Submodular meets Spectral:

Greedy Algorithms for Subset Selections, Sparse Approximation and Dictionary Selection

Abhimanyu Das, David Kempe
ICML 2011

Presenter: Wonryong Ryou

Introduction:

Subset Selection

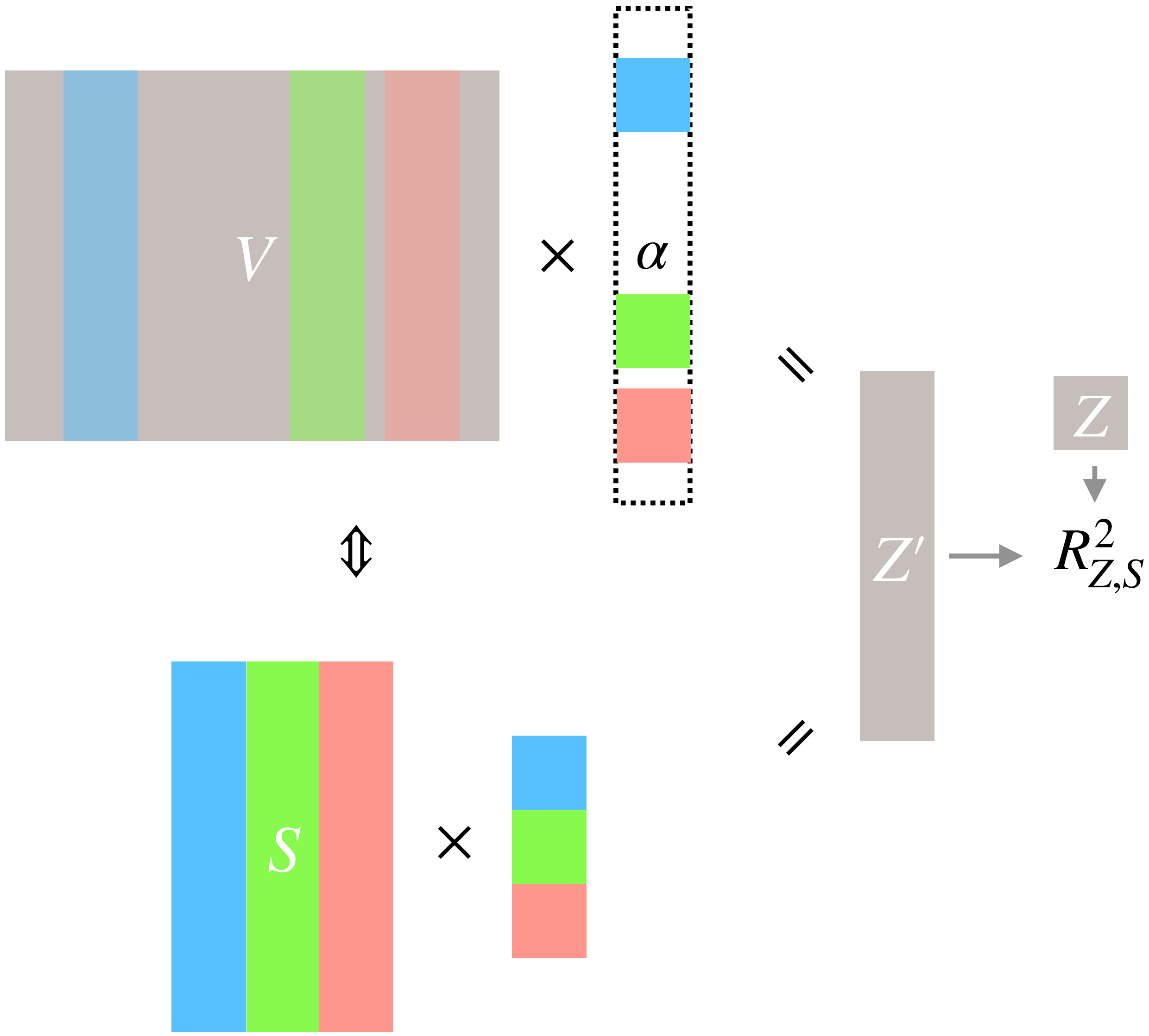
Given: $V = \{X_1, \cdots, X_n\}, k, Z$

Find: $S \subseteq V, s.t. |S| \leq k$

$$Z' = \sum_{i \in S} \alpha_i X_i$$

subject to: $S = \arg\max_S R^2_{Z,S}$

where $R^2_{Z,S} = \frac{\mathbb{V}Z - \mathbb{E}[(Z - Z')^2]}{\mathbb{V}Z}$



Introduction:

Dictionary Selection

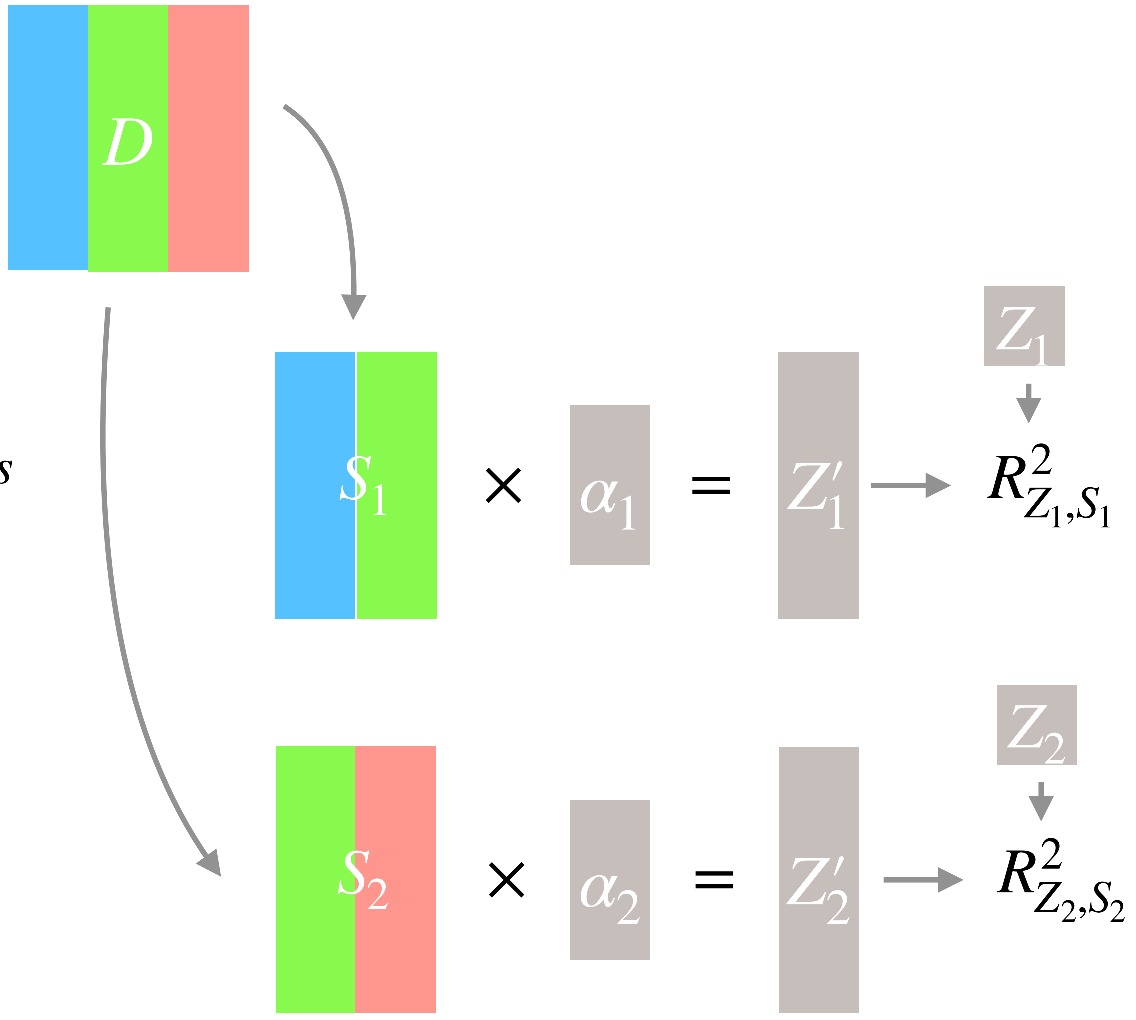
(Krause and Cevher 2010)

