Truthful Linear Regression

Rachel Cummings, Katrina Ligett

Caltech

Stratis Ioannidis

Northeastern University

$$y_i = \theta^\top x_i + z_i$$

(x_1, y_1)

(x_2, y_2)

$\vdots$

(x_n, y_n)

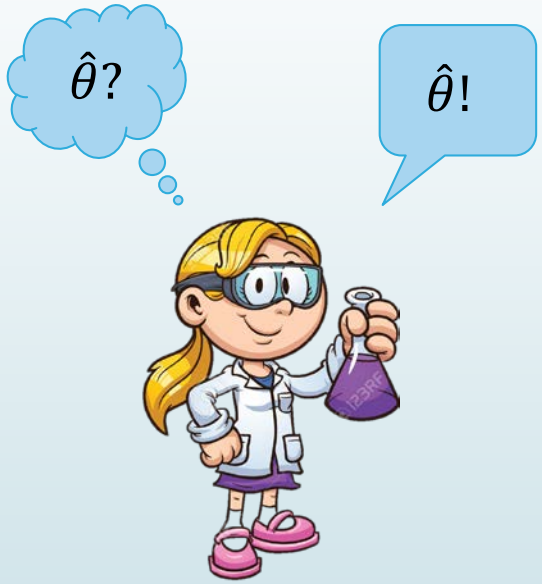
$$y_i = \theta^\top x_i + z_i$$

(x_1, y_1)

(x_2, y_2)

(x_n, y_n)

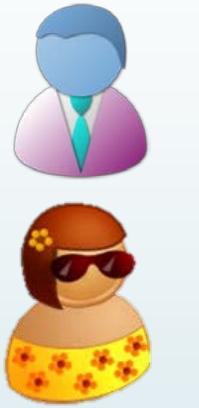
$$y_i = \theta^\top x_i + z_i$$



(x_1, y_1)

(x_2, y_2)

(x_n, y_n)



$$y_i = \theta^\top x_i + z_i$$

(x_1, y_1)

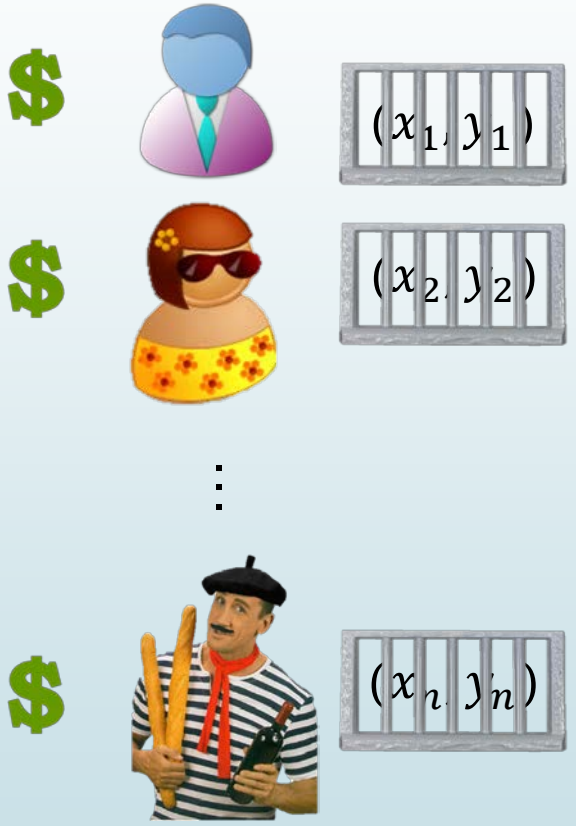
(x_2, y_2)

$\vdots$

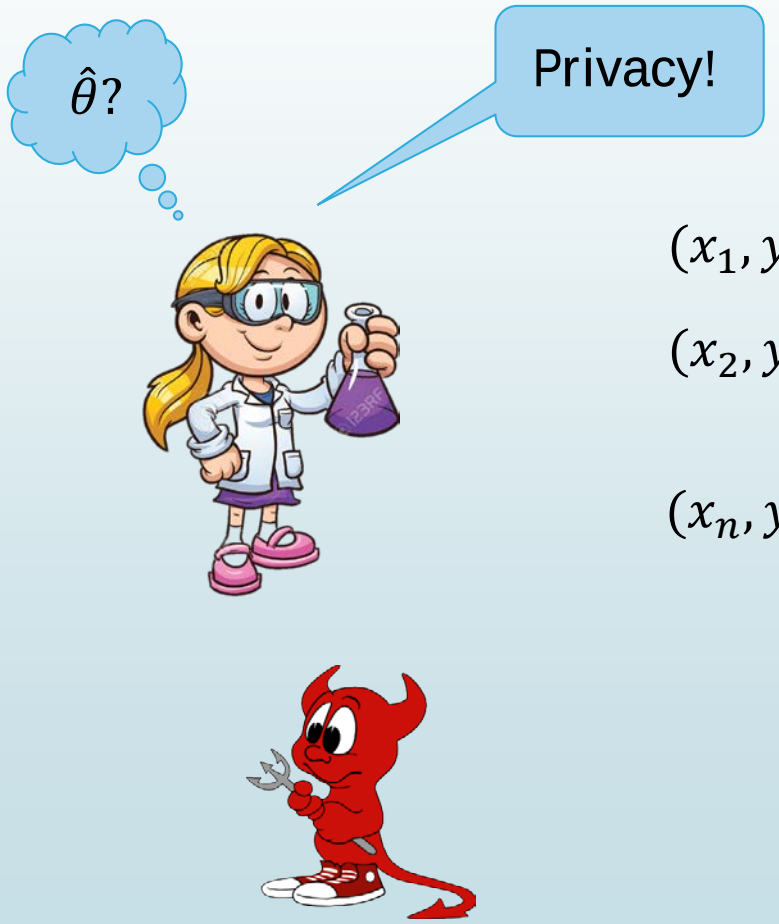
(x_n, y_n)

