CS-GY 6763/CS-UY 3943: Lecture 7
Submodularity

NYU Tandon School of Engineering, R. Teal Witter

SET FUNCTIONS

Consider a set function $f : 2^{[n]} \to R$.

Example: There are $n = 3$ classes and f represents the knowledge gained from a set of classes.

$[3] = \{1,2,3\}$

S	$f(S)$
$\{\}$	0
1	5
2	3
3	10
1,2	7
1,3	11
2,3	11
1,2,3	12

normalized

monotone

"diminishing returns"

$\max\{f(1), f(2)\} \leq \qquad \leq f(1) + f(2)$

SUBMODULARITY

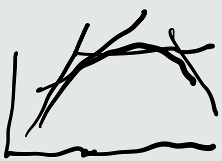
For $e \in [n]$ and $S \subseteq [n]$, the marginal gain of element e with respect to set S is

$$f(e|S) = f(\{e\} \cup S) - f(S).$$

discrete derivative

Definition (Submodular set function)

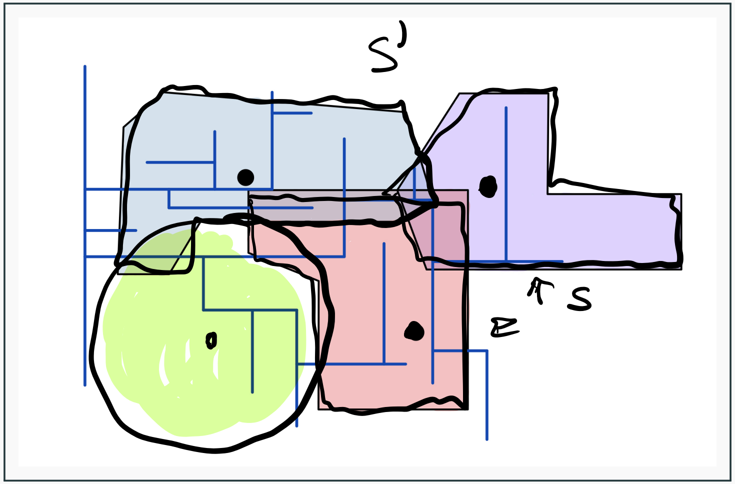
A set function $f : 2^{[n]} \to \mathbb{R}$ is submodular if, for all $e \in [n]$ and $S \subseteq S' \subseteq [n]$,



$$f(e|S) \geq f(e|S').$$

non-increasing derivative

In coverage problem, $f(S)$ is the amount of water "covered" by S.



$f(e|S) \geq f(e|S')$

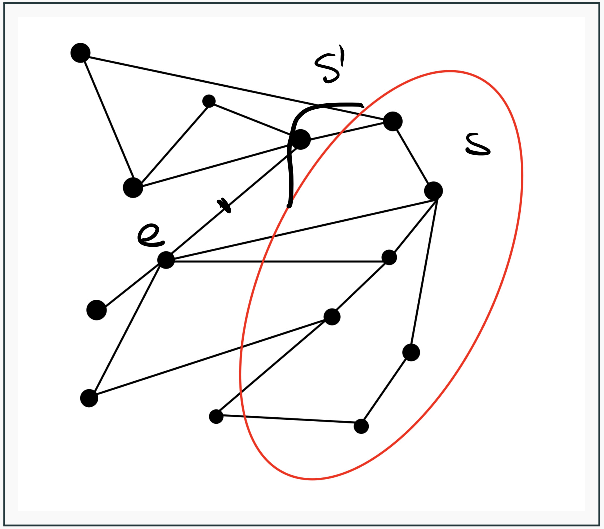
for $S \subseteq S'$

Why is coverage *submodular*?

[1]Finding a maximum of at most k hyper-edges is NP-Hard.

APPLICATION: GRAPH CUT[2]

In graph cut, $f(S)$ is the number of edges between S and $[n] \setminus S$.



$f(S) = 7$

$f(e|S) \geq f(e|S')$

$S \subseteq S'$

Why is graph cut *submodular*?

$f(\{\}) = 0 = f([n])$

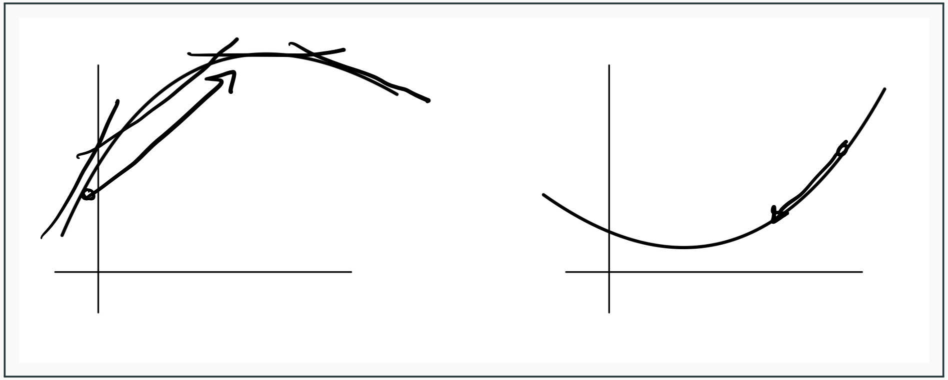
How is graph cut different from set cover?

[2] Finding a maximum graph cut is NP-Hard.

APPLICATION: OTHERS!

- **Combinatorial optimization** (1970-)
  - rank of a matroid
  - submodular flows
- **Algorithmic game theory** (2000-)
  - marketing on networks
  - combinatorial auctions
- **Machine learning** (2005-)
  - document summarization
  - active learning

IS SUBMODULARITY MORE LIKE CONCAVITY OR CONVEXITY?