Given: $V = \{X_1, \cdots, X_n\}, k, d, Z_1, \cdots, Z_s$

Find: $D \subseteq V, s.t. |D| \leq d$

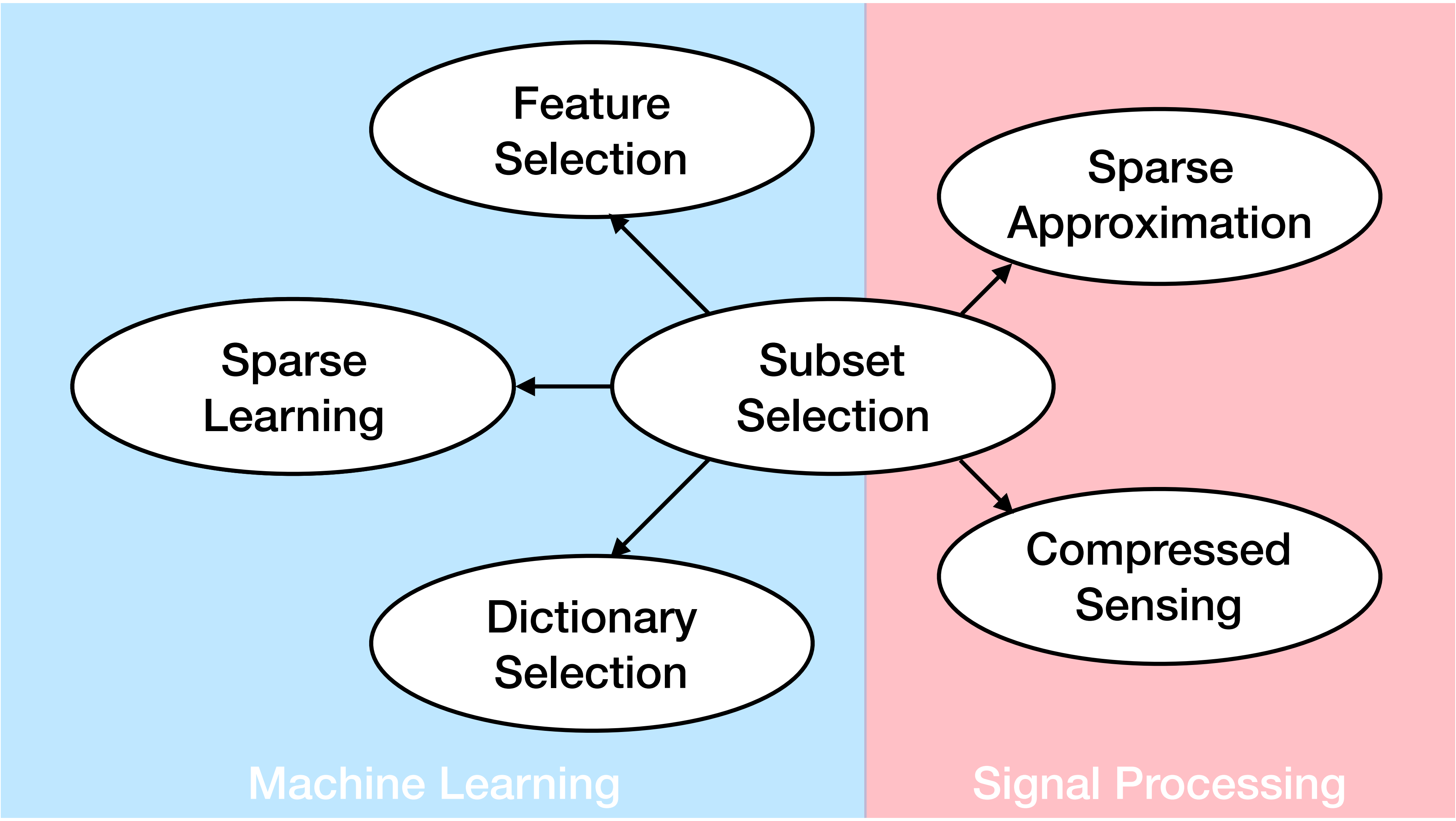
subject to: $D = \arg\max_D F(D)$

$$= \arg\max_D \sum_j \max_{S \subset D, |S|=k} R^2_{Z_j,S}$$



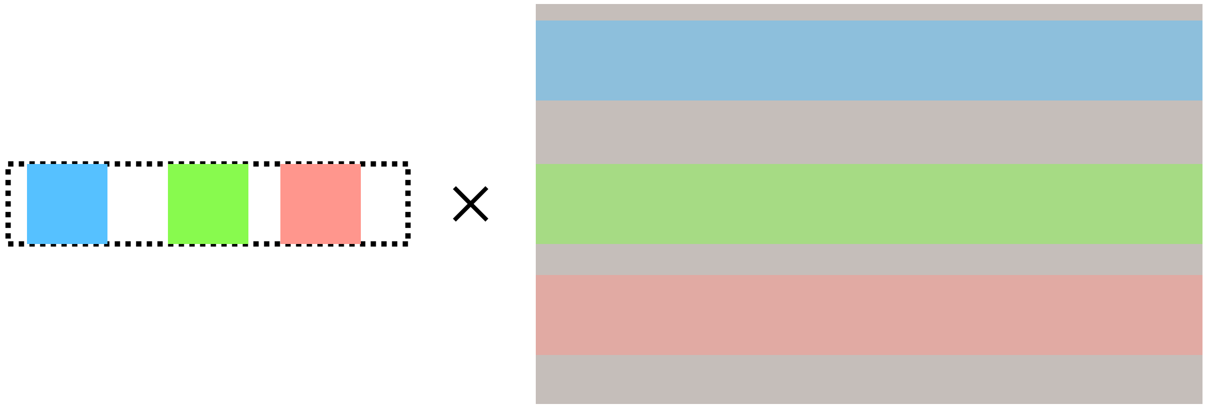
Introduction:

and More…

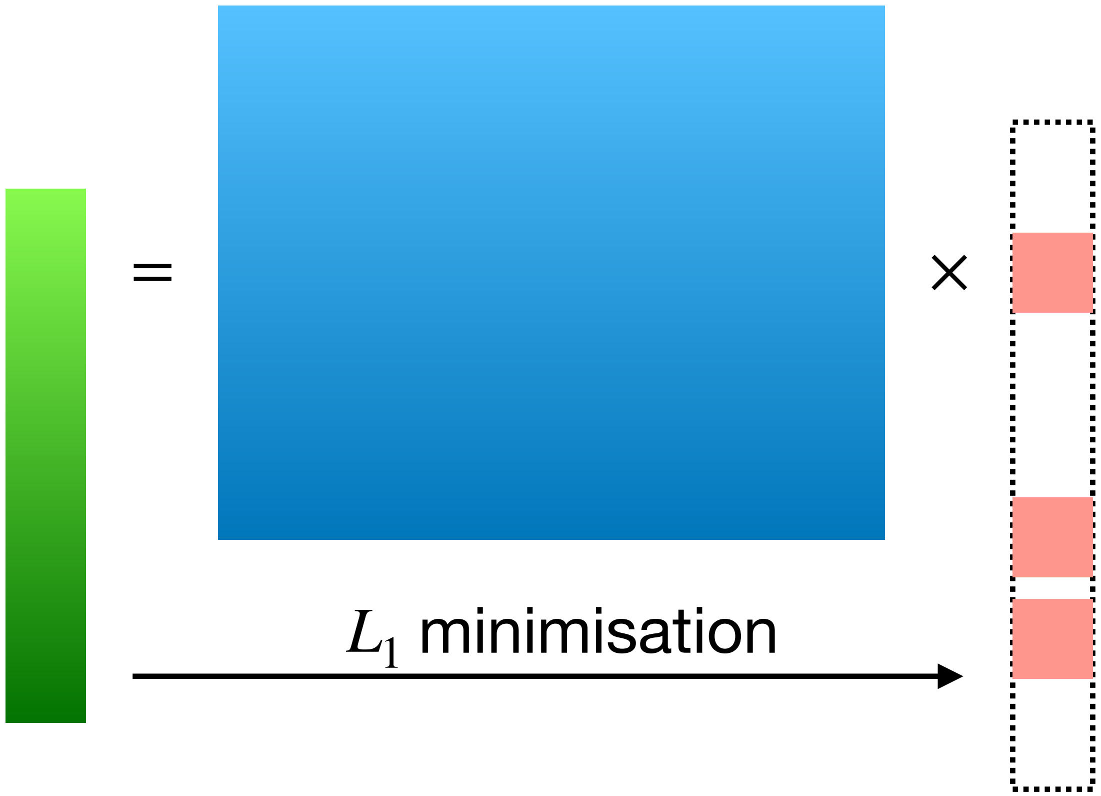


Introduction:

and More…



Feature selection



Compressed sensing

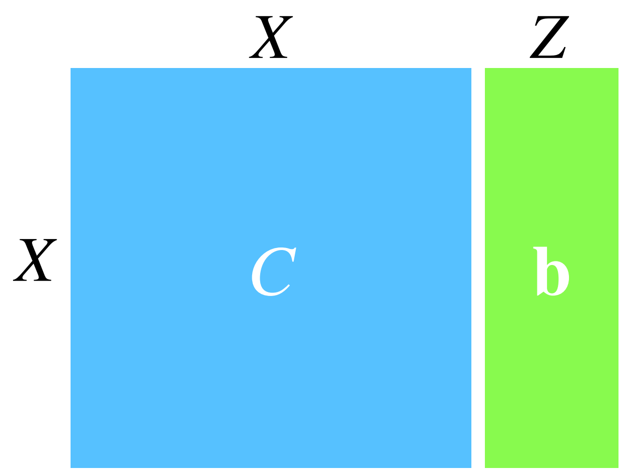
Introduction:

Those problems are

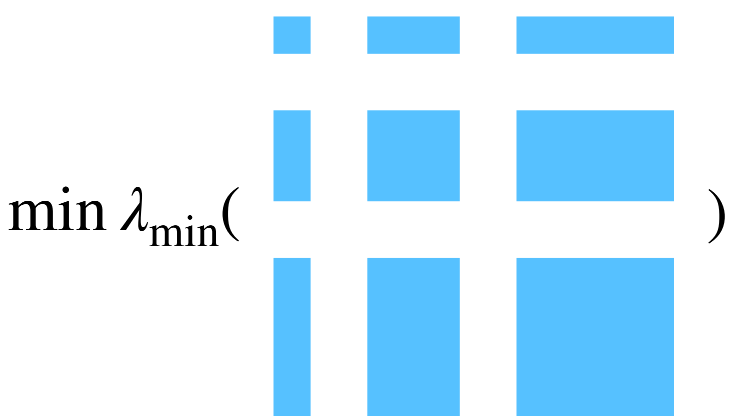
NP-hard

“Getting better theoretical bounds for subset selections, with the novel concept of ***submodularity ratio.***”

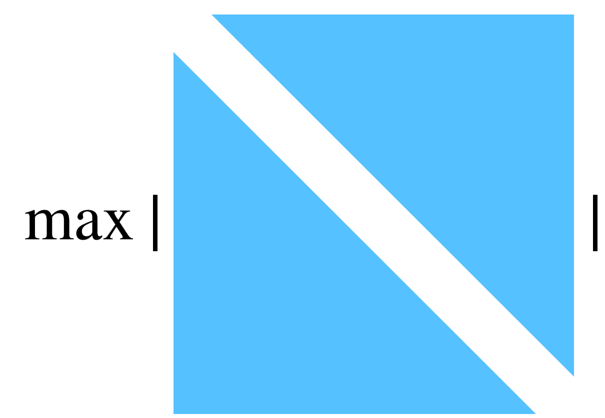
Preliminary:
Notations & Definitions



Covariance matrices $C, \mathbf{b}$



Smallest k**-sparse eigenvalue** $\lambda_{\min}(C, k)$



Coherence $\mu(C)$

Preliminary:

Prior Works

- Sparse recovery
  - L1 relaxation: near-optimal under certain conditions.
    (Tropp 2006, E.J.Candès et al. 2005)

- Subset selection
  - Greedy: $(1 - \Theta(\mu k)) \cdot OPT$ for R^2 under $\mu = O(1/k)$,
    $(1 + \Theta(\mu^2 k)) \cdot OPT$ for MSE.
    (Das and Kempe 2008, Gilbert et al. 2003, Tropp et al. 2003, 2004)

- Dictionary selection
  - Greedy: additive approximation guarantee.
    (Krause and Cevher 2010)

Main Idea:

Submodular Function

$f : 2^{\Omega} \to \mathbb{R}$ is submodular if

$$\forall A, B \subseteq \Omega,\ A \subseteq B,\ \forall s \notin B$$

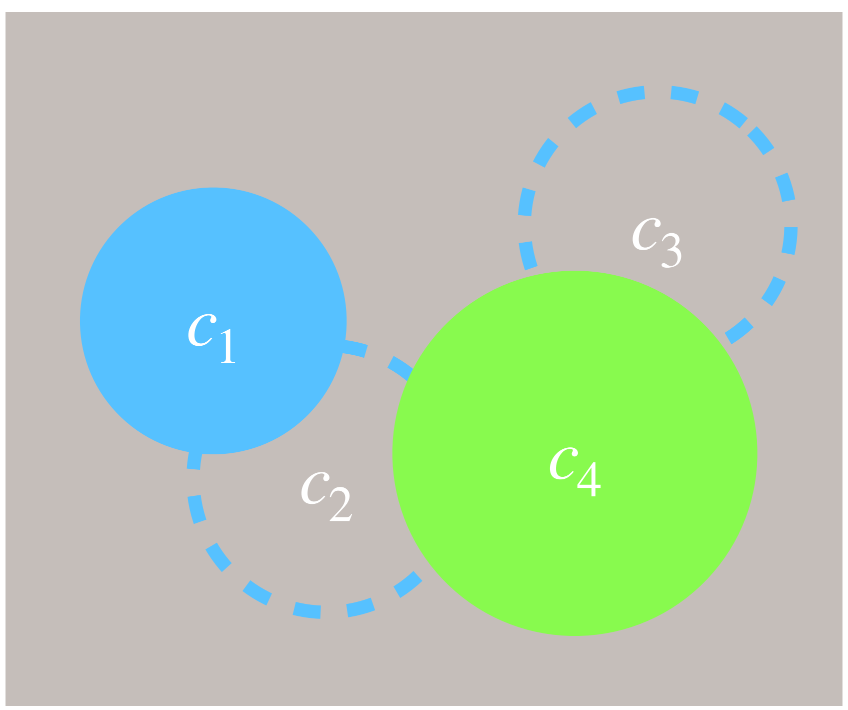
$$f(A \cup \{s\}) - f(A) \geq f(B \cup \{s\}) - f(B)$$

Main Idea:
Submodular Function

$f : 2^{\Omega} \rightarrow \mathbb{R}$ is submodular if

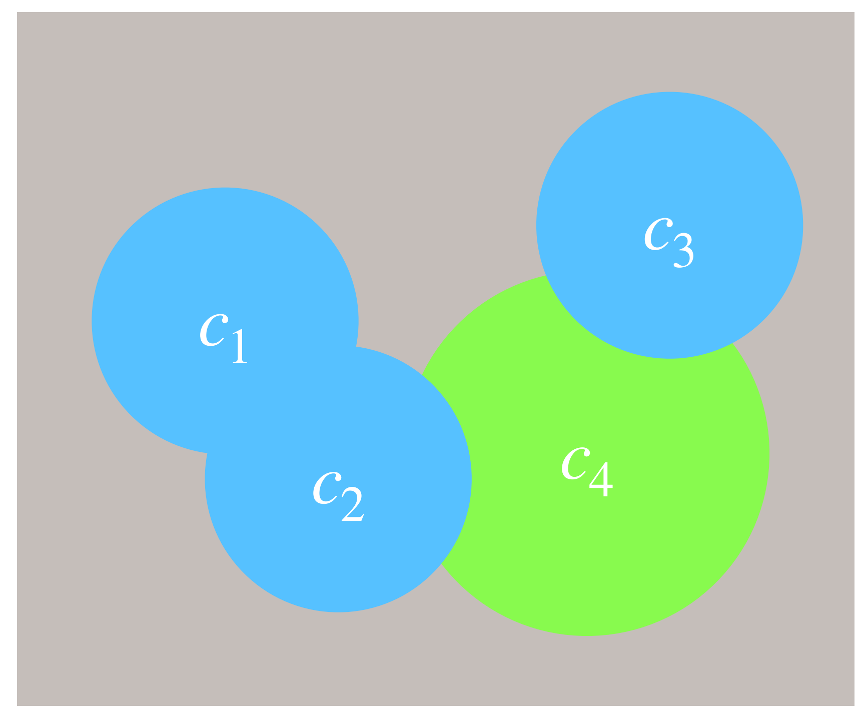
$$\forall A, B \subseteq \Omega,\ A \subseteq B,\ \forall s \notin B$$

$$f(A \cup \{s\}) - f(A) \geq f(B \cup \{s\}) - f(B)$$



$f(\{c_1, c_4\}) - f(\{c_1\})$

$\geq$



$f(\{c_1, c_2, c_3, c_4\}) - f(\{c_1, c_2, c_3\})$

Main Idea:
Submodular Function

$f : 2^{\Omega} \rightarrow \mathbb{R}$ is submodular if

$$\forall A, B \subseteq \Omega,\ A \subseteq B,\ \forall s \notin B$$

$$f(A \cup \{s\}) - f(A) \geq f(B \cup \{s\}) - f(B)$$

Greedy algorithm performs $\left(1 - \frac{1}{e}\right) \cdot OPT$

(Nemhauser et al. 1978)

