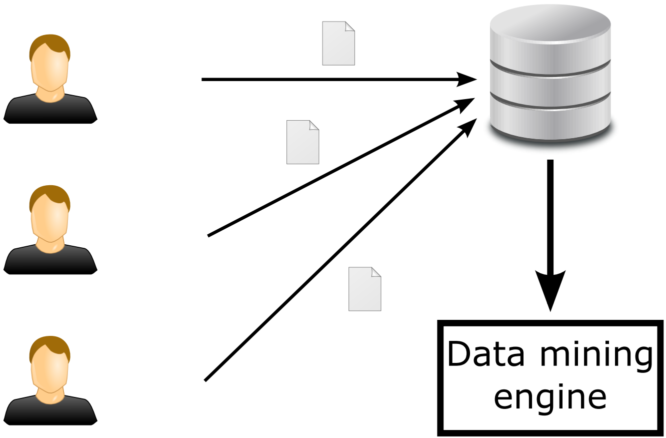
Budget Feasible Mechanisms for Experimental Design

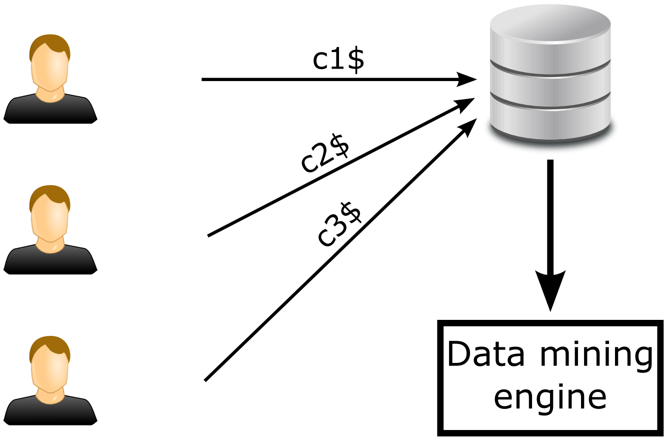
Thibaut Horel
Joint work with Stratis Ioannidis and S. Muthukrishnan

February 26, 2013

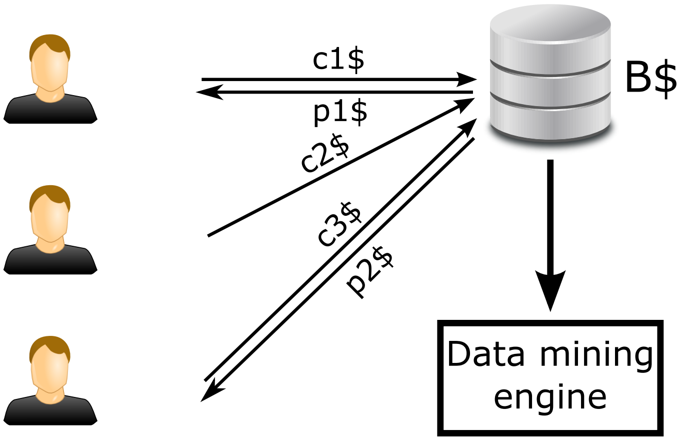
Motivation



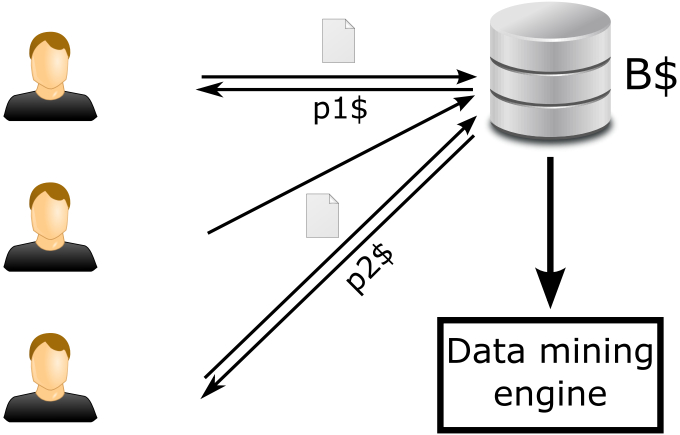
Motivation



Motivation



Motivation



Challenges

- Value of data?
- How to optimize it?
- Strategic users?

