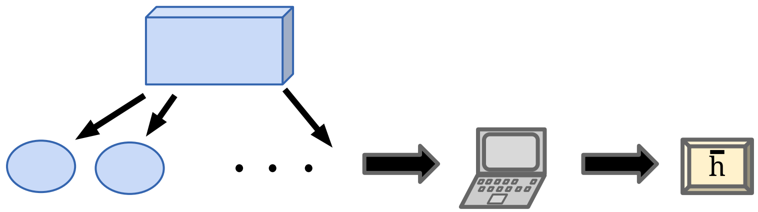
1. price data actively based on value
2. machine-learning style bounds
3. transform learning algs to mechanisms

learning alg  mechanism

What is the “classic” learning problem?



Classic ML bounds



measure of problem difficulty

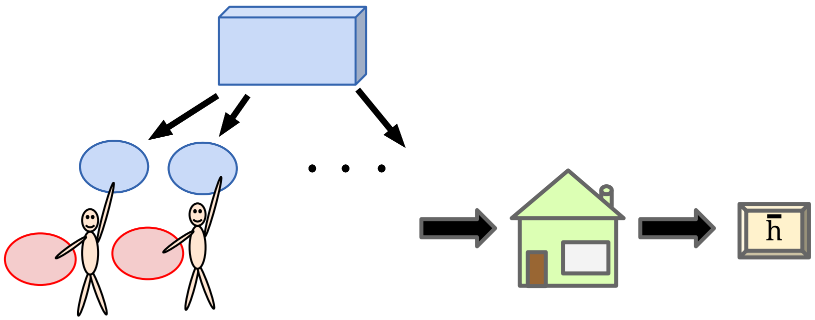
$$\mathrm{E}\,\mathrm{loss}(\bar{h}) \le \mathrm{E}\,\mathrm{loss}(h^*) + O\left(\sqrt{\frac{\text{VC-dim}}{T}}\right)$$

alg's hypothesis

optimal hypothesis

# of data points

Main result



measure of "problem difficulty", in [0,1].

For a variety of learning problems:

$$\mathrm{E}\,\mathrm{loss}(\,\bar{h}\,) \le \mathrm{E}\,\mathrm{loss}(\,h^*\,) + O\left(\sqrt{\frac{\gamma}{B}}\right)$$

our hypothesis

optimal hypothesis

Budget constraint

(Assume: γ is approximately known in advance)

Main result

$\gamma \approx$ average cost * difficulty

"if problem is **cheap** or **easy** or has **good correlations**, we do well"

For a variety of learning problems:

$$\mathrm{E}\,\mathrm{loss}(\,\bar{h}\,) \le \mathrm{E}\,\mathrm{loss}(\,h^*\,) + O\left(\sqrt{\frac{\gamma}{B}}\right)$$

our hypothesis

optimal hypothesis

Budget constraint

(Assume: γ is approximately known in advance)