Main Idea:

Submodularity Ratio

$$\gamma_{U,k}(f) = \min_{L \subseteq U, S : |S| \leq k, S \cap L = \emptyset} \frac{\sum_{x \in S} \left(f(L \cup \{x\}) - f(L) \right)}{f(L \cup S) - f(L)}$$

Main Idea:
Submodularity Ratio

Lemma 2.4

$$\gamma_{U,k}(f) \geq \lambda_{\min}(C, k + |U|) \geq \lambda_{\min}(C)$$

Main Idea:

Submodularity Ratio

$$\gamma_{U,k}(f) = \min_{L \subseteq U, S : |S| \leq k, S \cap L = \emptyset} \frac{\sum_{x \in S} \left(f(L \cup \{x\}) - f(L) \right)}{f(L \cup S) - f(L)}$$

$\gamma_{U,k} \geq 1$

f is submodular

$\gamma_{U,k} < 1$

Still works well!

Analysis:

Algorithms

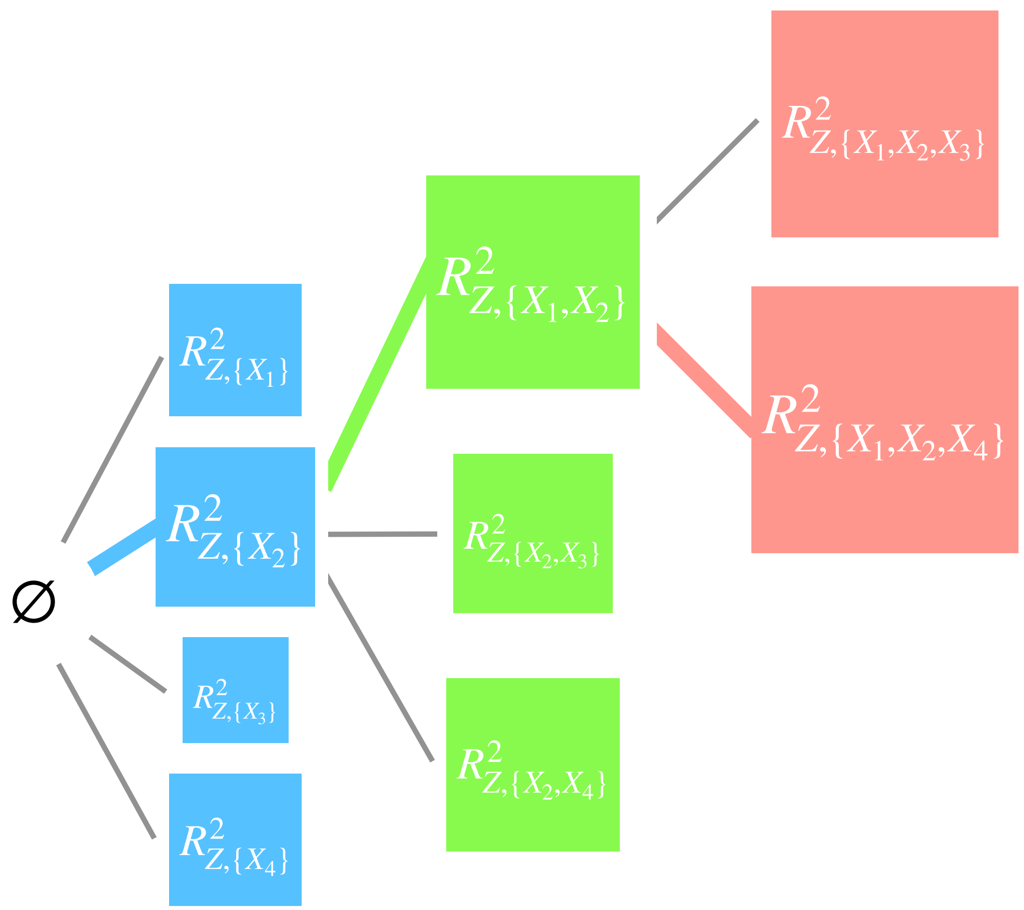
Forward Regression (FR)
Orthogonal Matching Pursuit (OMP)
Oblivious Algorithm (OBL)

Subset Selection

Submodular Dictionary Selection -
Modular Approximation (SDS_{MA})
Orthogonal Matching Pursuit (SDS_{OMP})

Dictionary Selection

Analysis - Subset Selection:
Forward Regression



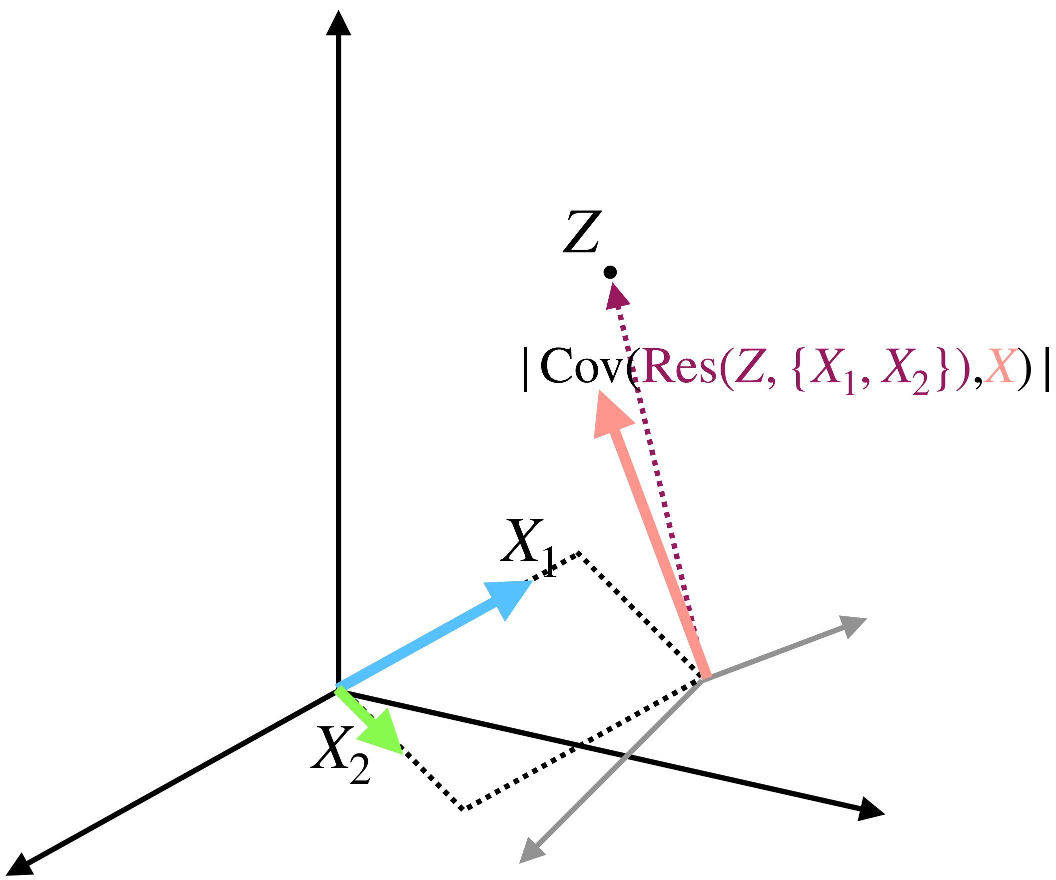
$$R^2_{Z,S^{FR}} \geq (1 - e^{-\gamma_{S^{FR},k}}) \cdot OPT$$

$$\geq (1 - e^{-\lambda_{\min}(C,2k)}) \cdot OPT$$

$$\geq (1 - e^{-\lambda_{\min}(C,k)}) \cdot \Theta\left((\frac{1}{2})^{1/\lambda_{\min}(C,k)} \right) \cdot OPT$$

