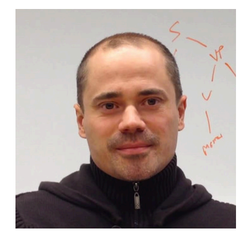
Greedy Minimization of Weakly Supermodular Set Functions

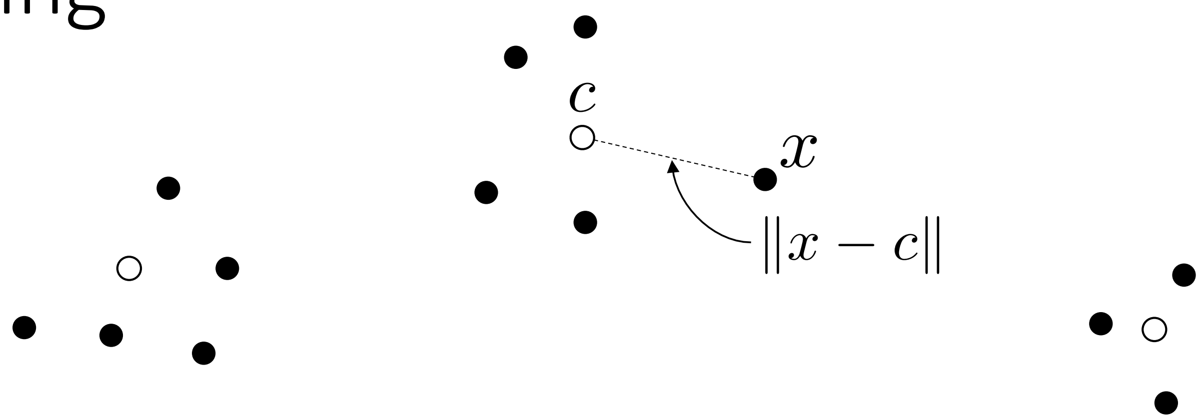
Edo Liberty (Amazon)
Maxim Sviridenko (Yahoo)



High level view

1. Machine learning involves **optimization**
2. Often, **minimizing** a **set function** with **cardinality constraints**
3. Many of which are **weakly supermodular**
4. A **greedy extension algorithm** works well for those