Related work in purchasing data

Type of goal

Roth, Schoenebeck 2012

Ligett, Roth 2012

Horel, Ionnadis, Muthukrishnan 2014

Cummings, Ligett, Roth, Wu, Ziani 2015

Cai, Daskalakis, Papadimitriou 2015

Dekel, Fisher, Procaccia 2008

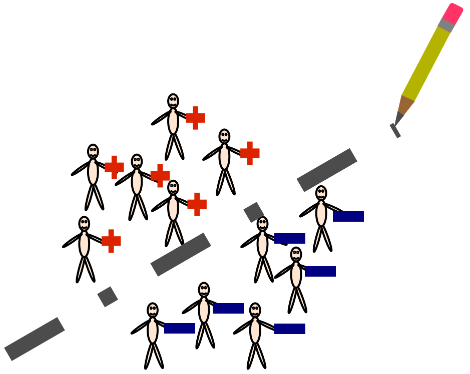
Ghosh, Ligett, Roth, Schoenebeck 2014

Meir, Procaccia, Rosenschein 2012

this work

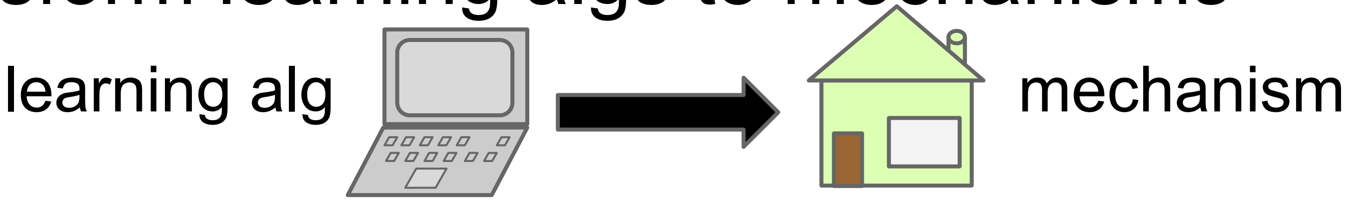
Model

This paper:



Key features/ideas:

1. price data actively based on value
2. machine-learning style bounds
3. transform learning algs to mechanisms

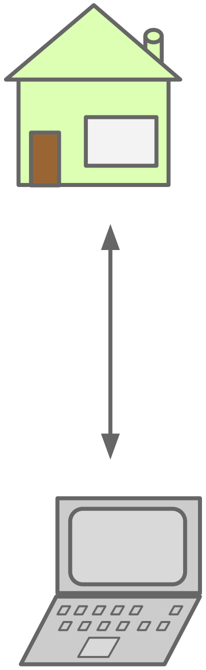


Learning algorithms: FTRL

- Follow-The-Regularized-Leader (FTRL)
  (Multiplicative Weights, Online Gradient Descent, ….)

- FTRL algs do “no regret” learning:
  - output a hypothesis at *each* time
  - want low total loss

- we interface with FTRL as a black box…
  … but analysis relies on “opening the box”

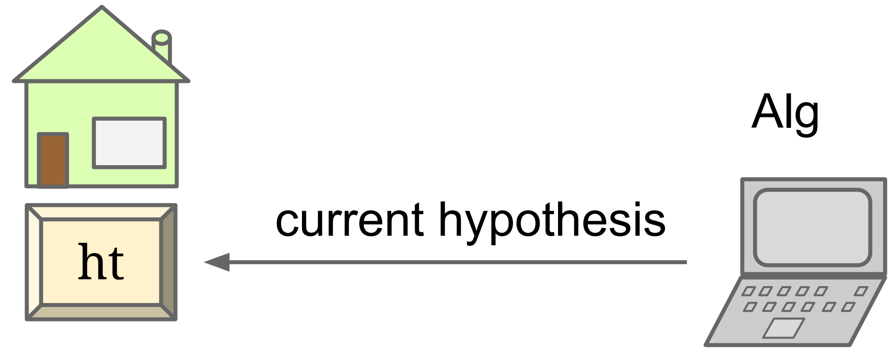


Our mechanism

At each time t = 1, …, T:

1. post menu

price(z) ~ distribution(ht, z)