Challenges

- Value of data?
- How to optimize it?
- Strategic users?

Challenges

- Value of data?
- How to optimize it?
- Strategic users?

Challenges

- Value of data?
- How to optimize it?
- Strategic users?

Contributions

- case of the linear regression
- deterministic mechanism
- generalization (randomized mechanism)

Contributions

- case of the linear regression
- deterministic mechanism
- generalization (randomized mechanism)

Contributions

- case of the linear regression
- deterministic mechanism
- generalization (randomized mechanism)

Contributions

- case of the linear regression
- deterministic mechanism
- generalization (randomized mechanism)

Outline

Outline

Reverse auction

- set of N sellers: $\mathcal{A} = \{1, \ldots, N\}$; a buyer
- V value function of the buyer, $V : 2^{\mathcal{A}} \rightarrow \mathbb{R}^+$
- $c_i \in \mathbb{R}^+$ price of seller's i good
- B budget constraint of the buyer

Goal

- Find $S \subset \mathcal{A}$ maximizing $V(S)$
- Find payment p_i to seller $i \in S$

Reverse auction

- set of N sellers: $\mathcal{A} = \{1, \dots, N\}$; a buyer
- V value function of the buyer, $V : 2^{\mathcal{A}} \to \mathbb{R}^+$
- $c_i \in \mathbb{R}^+$ price of seller's i good
- B budget constraint of the buyer

Goal

- Find $S \subset \mathcal{A}$ maximizing $V(S)$
- Find payment p_i to seller $i \in S$

Reverse auction

- set of N sellers: $\mathcal{A} = \{1, \ldots, N\}$; a buyer
- V value function of the buyer, $V : 2^{\mathcal{A}} \to \mathbb{R}^+$
- $c_i \in \mathbb{R}^+$ price of seller's i good
- B budget constraint of the buyer

Goal

- Find $S \subset \mathcal{A}$ maximizing $V(S)$
- Find payment p_i to seller $i \in S$

Reverse auction

- set of N sellers: $\mathcal{A} = \{1, \ldots, N\}$; a buyer
- V value function of the buyer, $V : 2^{\mathcal{A}} \to \mathbb{R}^+$
- $c_i \in \mathbb{R}^+$ price of seller's i good
- B budget constraint of the buyer

Goal

- Find $S \subset \mathcal{A}$ maximizing $V(S)$
- Find payment p_i to seller $i \in S$

Reverse auction

- set of N sellers: $\mathcal{A} = \{1, \dots, N\}$; a buyer
- V value function of the buyer, $V : 2^{\mathcal{A}} \rightarrow \mathbb{R}^+$
- $c_i \in \mathbb{R}^+$ price of seller's i good
- B budget constraint of the buyer

Goal

- Find $S \subset \mathcal{A}$ maximizing $V(S)$
- Find payment p_i to seller $i \in S$

Objectives

Payments $(p_i)_{i \in S}$ must be:

- individually rational: $p_i \geq c_i, \; i \in S$
- truthful: reporting one's true cost is a dominant strategy
- budget feasible: $\sum_{i \in S} p_i \leq B$

Mechanism must be:

- computationally efficient: polynomial time
- good approximation: $V(OPT) \leq \alpha V(S)$ with:

$$OPT = \arg\max_{S \subset \mathcal{A}} \left\{ V(S) \mid \sum_{i \in S} c_i \leq B \right\}$$

Objectives

Payments $(p_i)_{i \in S}$ must be:

- individually rational: $p_i \geq c_i, \; i \in S$
- truthful: reporting one's true cost is a dominant strategy
- budget feasible: $\sum_{i \in S} p_i \leq B$

Mechanism must be:

- computationally efficient: polynomial time
- good approximation: $V(OPT) \leq \alpha V(S)$ with:

$$OPT = \arg\max_{S \subset \mathcal{A}} \left\{ V(S) \mid \sum_{i \in S} c_i \leq B \right\}$$

Objectives

Payments $(p_i)_{i \in S}$ must be:

- individually rational: $p_i \geq c_i, \; i \in S$
- truthful: reporting one's true cost is a dominant strategy
- budget feasible: $\sum_{i \in S} p_i \leq B$

Mechanism must be:

- computationally efficient: polynomial time
- good approximation: $V(OPT) \leq \alpha V(S)$ with:

$$OPT = \arg\max_{S \subset \mathcal{A}} \left\{ V(S) \mid \sum_{i \in S} c_i \leq B \right\}$$

Objectives

Payments $(p_i)_{i \in S}$ must be:

- individually rational: $p_i \geq c_i, \; i \in S$
- truthful: reporting one's true cost is a dominant strategy
- budget feasible: $\sum_{i \in S} p_i \leq B$

Mechanism must be:

- computationally efficient: polynomial time
- good approximation: $V(OPT) \leq \alpha V(S)$ with:

$$OPT = \arg\max_{S \subset \mathcal{A}} \left\{ V(S) \mid \sum_{i \in S} c_i \leq B \right\}$$

Objectives

Payments $(p_i)_{i\in S}$ must be:

- individually rational: $p_i \geq c_i,\ i \in S$
- truthful: reporting one's true cost is a dominant strategy
- budget feasible: $\sum_{i\in S} p_i \leq B$

Mechanism must be:

- computationally efficient: polynomial time
- good approximation: $V(OPT) \leq \alpha V(S)$ with:

$$OPT = \arg\max_{S\subset\mathcal{A}} \left\{ V(S) \mid \sum_{i\in S} c_i \leq B \right\}$$

Objectives

Payments $(p_i)_{i \in S}$ must be:

- individually rational: $p_i \geq c_i, \ i \in S$
- truthful: reporting one's true cost is a dominant strategy
- budget feasible: $\sum_{i \in S} p_i \leq B$

Mechanism must be:

- computationally efficient: polynomial time
- good approximation: $V(OPT) \leq \alpha V(S)$ with:

$$OPT = \arg\max_{S \subset \mathcal{A}} \left\{ V(S) \mid \sum_{i \in S} c_i \leq B \right\}$$

Objectives

Payments $(p_i)_{i \in S}$ must be:

- individually rational: $p_i \geq c_i, \; i \in S$
- truthful: reporting one's true cost is a dominant strategy
- budget feasible: $\sum_{i \in S} p_i \leq B$

Mechanism must be:

- computationally efficient: polynomial time
- good approximation: $V(OPT) \leq \alpha V(S)$ with:

$$OPT = \arg\max_{S \subset \mathcal{A}} \left\{ V(S) \mid \sum_{i \in S} c_i \leq B \right\}$$

Known results

When V is submodular:

- randomized budget feasible mechanism, approximation ratio: 7.91 (Chen et al., 2011)
- deterministic mechanisms for:
  - Knapsack: $2 + \sqrt{2}$ (Chen et al., 2011)
  - Matching: 7.37 (Singer, 2010)
  - Coverage: 31 (Singer, 2012)

Known results

When V is submodular:

- randomized budget feasible mechanism, approximation ratio: 7.91 (Chen et al., 2011)

- deterministic mechanisms for:
  - Knapsack: $2 + \sqrt{2}$ (Chen et al., 2011)
  - Matching: 7.37 (Singer, 2010)
  - Coverage: 31 (Singer, 2012)

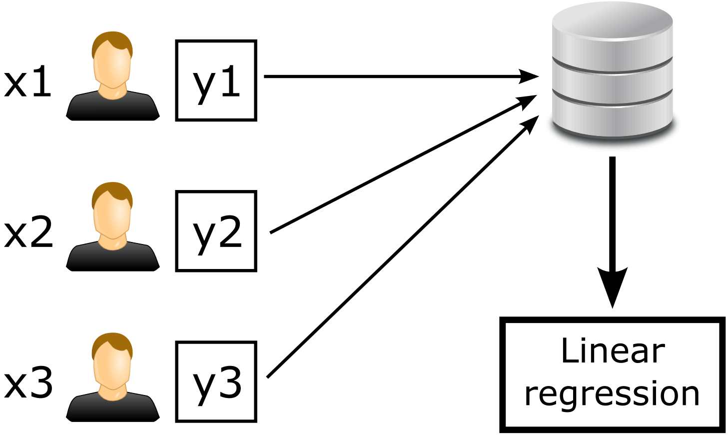
Known results

When V is submodular:

- randomized budget feasible mechanism, approximation ratio: 7.91 (Chen et al., 2011)
- deterministic mechanisms for:
  - Knapsack: $2 + \sqrt{2}$ (Chen et al., 2011)
  - Matching: 7.37 (Singer, 2010)
  - Coverage: 31 (Singer, 2012)

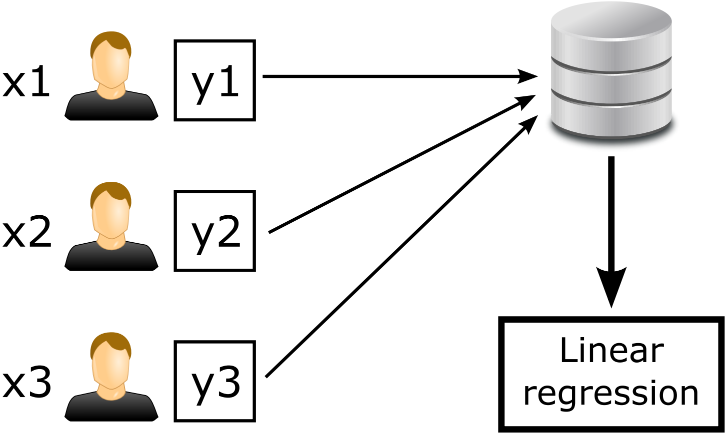
Outline

Linear Regression



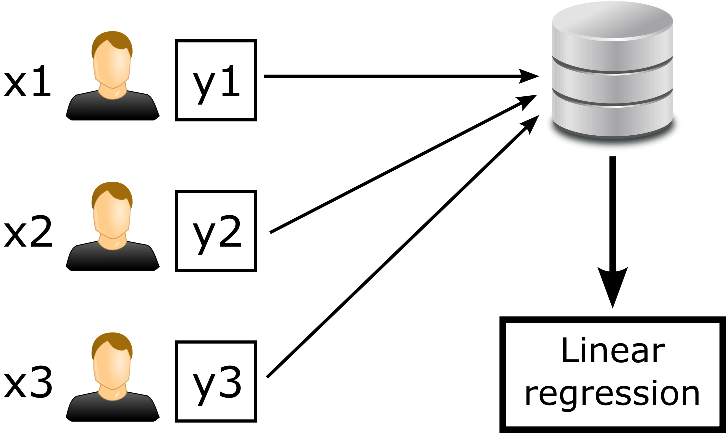
N users

Linear Regression



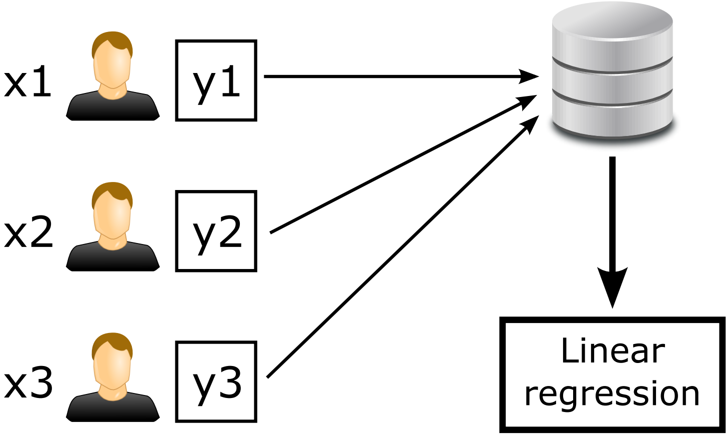
x_i: public features (e.g. age, gender, height, etc.)

Linear Regression



y_i: private data (e.g. disease, etc.)