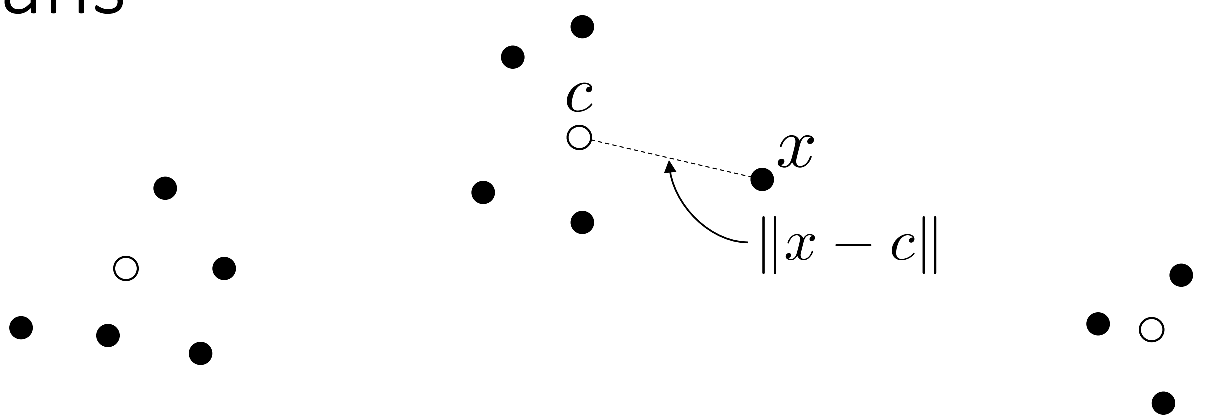
Clustering



$$f(S) = \sum_{x \in X} \min_{c \in S} w(x,c)$$

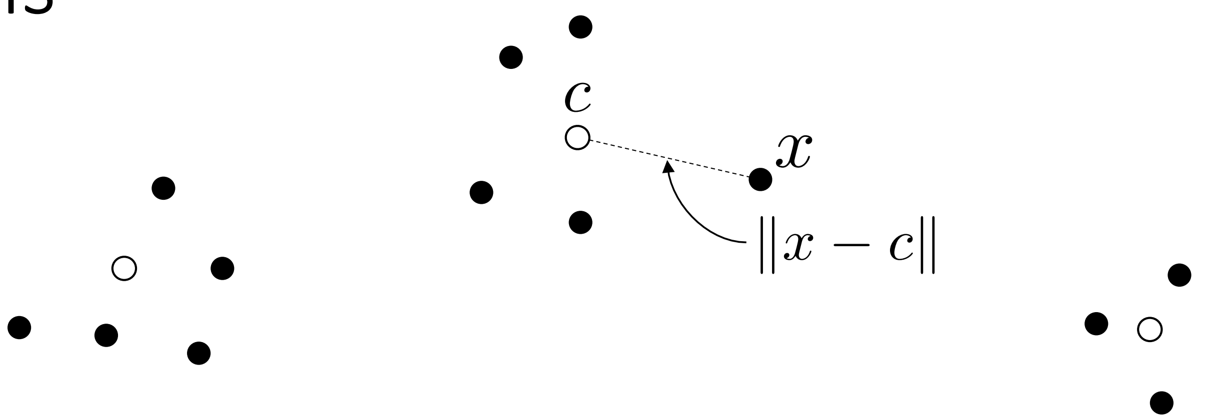
Subject to $|S| \leq k$

K-Medians



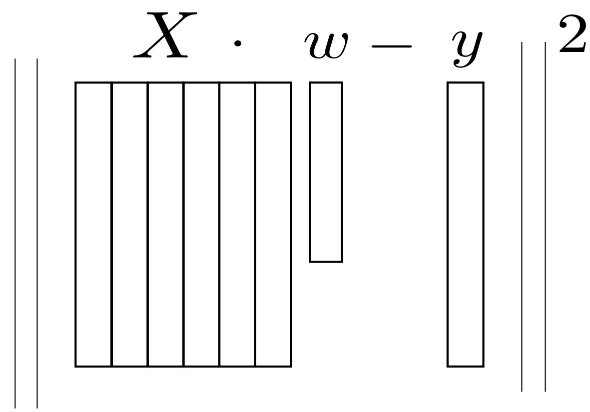
$$f(S) = \sum_{x \in X} \min_{c \in S} \|x - c\| \qquad \text{Subject to} \quad |S| \leq k$$

K-Means



$$f(S) = \sum_{x \in X} \min_{c \in S} \|x - c\|^2 \qquad \text{Subject to} \quad |S| \leq k$$