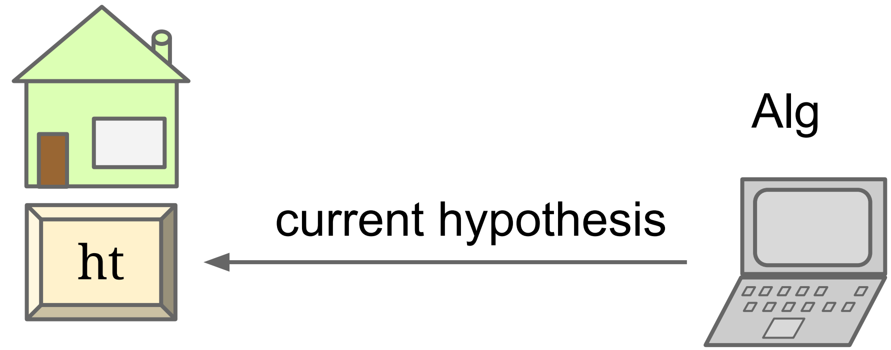


Our mechanism

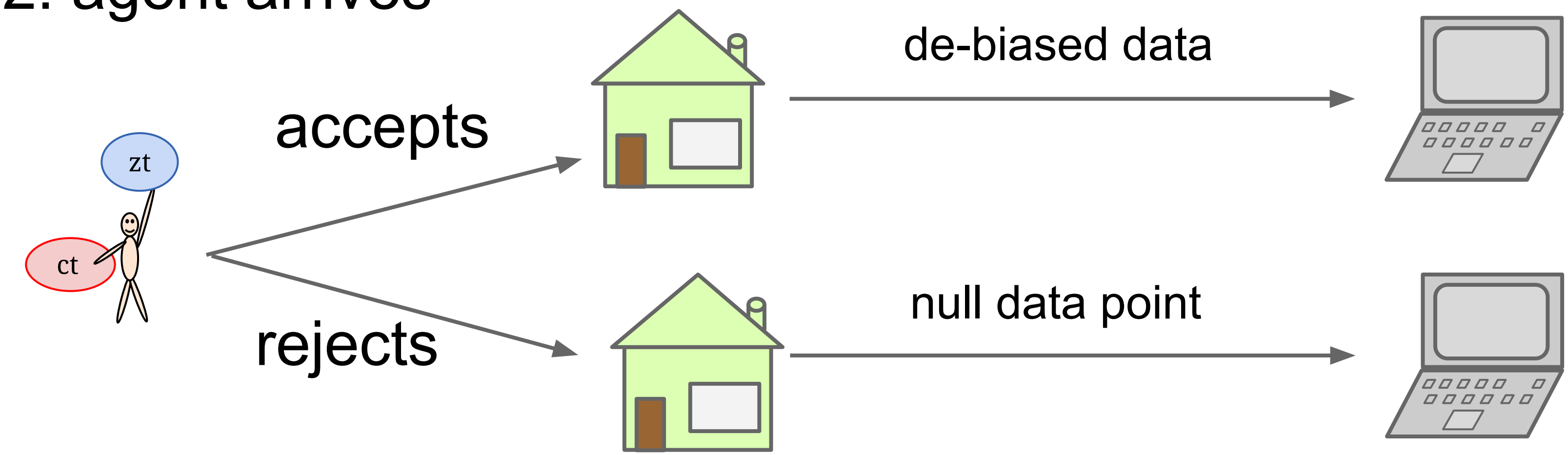
At each time t = 1, …, T:

1. post menu

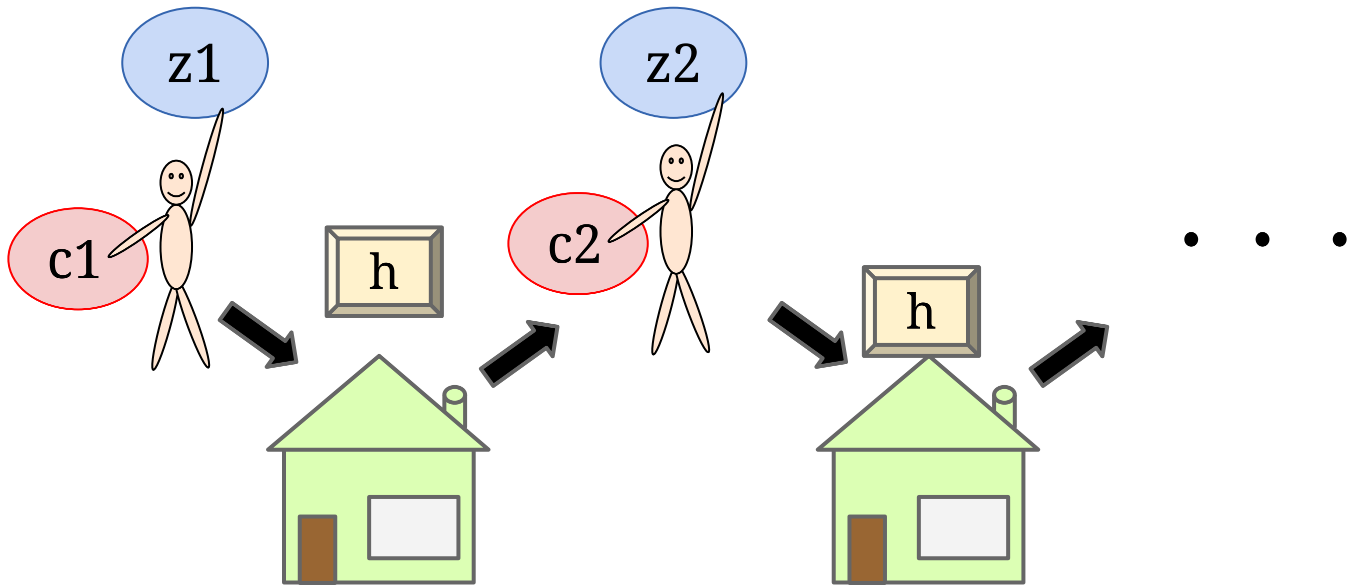
price(z) ~ distribution(ht, z)