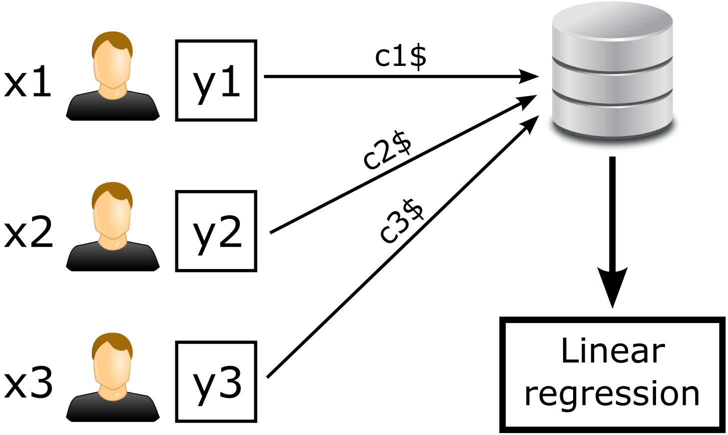
Linear Regression



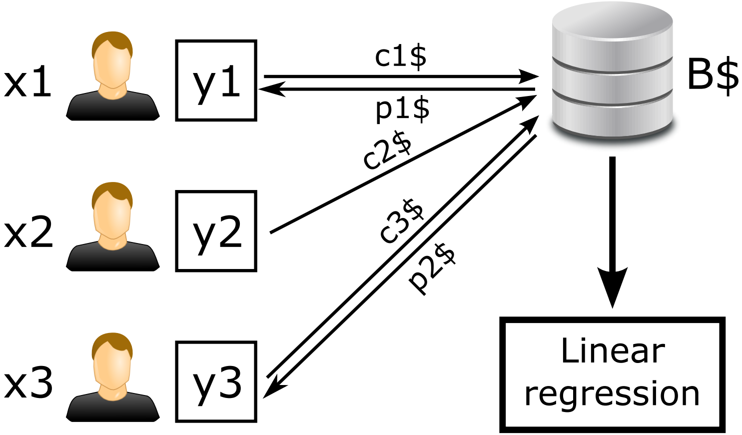
Gaussian Linear model: $y_i = \beta^T x_i + \varepsilon_i$

$$\beta^* = \arg\min_{\beta} \sum_i |y_i - \beta^T x_i|^2$$

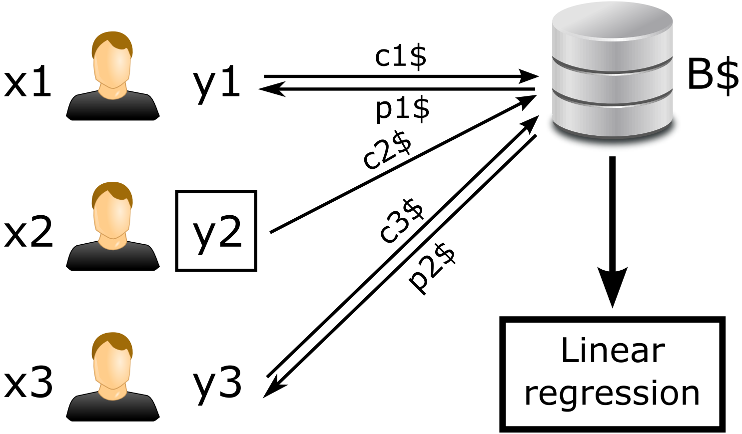
Linear Regression



Linear Regression



Linear Regression



Experimental design

- Public vector of features $x_i \in \mathbb{R}^d$
- Private data $y_i \in \mathbb{R}$

Gaussian linear model:

$$y_i = \beta^T x_i + \varepsilon_i, \quad \beta \in \mathbb{R}^d,\ \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Which users to select? Experimental design $\Rightarrow$ D-optimal criterion

Experimental Design

$$\text{maximize} \quad V(S) = \log\det\left(I_d + \sum_{i \in S} x_i x_i^T\right) \quad \text{subject to} \quad |S| \leq k$$

Experimental design

- Public vector of features $x_i \in \mathbb{R}^d$
- Private data $y_i \in \mathbb{R}$

Gaussian linear model:

$$y_i = \beta^T x_i + \varepsilon_i, \quad \beta \in \mathbb{R}^d,\ \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Which users to select? Experimental design $\Rightarrow$ D-optimal criterion