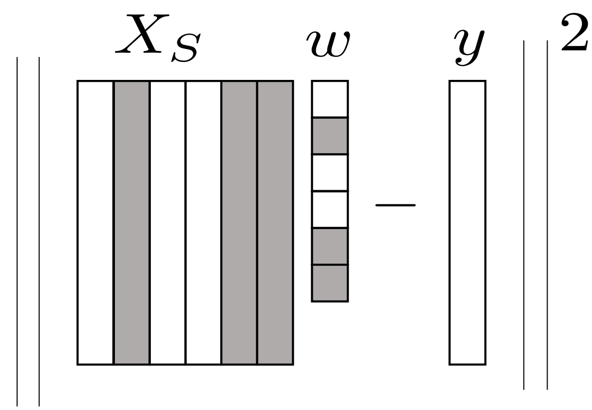
Sparse Regression



$$\min_{w} \|Xw - y\|^2 \text{ such that } |\mathrm{supp}(w)| \le k$$

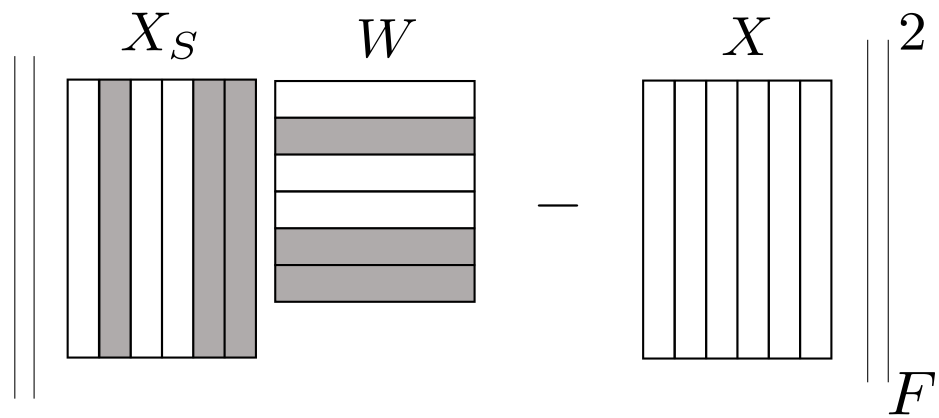
- Bi-criteria – [Natarajan 95]
- NP hard – [Foster, Karloff, Thaler 15]

Sparse Regression



$$f(S) = \|X_S X_S^{+} y - y\|^2 \quad \text{Subject to} \quad |S| \leq k$$

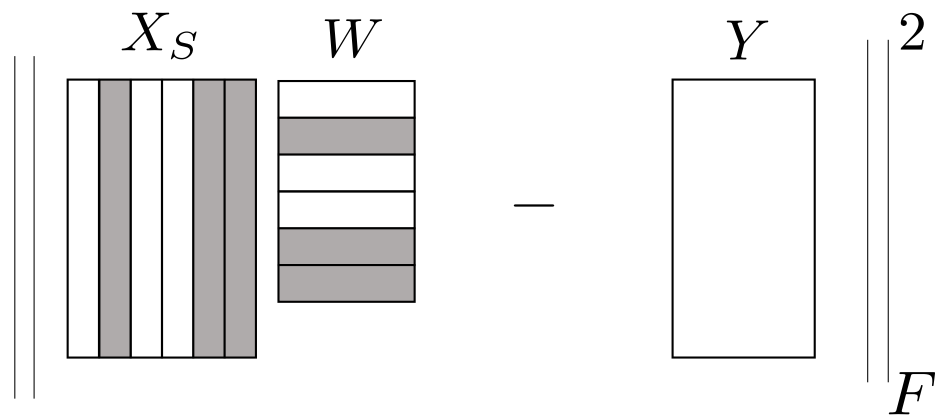
Columns Subset Selection



$$f(S) = \|X_S X_S^+ X - X\|_F^2 \quad \text{Subject to} \quad |S| \le k$$

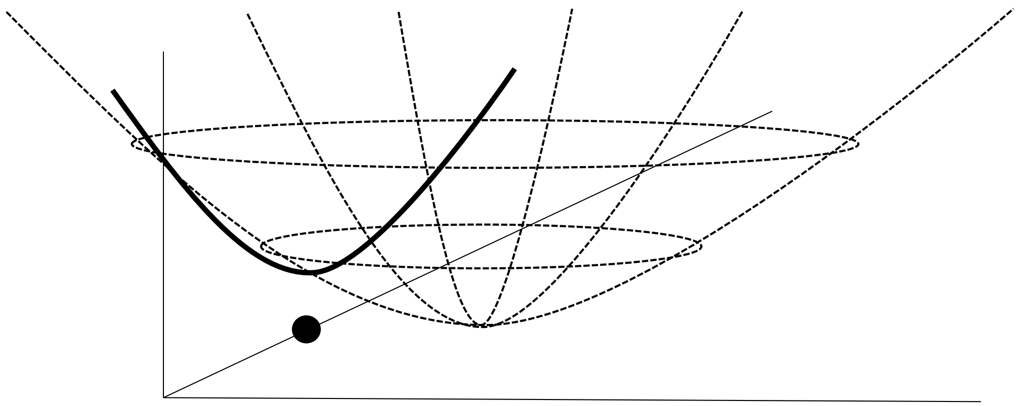
- [Deshpande, Rademacher 10]
- [Boutsidis, Drineas, Magdon-Ismail 14]