2. agent arrives

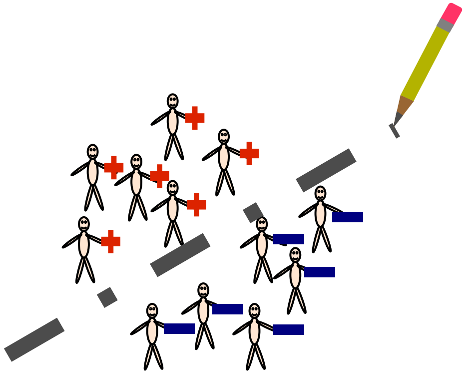


Analysis idea: use no-regret setting!



- Propose **regret minimization with purchased data**
- Prove upper and lower bounds on regret
- low regret ⇒ good prediction on new data (main result)

Summary

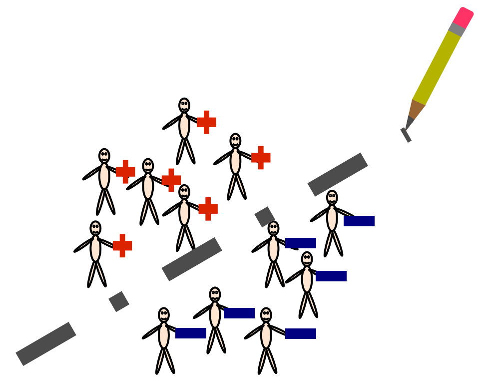


Problem: learn a good hypothesis by buying data from arriving agents

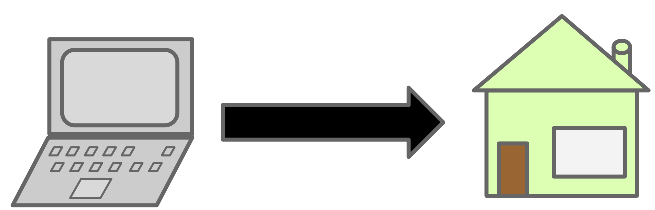
For a variety of learning problems:

$$\mathbf{E}\,\text{loss}(\,\bar{h}\,) \le \mathbf{E}\,\text{loss}(\,h^*\,) + O\left(\sqrt{\frac{\gamma}{B}}\right)$$

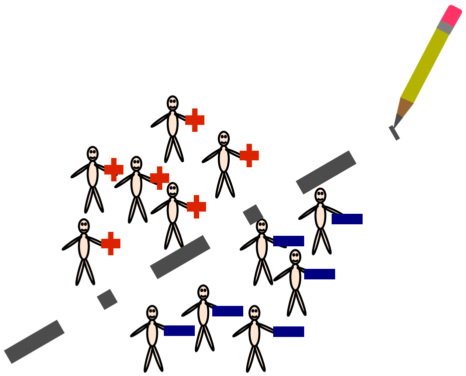
Key ideas



1. price data actively based on value
2. machine-learning style bounds
3. transform learning algs to mechanisms