Experimental Design

$$\text{maximize} \quad V(S) = \log\det\left(I_d + \sum_{i \in S} x_i x_i^T\right) \quad \text{subject to} \quad |S| \le k$$

Experimental design

- Public vector of features $x_i \in \mathbb{R}^d$
- Private data $y_i \in \mathbb{R}$

Gaussian linear model:

$$y_i = \beta^T x_i + \varepsilon_i, \quad \beta \in \mathbb{R}^d,\ \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Which users to select? Experimental design $\Rightarrow$ D-optimal criterion

Experimental Design

$$\text{maximize} \quad V(S) = \log\det\left(I_d + \sum_{i \in S} x_i x_i^T\right) \quad \text{subject to} \quad |S| \leq k$$

Budgeted Experimental design

$$\text{maximize} \quad V(S) = \log\det\left(I_d + \sum_{i\in S} x_i x_i^T\right) \quad \text{subject to} \quad \sum_{i\in S} c_i \leq B$$

- the non-strategic optimization problem is NP-hard
- V is submodular
- previous results give a randomized budget feasible mechanism
- deterministic mechanism?

Budgeted Experimental design

$$\text{maximize} \quad V(S) = \log\det\left(I_d + \sum_{i\in S} x_i x_i^T\right) \quad \text{subject to} \quad \sum_{i\in S} c_i \leq B$$

- the non-strategic optimization problem is NP-hard
- V is submodular
- previous results give a randomized budget feasible mechanism
- deterministic mechanism?

Budgeted Experimental design

$$\text{maximize} \quad V(S) = \log\det\left(I_d + \sum_{i\in S} x_i x_i^T\right) \quad \text{subject to} \quad \sum_{i\in S} c_i \leq B$$

- the non-strategic optimization problem is NP-hard
- V is submodular
- previous results give a randomized budget feasible mechanism
- deterministic mechanism?

Budgeted Experimental design

$$\text{maximize} \quad V(S) = \log\det\left(I_d + \sum_{i\in S} x_i {x_i}^T\right) \quad \text{subject to} \quad \sum_{i\in S} c_i \leq B$$

- the non-strategic optimization problem is NP-hard
- V is submodular
- previous results give a randomized budget feasible mechanism
- deterministic mechanism?

Main result

Theorem

There exists a budget feasible, individually rational and truthful mechanism for budgeted experimental design which runs in polynomial time. Its approximation ratio is:

$$\frac{10e - 3 + \sqrt{64e^2 - 24e + 9}}{2(e-1)} \simeq 12.98$$

Sketch of proof

Mechanism (Chen et. al, 2011) for submodular V

- Find $i^* = \arg\max_i V(\{i\})$
- Compute S_G greedily
- Return:
  - $\{i^*\}$ if $V(\{i^*\}) \geq V(OPT_{-i^*})$
  - S_G otherwise

Valid mechanism, approximation ratio: 8.34

Problem: OPT_{-i^*} is NP-hard to compute

Solution: Replace $V(OPT_{-i^*})$ with L^*:

- computable in polynomial time
- close to $V(OPT_{-i^*})$

- Knapsack (Chen et al., 2011)
- Coverage (Singer, 2012)

Sketch of proof

Mechanism (Chen et. al, 2011) for submodular V

- Find $i^* = \arg\max_i V(\{i\})$
- Compute S_G greedily
- Return:
  - $\{i^*\}$ if $V(\{i^*\}) \geq V(OPT_{-i^*})$
  - S_G otherwise

Valid mechanism, approximation ratio: 8.34

Problem: OPT_{-i^*} is NP-hard to compute

Solution: Replace $V(OPT_{-i^*})$ with L^*:

- computable in polynomial time
- close to $V(OPT_{-i^*})$

- Knapsack (Chen et al., 2011)
- Coverage (Singer, 2012)

Sketch of proof

Mechanism (Chen et. al, 2011) for submodular V

- Find $i^* = \arg\max_i V(\{i\})$
- Compute S_G greedily
- Return:
  - $\{i^*\}$ if $V(\{i^*\}) \geq V(OPT_{-i^*})$
  - S_G otherwise