Sparse Multiple Linear Regression



$$f(S) = \|X_S X_S^+ Y - Y\|_F^2 \quad \text{Subject to} \quad |S| \leq k$$

Sparse Convex Function Minimization



$$\min_x R(x) \quad \text{such that} \quad |\operatorname{supp}(x)| \leq k$$

- [Shalev-Shwartz, Srebro, Zhang 10]

Weak Supermodularity

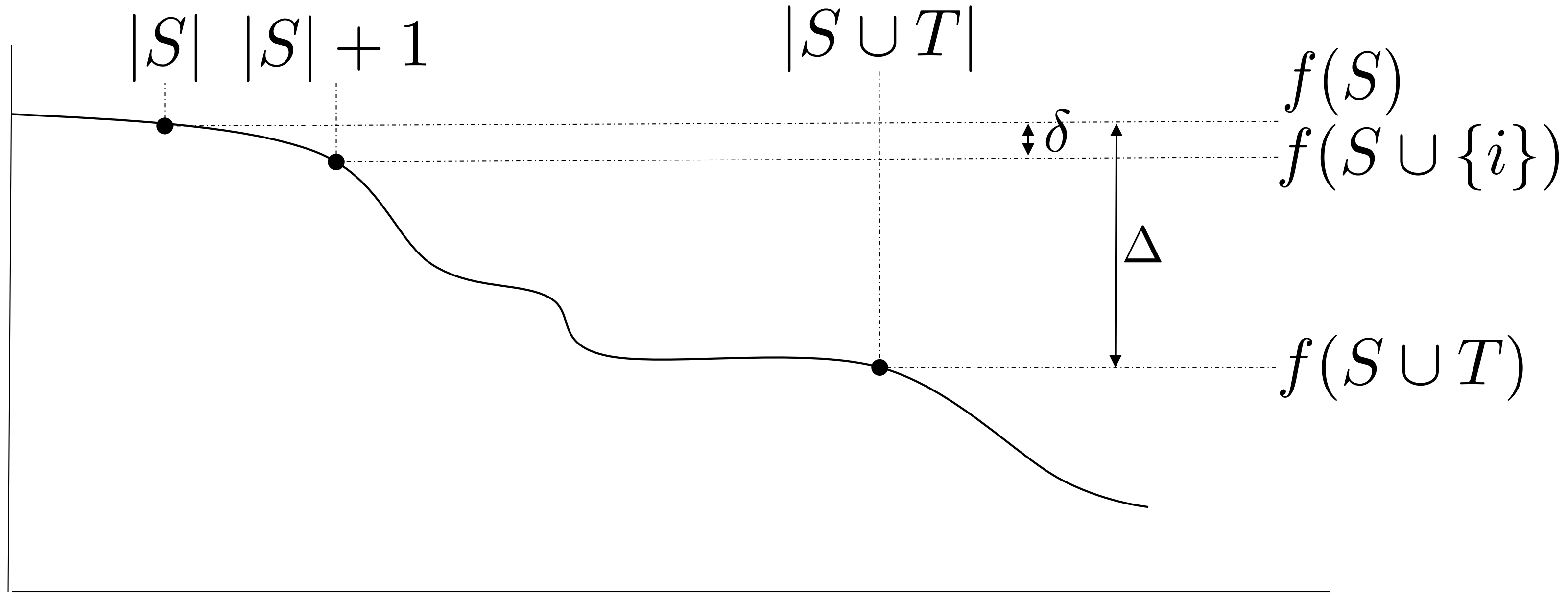
Definition 1. *A set function* $f(S) : 2^{[n]} \to \mathbb{R}_+$ *which is*

- *Non-negative* - $f(S) \geq 0$
- *non-increasing* - $f(S) \geq f(S \cup T)$

is said to be weakly-α*-supermodular if there exists* $\alpha \geq 1$ *such that for any two sets* $S, T \subseteq [n]$

$$f(S) - f(S \cup T) \leq \alpha \sum_{i \in T \setminus S} (f(S) - f(S \cup \{i\}))$$

Weak Supermodularity



$$\exists i \in T \setminus S \;\; s.t. \;\; \delta \geq \frac{\Delta}{\alpha |T \setminus S|}$$