learning alg  mechanism

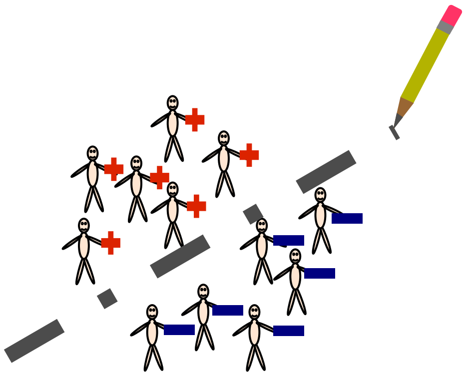
Future work



- Improve bounds (no-regret: gap between lower and upper bounds)
- Propose “universal quantity” to replace γ in bounds (analogue of VC-dimension)
- Variants of the model, better batch mechanisms
- Explore black-box use of learning algs in mechanisms

Future work

- Improve bounds (no-regret: gap between lower and upper bounds)
- Propose "universal quantity" to replace γ in bounds (analogue of VC-dimension)
- Variants of the model, better batch mechanisms
- Explore black-box use of learning algs in mechanisms



Thanks!

Additional slides

What would you do before this work?

Naive 1: post price of 1, obtain B points, run a learner on them.

Naive 2: post lower prices, obtain biased data, do what??

Roth-Schoenebeck (EC 2012): draw prices from a distribution, obtain biased data, de-bias it.

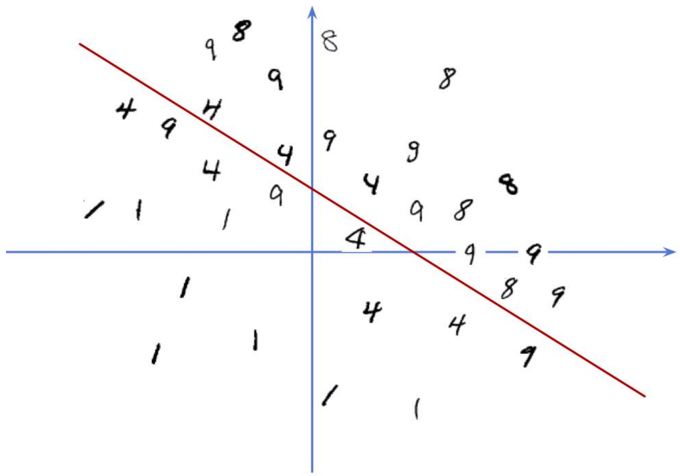
- **Batch setting** (offer each data point the same price distribution)
- Each agent has a number. Task is to **estimate the mean**
- Derives price distribution to **minimize variance** of estimate

Related work

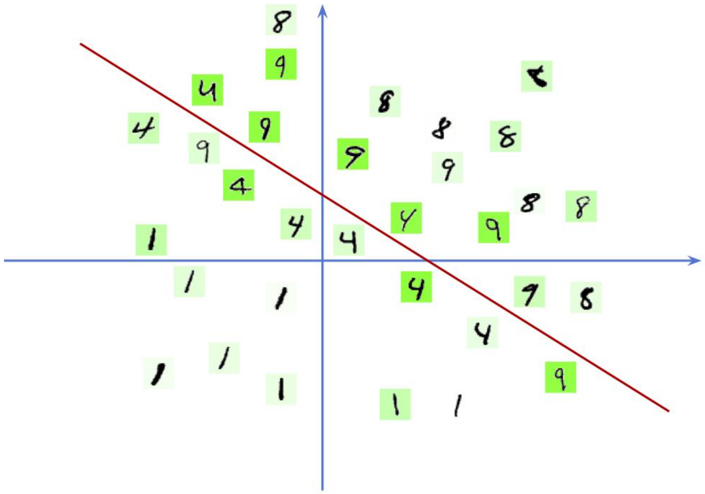
	ML-style risk bounds	**Minimize variance or related goal**
agents cannot fabricate data, have costs	this work	Roth, Schoenebeck 2012 Ligett, Roth 2012 Horel, Ionnadis, Muthukrishnan 2014
principal-agent style, data depends on effort		Cummings, Ligett, Roth, Wu, Ziani 2015 Cai, Daskalakis, Papadimitriou 2015
can fabricate data (like in peer-prediction)	Meir, Procaccia, Rosenschein 2012	Dekel, Fisher, Procaccia 2008 Ghosh, Ligett, Roth, Schoenebeck 2014