$$y_i = \theta^\top x_i + z_i$$

$$y_i = \theta^\top x_i + z_i$$



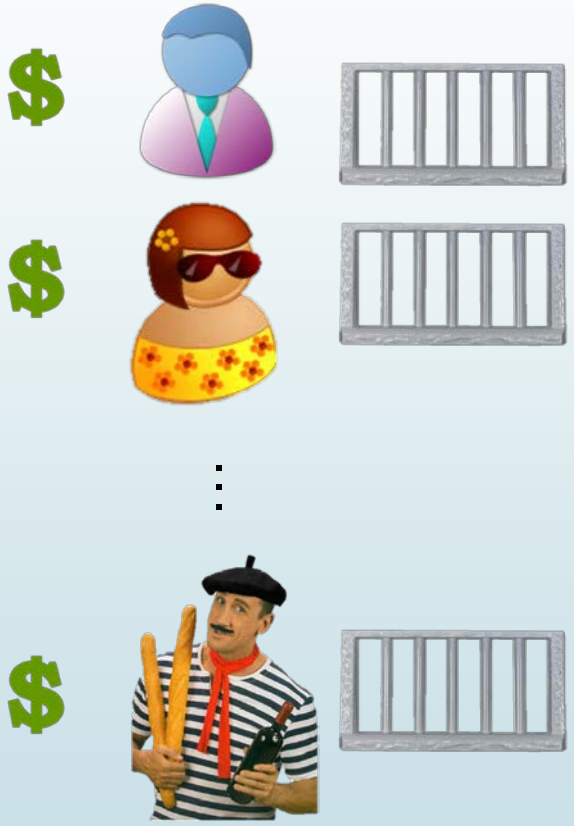
$$y_i = \theta^\top x_i + z_i$$



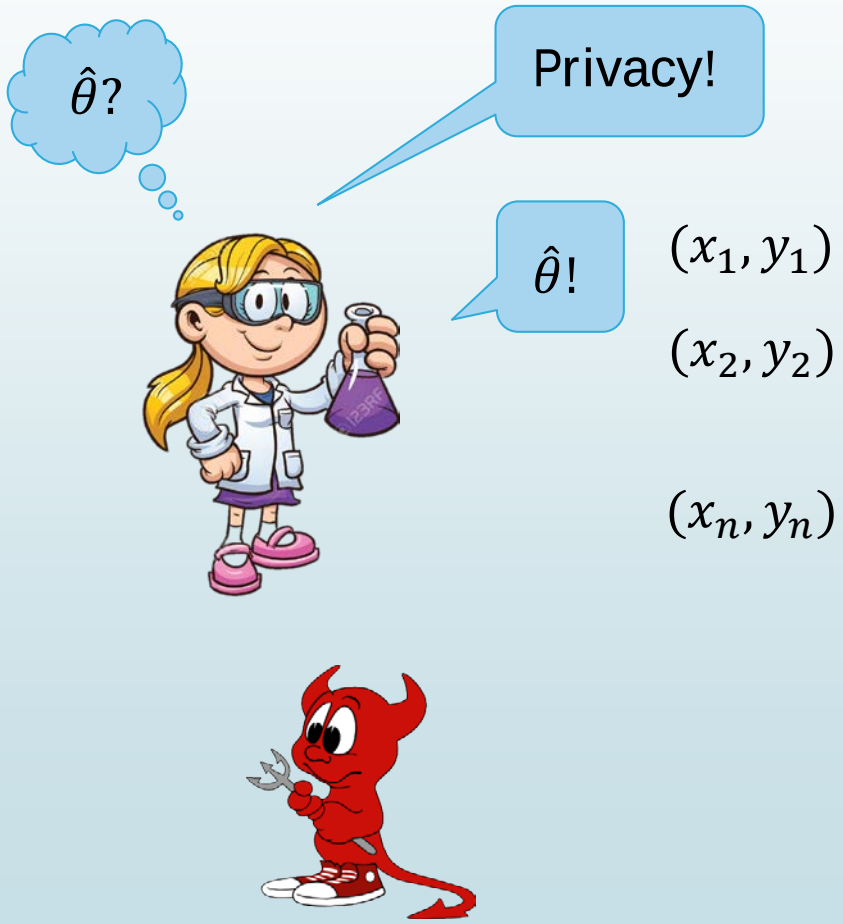
(x_1, y_1)