Weakly Supermodular Problems

Problem	alpha
k-medians	1
k-means	1
Sparse Regression	$\max_{S'} \|X_{S'}^{+}\|_2^2$
Column subset Selection	$\max_{S'} \|X_{S'}^{+}\|_2^2$
Sparse Multiple Linear Regression	$\max_{S'} \|X_{S'}^{+}\|_2^2$
Sparse Convex Function Minimization (for λ strongly convex and β smooth)	β/λ

Greedy algorithms and Sub/Supermodularity

- Nemhauser, Wolsey, Fisher 78
  - $(1 - 1/e)$ approx for greedy algorithm on maximizing supermodular functions
  - $(1 - \varepsilon)$ approx using $|S| \leq k \log(1/\varepsilon)$

- Das, Kempe 11
  - Define submodulairy-ratio which is analogues to our alpha
  - Give guaranties and bicriteria for maximization problem

- Folklore
  - Supermodular Minimization $\neq$ Submodular Maximization
  - Approximation for Supermodular Minimization can be NP hard.

Algorithm 1 Greedy Extension Algorithm

input: Weakly-α-supermodular function $f(S)$, initial set S_0, parameters $k \in \mathbb{Z}_+$ and the sequence $\Lambda_1, \Lambda_2, \ldots$

while $t \leq \lceil \alpha k \ln \Lambda_t \rceil$ **do**

$S_t \leftarrow S_{t-1} \cup \arg\min_{i \in [n]} f(S_{t-1} \cup \{i\})$

output: S_t

Lemma 1. *Let S_τ be the output of the greedy algorithm. Then $|S_\tau| \leq |S_0| + \lceil \alpha k \ln \Lambda_\tau \rceil$ and $f(S_\tau) \leq f(S^*) + \frac{f(S_0) - f(S^*)}{\Lambda_{\tau+1}}$ where S^* is an optimal solution of the optimization problem.*

Analysis

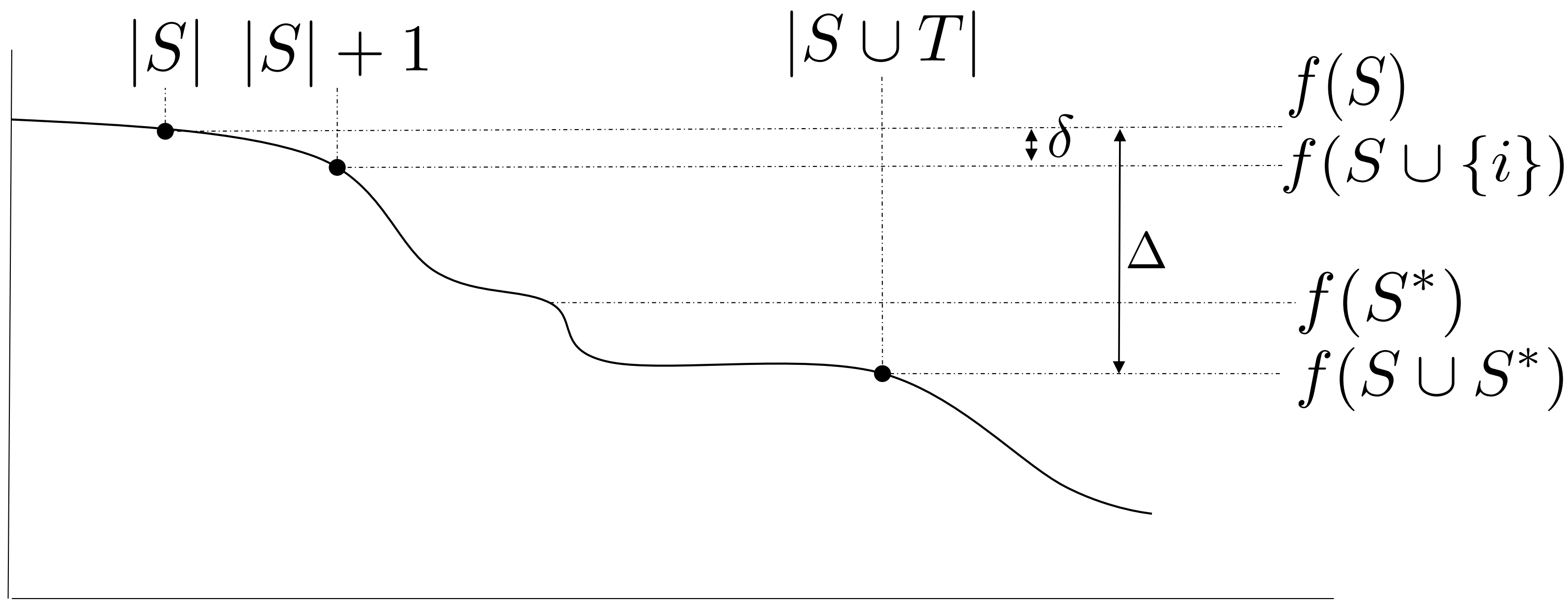
$$
\begin{aligned}
f(S_{t-1}) - f(S^*) &\leq f(S_{t-1}) - f(S_{t-1} \cup S^*) \\
&\leq \alpha \cdot \sum_{i \in S^* \setminus S_{t-1}} f(S_{t-1}) - f(S_{t-1} \cup \{i\}) \\
&\leq \alpha k \cdot \max_{i \in [n]} f(S_{t-1}) - f(S_{t-1} \cup \{i\}) \\
&= \alpha k \cdot (f(S_{t-1}) - f(S_t)) \ .
\end{aligned}
$$