Valid mechanism, approximation ratio: 8.34

Problem: OPT_{-i^*} is NP-hard to compute

Solution: Replace $V(OPT_{-i^*})$ with L^*:

- computable in polynomial time
- close to $V(OPT_{-i^*})$

- Knapsack (Chen et al., 2011)
- Coverage (Singer, 2012)

Sketch of proof

Mechanism (Chen et. al, 2011) for submodular V

- Find $i^* = \arg\max_i V(\{i\})$
- Compute S_G greedily
- Return:
  - $\{i^*\}$ if $V(\{i^*\}) \geq L^*$
  - S_G otherwise

Valid mechanism, approximation ratio: 8.34

Problem: OPT_{-i^*} is NP-hard to compute

Solution: Replace $V(OPT_{-i^*})$ with L^*:

- computable in polynomial time
- close to $V(OPT_{-i^*})$

- Knapsack (Chen et al., 2011)
- Coverage (Singer, 2012)

Sketch of proof

Mechanism (Chen et. al, 2011) for submodular V

- Find $i^* = \arg\max_i V(\{i\})$
- Compute S_G greedily
- Return:
  - $\{i^*\}$ if $V(\{i^*\}) \geq L^*$
  - S_G otherwise

Valid mechanism, approximation ratio: 8.34

Problem: OPT_{-i^*} is NP-hard to compute

Solution: Replace $V(OPT_{-i^*})$ with L^*:

- computable in polynomial time
- close to $V(OPT_{-i^*})$

- Knapsack (Chen et al., 2011)
- Coverage (Singer, 2012)

Sketch of proof

Mechanism (Chen et. al, 2011) for submodular V

- Find $i^* = \arg\max_i V(\{i\})$
- Compute S_G greedily
- Return:
  - $\{i^*\}$ if $V(\{i^*\}) \geq L^*$
  - S_G otherwise

Valid mechanism, approximation ratio: 8.34

Problem: OPT_{-i^*} is NP-hard to compute

Solution: Replace $V(OPT_{-i^*})$ with L^*:

- computable in polynomial time
- close to $V(OPT_{-i^*})$

- Knapsack (Chen et al., 2011)
- Coverage (Singer, 2012)

Sketch of proof

Mechanism (Chen et. al, 2011) for submodular V

- Find $i^* = \arg\max_i V(\{i\})$
- Compute S_G greedily
- Return:
  - $\{i^*\}$ if $V(\{i^*\}) \geq L^*$
  - S_G otherwise

Valid mechanism, approximation ratio: 8.34

Problem: OPT_{-i^*} is NP-hard to compute

Solution: Replace $V(OPT_{-i^*})$ with L^*:

- computable in polynomial time
- close to $V(OPT_{-i^*})$

- Knapsack (Chen et al., 2011)
- Coverage (Singer, 2012)

Sketch of proof (2)

$$L^* = \arg \max_{\lambda \in [0,1]^n} \left\{ \log\det \left(I_d + \sum_i \lambda_i x_i x_i^T \right) \mid \sum_{i=1}^n \lambda_i c_i \leq B \right\}$$

- polynomial time? convex optimization problem
- close to $V(OPT_{-i^*})$?

Technical lemma

$$L^* \leq 2V(OPT) + V(\{i^*\})$$

Sketch of proof (2)

$$L^* = \arg\max_{\lambda \in [0,1]^n} \left\{ \log\det\left(I_d + \sum_i \lambda_i x_i x_i^T \right) \mid \sum_{i=1}^{n} \lambda_i c_i \leq B \right\}$$

- polynomial time? convex optimization problem
- close to $V(OPT_{-i^*})$?

Technical lemma

$$L^* \leq 2V(OPT) + V(\{i^*\})$$

Sketch of proof (2)

$$L^* = \arg \max_{\lambda \in [0,1]^n} \left\{ \log\det\left(I_d + \sum_i \lambda_i x_i x_i^T \right) \mid \sum_{i=1}^n \lambda_i c_i \leq B \right\}$$

- polynomial time? convex optimization problem
- close to $V(OPT_{-i^*})$?