Analysis - Subset Selection:

Orthogonal Matching Pursuit



$$R^2_{Z,S^{OMP}} \geq (1 - e^{-(\gamma_{S^{OMP},k} \cdot \lambda_{\min}(C,2k))}) \cdot OPT$$

$$\geq (1 - e^{-\lambda_{\min}(C,2k)^2}) \cdot OPT$$

$$\geq (1 - e^{-\lambda_{\min}(C,k)^2}) \cdot \Theta\left((\frac{1}{2})^{1/\lambda_{\min}(C,k)} \right) \cdot OPT$$

Analysis - Subset Selection:

Oblivious Algorithm

$\mathrm{Cov}(Z, X_1)$

$\mathrm{Cov}(Z, X_2)$

$\mathrm{Cov}(Z, X_3)$

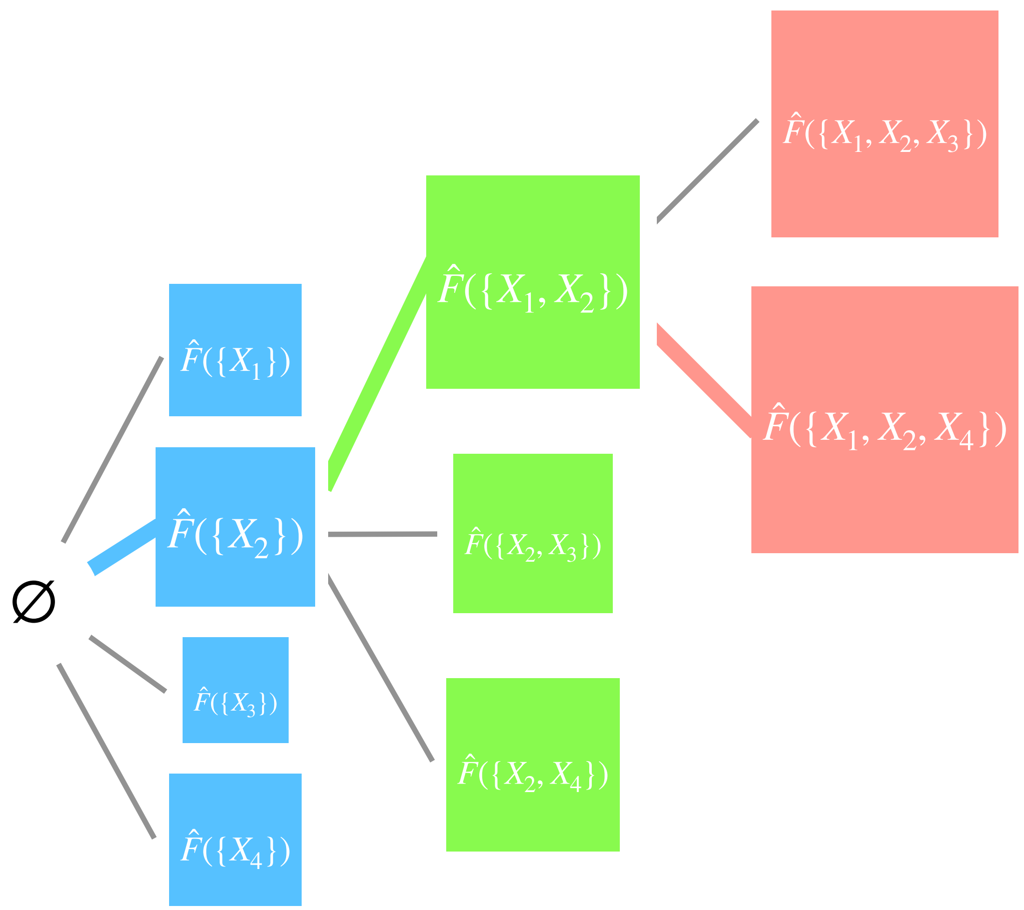
$\mathrm{Cov}(Z, X_4)$

$\mathrm{Cov}(Z, X_5)$

$$R^2_{Z,S^{OBL}} \geq \frac{\gamma_{\emptyset,k}}{\lambda_{\max}(C,k)} \cdot OPT$$

$$\geq \frac{\lambda_{\min}(C,k)}{\lambda_{\max}(C,k)} \cdot OPT$$

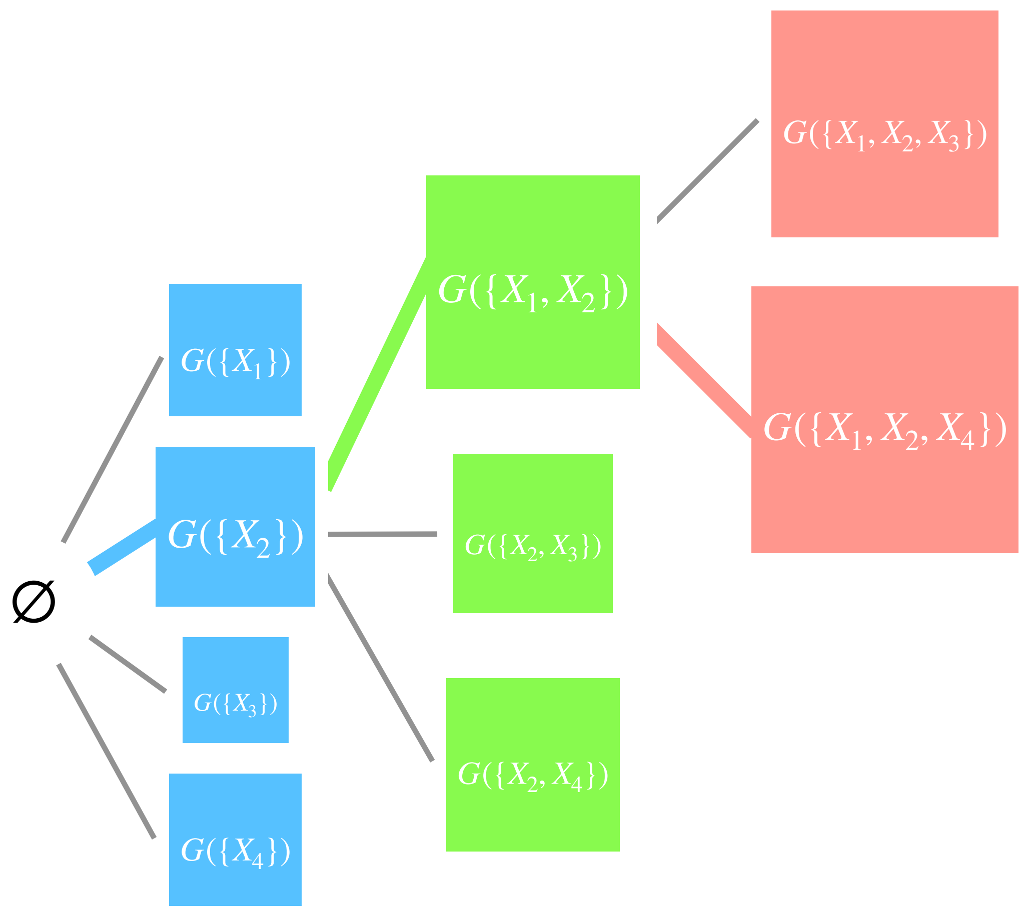
Analysis - Dictionary Selection: SDS$_{MA}$



$$\hat{F}(D) = \sum_{j=1}^{s} \max_{S \subset D, |S|=k} \left(\sum_{i \in S} R^2_{Z_j, X_i} \right)$$

$$\begin{aligned} F(D^{MA}) &\geq \frac{\gamma_{\emptyset,k}}{\lambda_{\max}(C,k)} \left(1 - \frac{1}{e}\right) F(D^{OPT}) \\ &\geq \frac{\lambda_{\min}(C,k)}{\lambda_{\max}(C,k)} \left(1 - \frac{1}{e}\right) F(D^{OPT}) \\ &\geq \left(1 - \frac{1}{e}\right) F(D^{OPT}) - \left(2 - \frac{1}{e}\right) k \cdot \mu(C) \end{aligned}$$