By rearranging the above equation and recursing over t we get

$$
\begin{aligned}
f(S_t) - f(S^*) &\leq (f(S_{t-1}) - f(S^*)) (1 - 1/\alpha k) \\
&\leq (f(S_0) - f(S^*)) (1 - 1/\alpha k)^t
\end{aligned}
$$

Substituting $\tau + 1 > \lceil \alpha k \ln \Lambda_{\tau+1} \rceil$ completes the proof

Weak Supermodularity



Every element added cuts the distance to $f(S^*)$ by fraction $(1 - 1/\alpha k)$

Algorithm 2 Greedy Extension Algorithm

input: Weakly-α-supermodular function $f(S)$, initial set S_0, $k \in \mathbb{Z}_+$
while $t \leq \lceil \alpha k \ln(f(S_0)/\varepsilon f(S_{t-1})) \rceil$ **do**
  $S_t \leftarrow S_{t-1} \cup \arg\min_{i \in [n]} f(S_{t-1} \cup \{i\})$
output: S_t

- this is instance of Algorithm 1 with $\Lambda_t = f(S_0)/\varepsilon f(S_{t-1})$
- Then we have $f(S_\tau) \leq f(S^*)/(1-\varepsilon)$
- And $|S_t| \leq |S_0| + \lceil \alpha k \ln(\frac{1}{\varepsilon} \frac{f(S_0)}{f(S^*)}) \rceil$

Algorithm 3 Greedy Extension Algorithm; an alternative stopping criterion

input: Weakly-α-supermodular function f, S_0, f_{stop}

repeat

$S_t \leftarrow S_{t-1} \cup \arg\min_i f(S_{t-1} \cup \{i\})$

until $f(S_t) \leq f_{\text{stop}}$

output: $S = S_t$

- He have
$$|S| \leq |S_0| + \left\lceil \alpha k' \left(\ln \frac{f(S_0) - f'}{f_{\text{stop}} - f'} \right) \right\rceil$$