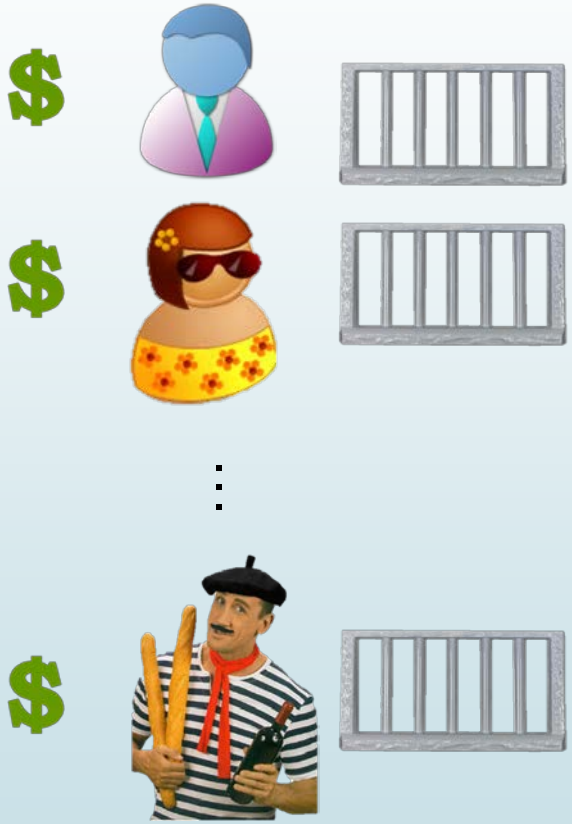
(x_2, y_2)

(x_n, y_n)



$$y_i = \theta^\top x_i + z_i$$





$$y_i = \theta^\top x_i + z_i$$

(x_1, y_1)

(x_2, y_2)

(x_n, y_n)

Challenges:

- Privacy guarantee
- Data is not verifiable

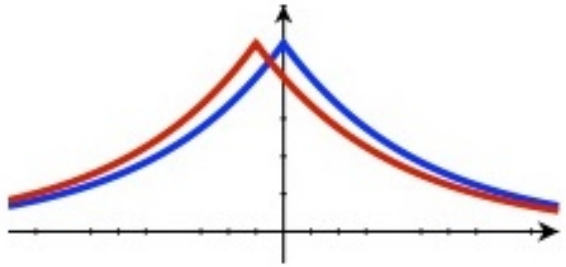
Privacy

- A

$(t_1, \ldots, t_i, \ldots, t_n)$

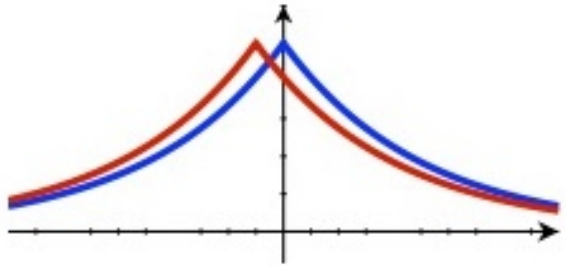
$(t_1, \ldots, t'_i, \ldots, t_n)$

$\Longrightarrow$



-
-
  -
  -

Privacy

- A is biased estimator
- $(t_1, \ldots, t_i, \ldots, t_n)$
  $(t_1, \ldots, t_i', \ldots, t_n)$ $\Longrightarrow$ 
- Getting non-trivial privacy parameter requires regularization
  - Ridge regression
  - θ is biased estimator

Truthfulness

- **Peer prediction** – Ask
-
-
- _____

Truthfulness

- **Peer prediction** – Ask is biased
- Technical challenge: $\boldsymbol{\theta}$ is biased

Biased!

Main result

We design a mechanism that is:

- private
- truthful
- accurately estimates $\hat{\theta}$
- individually rational
- analyst's budget is asymptotically small

Main result

We design a mechanism that is:

- private
- truthful
- accurately estimates $\hat{\theta}$
- individually rational
- analyst's budget is asymptotically small

Truthful Linear Regression

Rachel Cummings, Katrina Ligett, Stratis Ioannidis