Technical lemma

$$L^* \leq 2V(OPT) + V(\{i^*\})$$

Sketch of proof (2)

$$L^* = \arg\max_{\lambda \in [0,1]^n} \left\{ \log\det\left(I_d + \sum_i \lambda_i x_i x_i^T \right) \mid \sum_{i=1}^n \lambda_i c_i \leq B \right\}$$

- polynomial time? convex optimization problem
- close to $V(OPT_{-i^*})$?

Technical lemma

$$L^* \leq 2V(OPT) + V(\{i^*\})$$

Outline

Generalization

- Generative model: $y_i = f(x_i) + \varepsilon_i,\ i \in \mathcal{A}$
- prior knowledge of the experimenter: f is a random variable
- uncertainty of the experimenter: entropy $H(f)$
- after observing $\{y_i,\ i \in S\}$, uncertainty: $H(f \mid S)$

Value function: Information gain

$$V(S) = H(f) - H(f \mid S), \quad S \subset \mathcal{A}$$

V is submodular $\Rightarrow$ randomized budget feasible mechanism

Generalization

- Generative model: $y_i = f(x_i) + \varepsilon_i, \; i \in \mathcal{A}$
- prior knowledge of the experimenter: f is a random variable
- uncertainty of the experimenter: entropy $H(f)$
- after observing $\{y_i, \; i \in S\}$, uncertainty: $H(f \mid S)$

Value function: Information gain

$$V(S) = H(f) - H(f \mid S), \quad S \subset \mathcal{A}$$

V is submodular $\Rightarrow$ randomized budget feasible mechanism

Generalization

- Generative model: $y_i = f(x_i) + \varepsilon_i,\ i \in \mathcal{A}$
- prior knowledge of the experimenter: f is a random variable
- uncertainty of the experimenter: entropy $H(f)$
- after observing $\{y_i,\ i \in S\}$, uncertainty: $H(f \mid S)$

Value function: Information gain

$$V(S) = H(f) - H(f \mid S), \quad S \subset \mathcal{A}$$

V is submodular $\Rightarrow$ randomized budget feasible mechanism

Generalization

- Generative model: $y_i = f(x_i) + \varepsilon_i,\ i \in \mathcal{A}$
- prior knowledge of the experimenter: f is a random variable
- uncertainty of the experimenter: entropy $H(f)$
- after observing $\{y_i,\ i \in S\}$, uncertainty: $H(f \mid S)$

Value function: Information gain

$$V(S) = H(f) - H(f \mid S), \quad S \subset \mathcal{A}$$

V is submodular $\Rightarrow$ randomized budget feasible mechanism

Generalization

- Generative model: $y_i = f(x_i) + \varepsilon_i,\ i \in \mathcal{A}$
- prior knowledge of the experimenter: f is a random variable
- uncertainty of the experimenter: entropy $H(f)$
- after observing $\{y_i,\ i \in S\}$, uncertainty: $H(f \mid S)$

Value function: Information gain

$$V(S) = H(f) - H(f \mid S), \quad S \subset \mathcal{A}$$

V is submodular $\Rightarrow$ randomized budget feasible mechanism

Generalization

- Generative model: $y_i = f(x_i) + \varepsilon_i,\ i \in \mathcal{A}$
- prior knowledge of the experimenter: f is a random variable
- uncertainty of the experimenter: entropy $H(f)$
- after observing $\{y_i,\ i \in S\}$, uncertainty: $H(f \mid S)$

Value function: Information gain

$$V(S) = H(f) - H(f \mid S), \quad S \subset \mathcal{A}$$

V is submodular $\Rightarrow$ randomized budget feasible mechanism

Conclusion

- Experimental design + Auction theory = powerful framework
- deterministic mechanism for the general case? other learning tasks?
- approximation ratio $\simeq 13$. Lower bound: 2

Conclusion

- Experimental design + Auction theory = powerful framework
- deterministic mechanism for the general case? other learning tasks?
- approximation ratio $\simeq 13$. Lower bound: 2

Conclusion

- Experimental design + Auction theory = powerful framework
- deterministic mechanism for the general case? other learning tasks?
- approximation ratio $\simeq 13$. Lower bound: 2