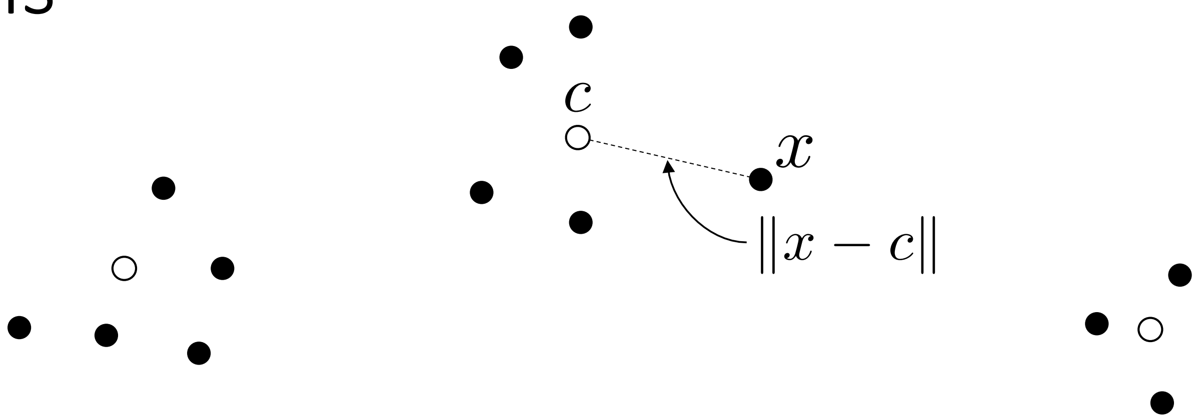
- Where
$$k' = \min |S'| \text{ such that } f(S') \leq f'$$

Recipe for New Bi-Criteria algorithms

- Bound alpha for your problem
- Generate S_0 such that $f(S_0)/f(S^*) \le \rho$ using a known ρ-approximation algorithm.
- Use the given greedy extension algorithm

- output S_t
- Such that $|S_t| \le |S_0| + \lceil \alpha k \ln(\rho/\varepsilon) \rceil$
- and $f(S_t) < (1+\varepsilon) f(S^*)$

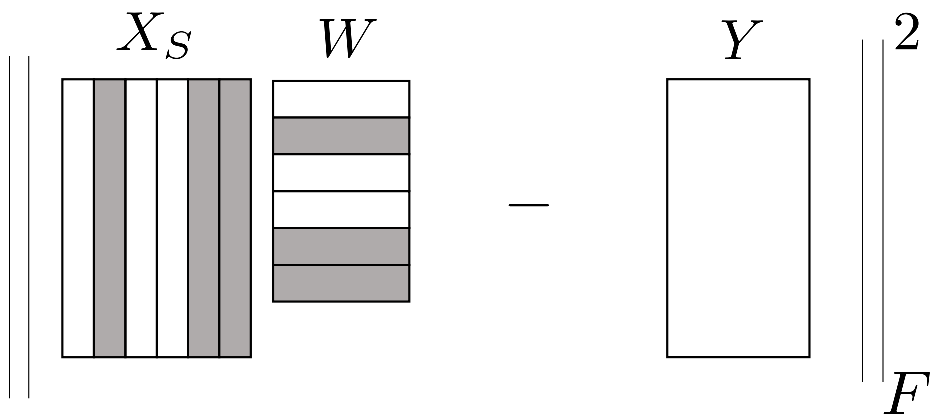
K-Means



▶ **Lemma 8.** *For the constrained k-means problem, one can find in $O(n^2 dk \log(1/\varepsilon))$ time a set S of size $|S| = O(k) + k\log(1/\varepsilon)$ such that $f(S) \leq (1+\varepsilon)f(S^*)$ where $f(S^*)$ is the optimal solution.*

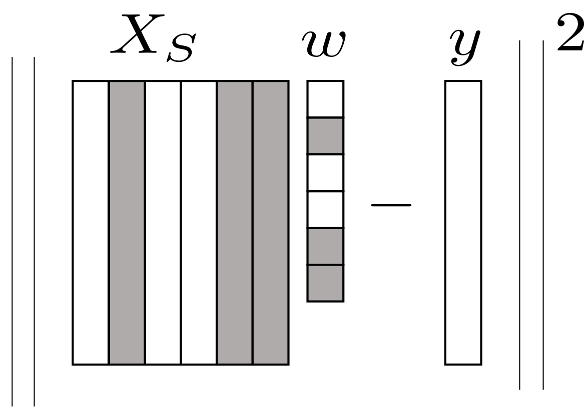
▶ **Lemma 12.** *Let $f(S^*)$ be the optimal solution to the unconstrained k-means problem. One can find in time $O(n^{O(\log(1/\varepsilon)/\varepsilon^2)}dk)$ a set $S \in \mathbb{R}^d$ of size $|S| = O(k) + k\log(1/\varepsilon)$ such that $f(S) \leq (1+\varepsilon)f(S^*)$.*

Sparse Multiple Linear Regression



▶ **Lemma 13.** *For* $X \in \mathbb{R}^{m \times n}$ *and* $Y \in \mathbb{R}^{m \times \ell}$ *the SMLR minimization function* $f(S) = \|Y - X_S X_S^+ Y\|_F^2$ *is* α*-weakly-supermodular with* $\alpha = \max_{S'} \|X_{S'}^+\|_2^2$.