Analysis - Dictionary Selection: SDS$_{OMP}$



$$G(D) = \sum_{j=1}^{s} R^2_{Z_j, OMP(D,Z,k)}$$

$$F(D^{OMP}) \geq \frac{\gamma_{\varnothing,k}}{\lambda_{\max}(C,k)} \cdot \frac{1 - e^{-(p \cdot \gamma_{\varnothing,k})}}{d - d \cdot p \cdot \gamma_{\varnothing,k} + 1} \cdot F(D^{OPT})$$

$$\geq \left(1 - \frac{1}{e}\right) F(D^{OPT}) - k\left(6n + 2 - \frac{1}{e}\right)\mu(C)$$

$$where\ p = \frac{1 - e^{-\lambda_{\min}(C,2k)^2}}{\lambda_{\max}(C,k)}$$

Experiments - Subset Selection:
Setup

Exhaustive Search (OPT)
Forward Regression (FR)
Orthogonal Matching Pursuit (OMP)
L1-Regularisation/Lasso (L1)
Oblivious Algorithm (OBL)

Algorithms

Boston Housing Data
World Development Indicators
Synthetic Data

Datasets

Experiments - Subset Selection:

Results: Boston Housing

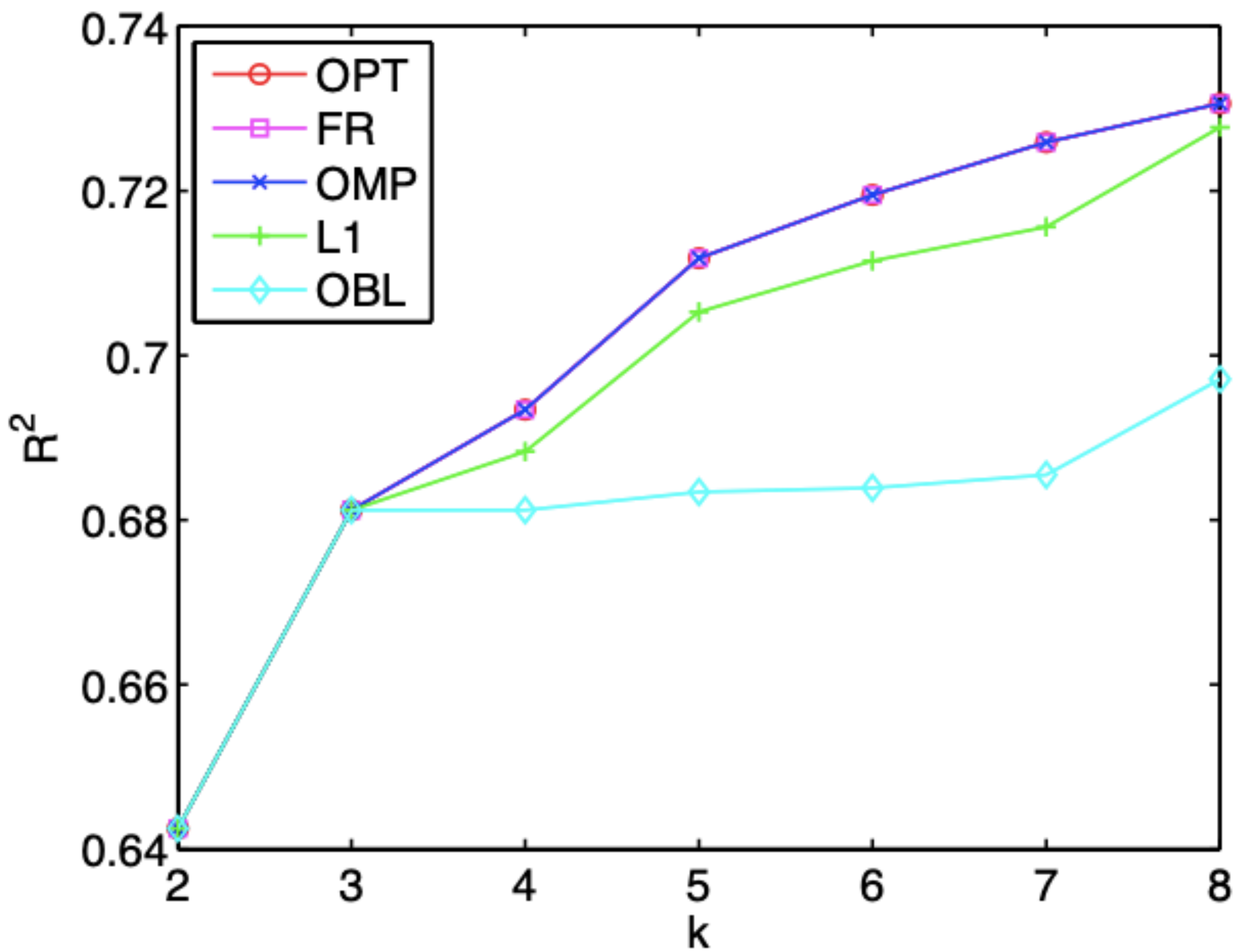


Figure 1: Boston Housing R^2

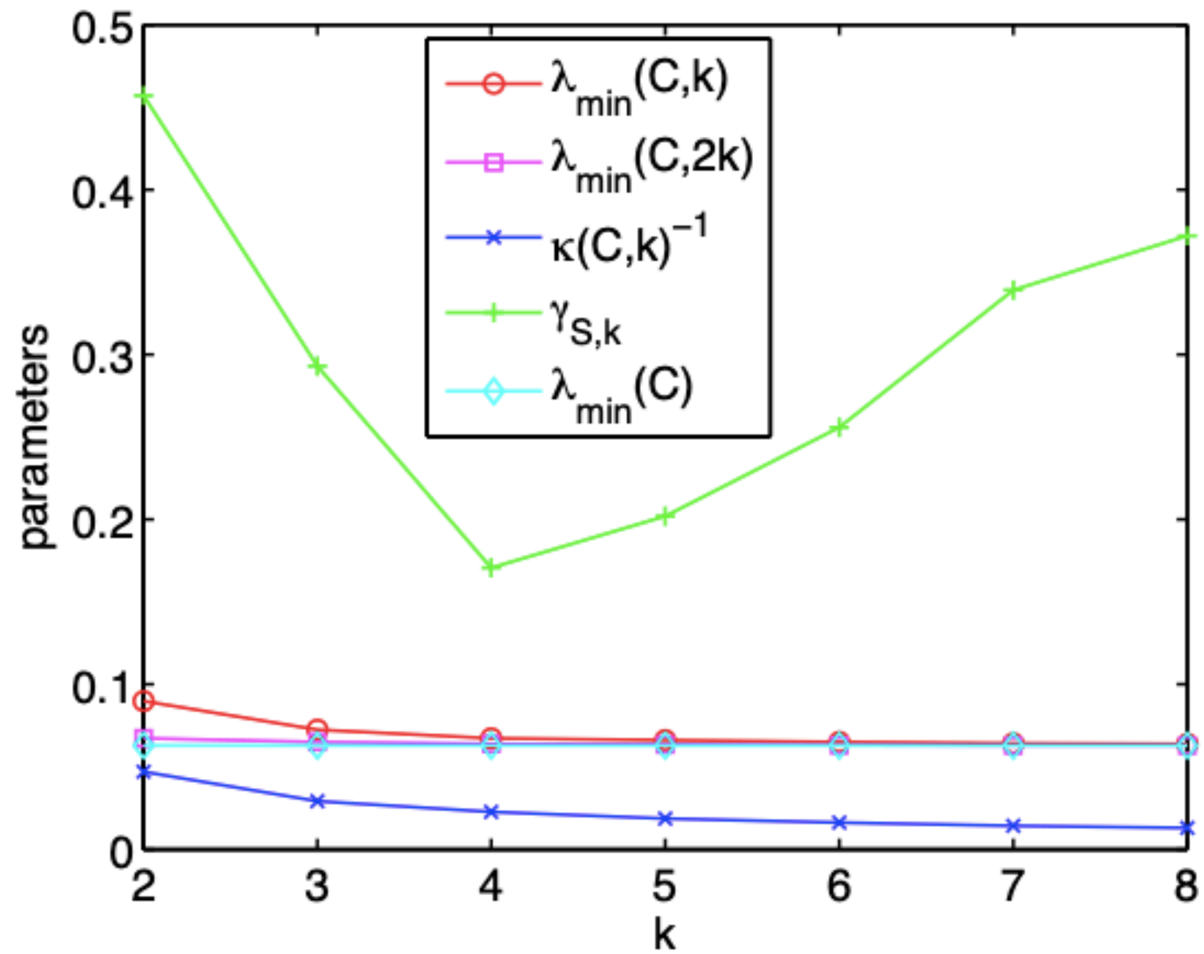


Figure 2: Boston Housing parameters

Experiments - Subset Selection:

Results: World Development Idc

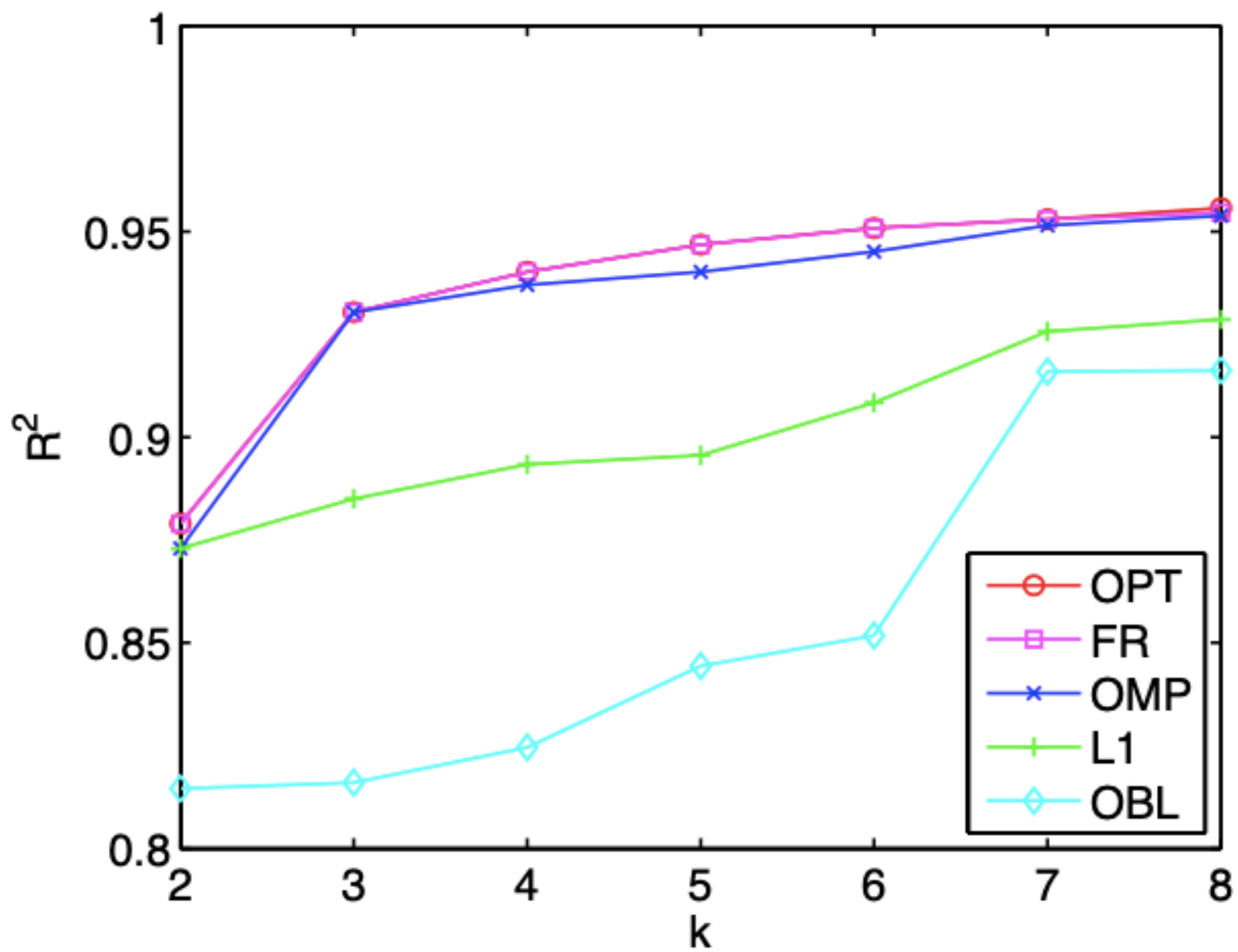


Figure 3: World Bank R^2

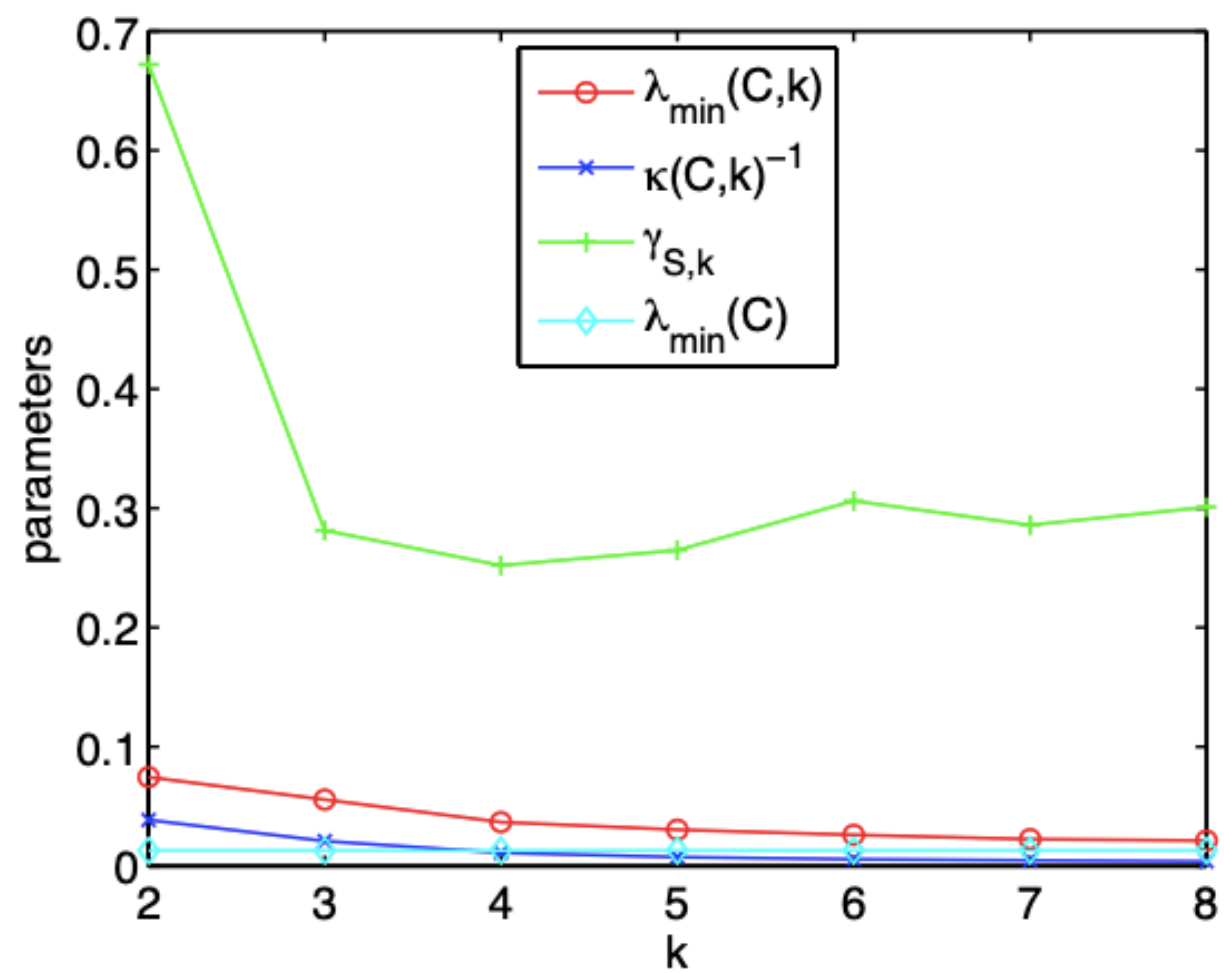


Figure 4: World Bank parameters

Experiments - Subset Selection:

Results: Synthetic Data

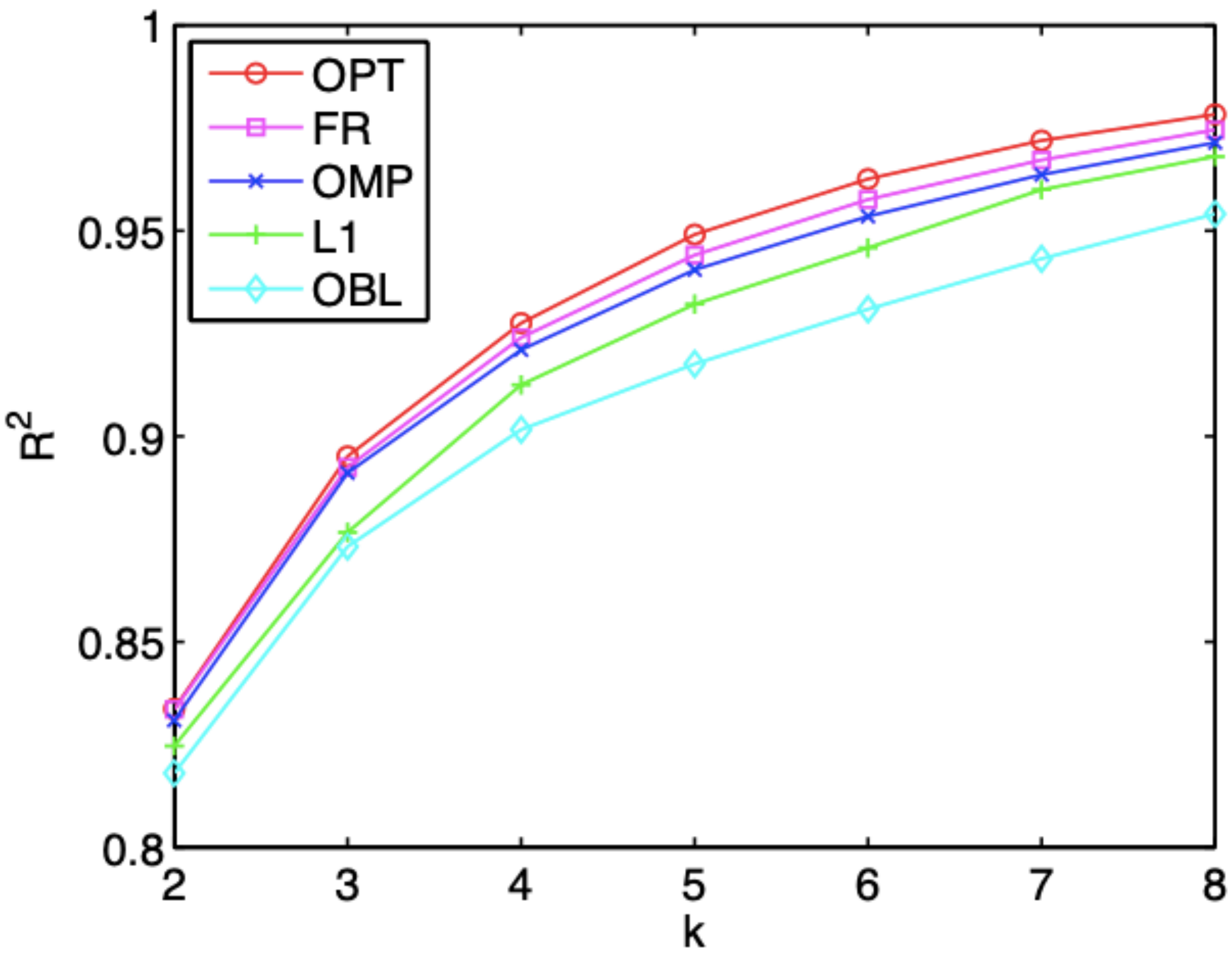


Figure 5: Synthetic Data R^2

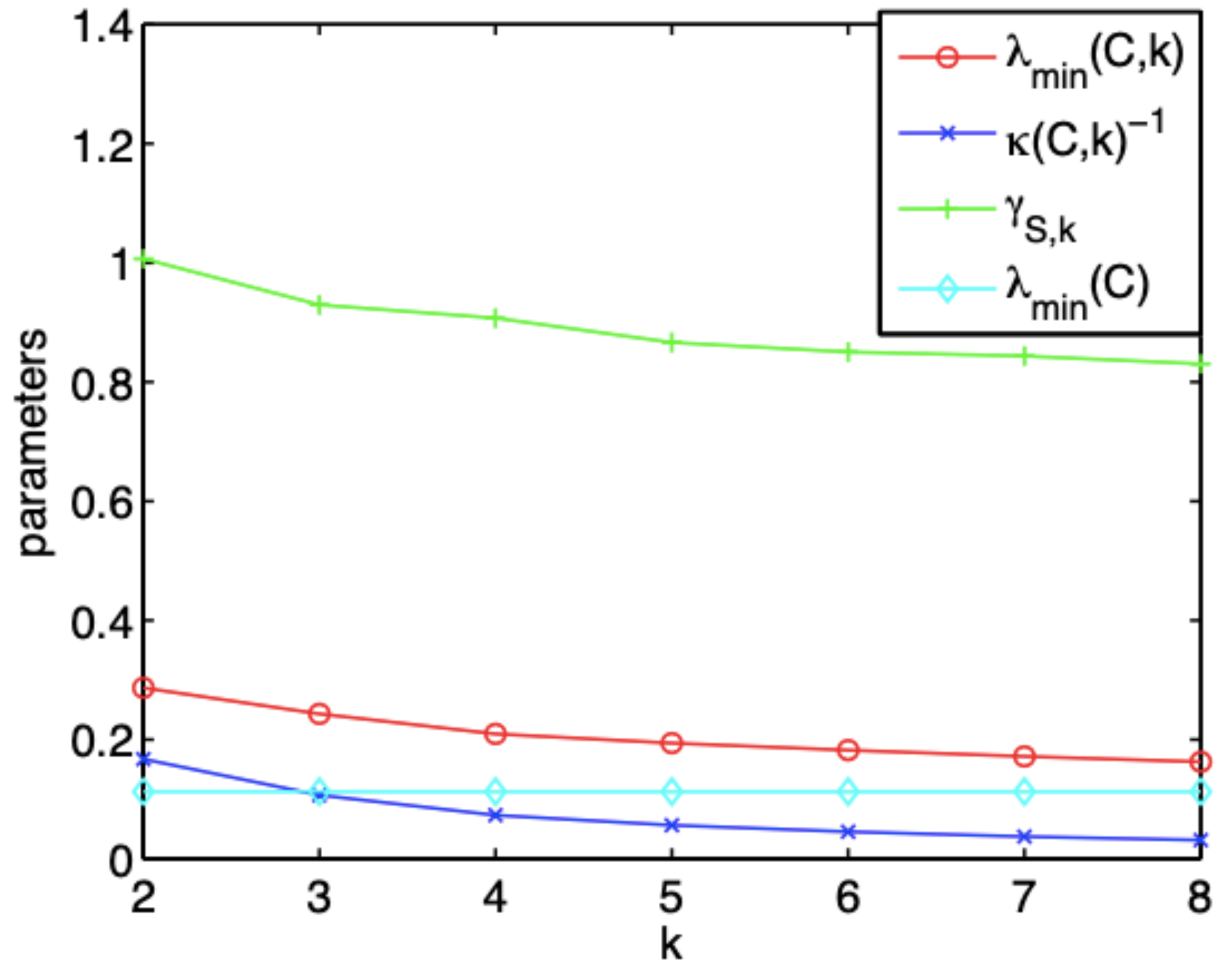


Figure 6: Synthetic Data parameters

Experiments - Subset Selection:

Pushing the Bounds

$$\gamma_{U,k}(f) = \min_{\substack{L \subseteq U, S: |S| \leq k, S \cap L = \emptyset \\ \frac{f(L \cup \{x\})}{f(L)} \geq 1+\epsilon,\ \forall x \in S}} \frac{\sum_{x \in S} \left(f(L \cup \{x\}) - f(L) \right)}{f(L \cup S) - f(L)}$$

$$\gamma \simeq 0.4 \quad \overset{\epsilon = 0.2}{\Rightarrow} \quad \gamma \simeq 0.8 \sim 1.74$$

$$1 - e^{\gamma} \simeq 33\,\% \quad \Rightarrow \quad 1 - e^{\gamma} \simeq 55\,\% \sim 82\,\%$$

Conclusion:

Further Applications

- Feature selection
  (Johnson et al. 2015)
- Maximum coverage / Minimum cover
  (Das and Campe 2018)
- Generalised Linear Models
  (Elenberg et al. 2016)

Conclusion:

Summary

- Introduced **submodularity ratio, γ**.
- Analysed greedy algorithms with γ, and got the strongest known bounds for subset selection and dictionary selection.
- Demonstrated the superiority of γ over spectral parameters by experiments.

Questions?

Thank you for your attention!