Simulation results

MNIST dataset -- handwritten digit classification



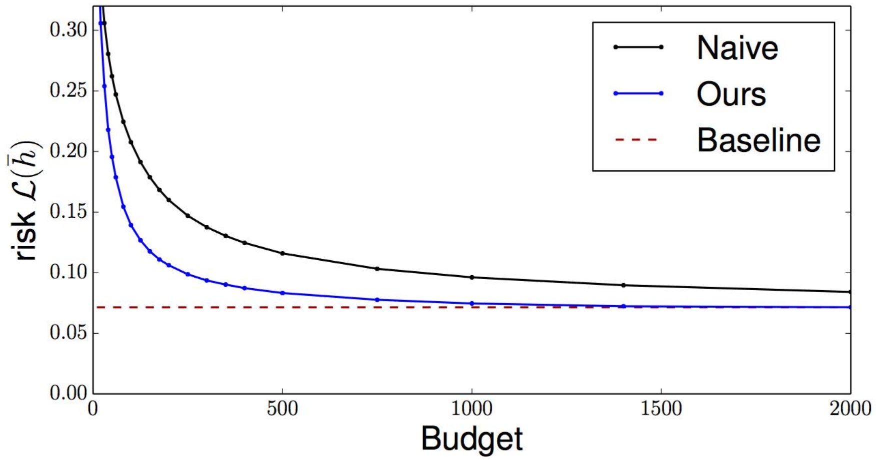
Toy problem: classify (1 or 4) vs (9 or 8)



Brighter green = higher cost

Simulation results

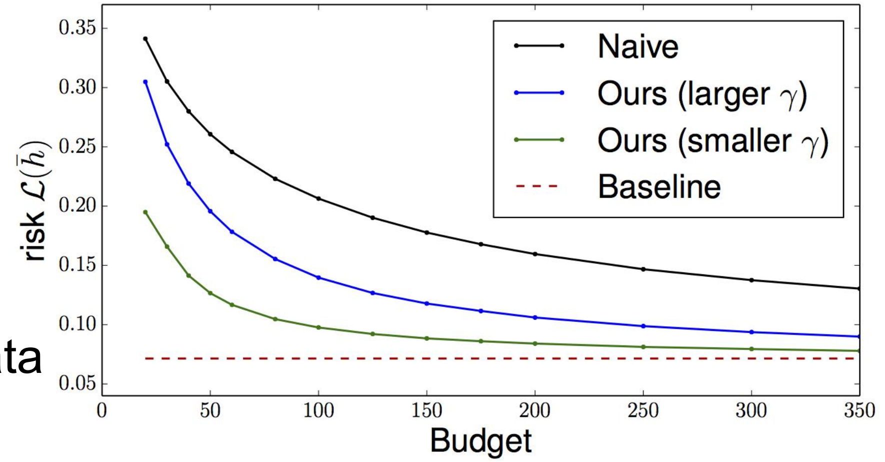
- T = 8503
- train on half,
  test on half
- Alg: Online Gradient Descent



Naive: pay 1 until budget is exhausted, then run alg

Baseline: run alg on all data points (no budget)

Large γ: bad correlations
Small γ: independent cost/data



“value” and pricing distribution?

- Value of data = ~~size of loss~~
  size of **gradient** of loss
  (“how much you learn from the loss”)

- Pricing distribution:

  $$\Pr[\text{ price} \geq x\text{ }] = \frac{\| \nabla \text{ loss}(h_t , z_t) \|}{K \sqrt{x}}$$

- K = normalization constant proportional to $\gamma = \frac{1}{T} \sum_t \| \nabla \text{loss}(h_t,z_t) \| \sqrt{c_t}$
  (assume approximate knowledge of K … in practice, can estimate it online)

- Distribution is derived by optimizing regret bound of mechanism for “at-cost” variant of no-regret setting

Pricing distribution