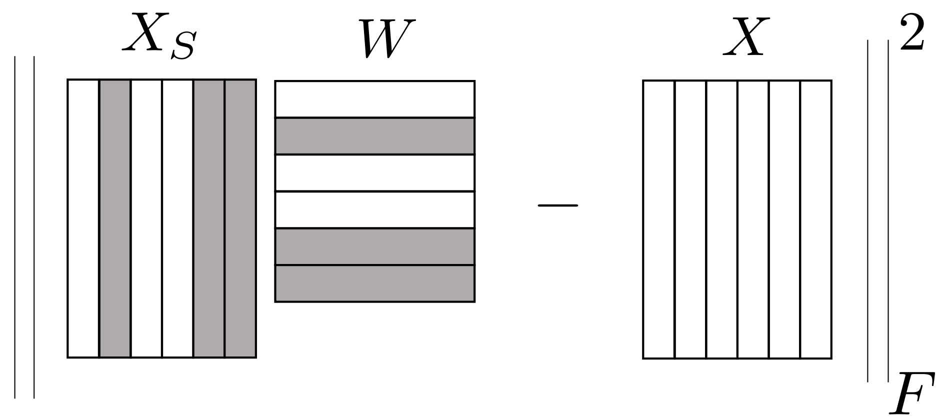
Sparse Regression



Natarajan's analysis gets $$|S| \leq \left\lceil 9k\alpha \ln \frac{\|y\|_2^2}{E} \right\rceil$$

Simply by invoking Algorithm 3 $$|S| \leq \left\lceil k\alpha \ln \frac{\|y\|_2^2 - E/4}{E - E/4} \right\rceil \leq \left\lceil \frac{4}{3} k\alpha \ln \frac{\|y\|_2^2}{E} \right\rceil$$

Columns Subset Selection

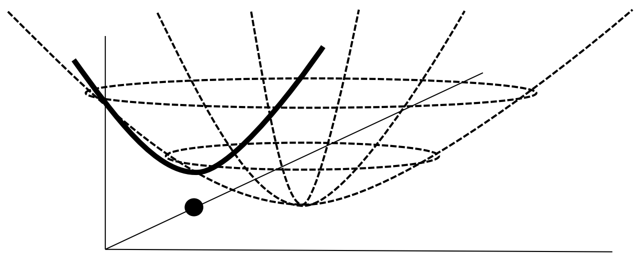


Initialing, for example, with [Boutsidis, Drineas, Magdon-Ismail 14]

$$f(S) \leq (1+\varepsilon) f(S^*) \qquad \text{and} \qquad |S| = O(\alpha k \ln(1/\varepsilon))$$

Previous results required polynomial dependence on epsilon

Sparse Convex Function Minimization



▶ **Theorem 19.** *Given the set function $f(S)$ defined in (6) corresponding to β-smooth λ-strongly convex function $R(w)$. The set function $f(S)$ is α-weakly-supermodular with $\alpha = \frac{\beta}{\lambda}$.*

▶ **Theorem 20.** *For any $\varepsilon > 0$, let $f_{\text{stop}} = R^* + \varepsilon$ then the Algorithm 3 outputs S such that*

$$|S| \leq \left\lceil \frac{\beta}{\lambda} k_f \left(\ln \frac{R(\emptyset) - R^*}{\varepsilon} \right) \right\rceil.$$

This reproves Theorem 2.8 in [Shalev-Shwartz, Srebro, Zhang 10]

Take home message

1. Machine learning involves **optimization**
2. Often, **minimizing** a **set function** with **cardinality constraints**
3. Many of which are **weakly supermodular**
4. A **greedy extension algorithm** works well for those