Arguments for concavity:

- Non-increasing derivative.

Arguments for convexity:

- Max cover and max cut are NP-hard.
- Exact minimization can be done in polynomial time.[3]

[3] M. Grötschel, L. Lovász & A. Schrijver. *Combinatorica* (1981).

APPROXIMATE SUBMODULAR MAXIMIZATION

$S \subseteq S'$

$f(\{\}) = 0$ $f(S) \leq f(S')$

Let $f : 2^{[n]} \to \mathbb{R}$ be a normalized, monotone, submodular set function.

$f(e|S) \geq f(e|S')$

We want to find

$$\arg\max_{S \subseteq [n]} f(S) \text{ subject to } |S| \overset{=}{\cancel{\leq}} k.$$

Greedy! Local update

… any ideas?

Let $\{\} = S_0 \subset S_1 \subset S_2 \subset \ldots \subset S_n = [n]$ with $S_i = \{s_1, s_2, \ldots, s_i\}$.

Theorem (Nemhauser-Wolsey 1981) Tight

Let $f : 2^{[n]} \to \mathbb{R}$ be a normalized, monotone, submodular set function. Fix positive integers ℓ and k. Choose # in optimal

$s_i = \arg\max_{e \in [n] \setminus S_i} f(e|S_i)$. *Then* # elements in greedy

$\ell = k$

$$f(S_\ell) \geq (1 - \boxed{e}^{-\ell/k}) f(S^*)$$

where

$$S^* = \arg\max_{S : |S| = k} f(S).$$

Claim: $\underbrace{f(S^*) - f(S_i)}_{\text{monotonicity}} \leq k\,[\underbrace{f(S_{i+1}) - f(S_i)}]$

$$\leq f(S^* \cup S_i) - f(S_i) \quad [= f((S^* \setminus S_i) \cup S_i) - f(S_i) = f(S^* \setminus S_i \mid S_i)]$$

$$= f(\{s_1^*, s_2^*, \dots, s_k^*\} \cup S_i) - f(\{s_1^*, \dots, s_{k-1}^*\} \cup S_i)$$

$$+ f(\{s_1^*, \dots, s_{k-1}^*\} \cup S_i) - f(\{s_1^*, \dots, s_{k-2}^*\} \cup S_i)$$

$$+ f(\{s_1^*\} \cup S_i) \qquad - f(S_i)$$

$$= \sum_{j=1}^{k} f(s_j^* \cup S_i \cup \{s_1^*, \dots, s_{j-1}^*\}) - f(S_i \cup \{s_1^*, \dots, s_{j-1}^*\})$$

$$\qquad\qquad \| \qquad f(s_j^* \mid S_i \cup \{s_1^*, \dots, s_{j-1}^*\})$$

$$\underset{\text{submod}}{\leq} \sum_{j=1}^{k} f(s_j^* \mid S_i) \underset{\text{greedy}}{\leq} \sum_{j=1}^{k} f(s_{i+1} \mid S_i) = k\,[f(\underbrace{s_{i+1} \cup S_i}_{S_{i+1}}) - f(S_i)]$$

proved

Claim: $\frac{1}{k}(f(S^*) - f(S_i)) \leq k\,[f(S_{i+1}) - f(S_i)]$

$$= k\left[\underbrace{(f(S^*) - f(S_i))} - (f(S^*) - f(S_{i+1}))\right]$$

$$f(S^*) - f(S_{i+1}) \leq \left(1 - \frac{1}{k}\right)[f(S^*) - f(S_i)]$$

$$\leq \left(1 - \frac{1}{k}\right)\left(1 - \frac{1}{k}\right)[f(S^*) - f(S_{i-1})]$$

$$\leq \left(1 - \frac{1}{k}\right)^{i+1}[f(S^*) - \cancel{f(S_{i-1})}]$$

0

$$f(S^*) - f(S_\ell) \leq \left(1 - \frac{1}{k}\right)^{\ell} f(S^*)$$

$$\left(\left(1 - \frac{1}{k}\right)^{k}\right)^{\ell/k} \leq \boxed{e}^{-\ell/k}$$

≈ 2.71

$$f(S^*)\left(1 - \boxed{e}^{-\ell/k}\right) \leq f(S_\ell)$$

GREEDY APPROXIMATION FACTOR

If we use the same number of sensors as optimal ($\ell = k$), then we get a $(1 - e^{-1}) \approx .63$ approximate solution.

If we use five times as many sensors as optimal ($\ell = 5k$), then we get a $(1 - e^{-5}) \approx .99$ approximate solution.

What if we have a set function is only *close* to submodular?

Definition (Weak Submodularity)

Fix a positive integer k. A set function $f : 2^{[n]} \to \mathbb{R}$ is γ_k-weakly submodular for k if, for all $S' \in [n]$ and $S \subset [n] \setminus S'$ where $|S| \leq k$,

$$\gamma_k(f) \leq \frac{\sum_{e \in S} f(e|S')}{f(S|S')}.$$

Intuition: How much f can increase by adding a set of size k vs. combined increase of each element.

Definition (Weak Submodularity)

Fix a positive integer k. A set function $f : 2^{[n]} \to \mathbb{R}$ is γ_k-weakly submodular for k if, for all $S' \in [n]$ and $S \subset [n] \setminus S'$ where $|S| \underset{=}{\leq} k$,

$$\gamma_k(\cancel{f}) \leq \frac{\sum_{e \in S} f(e|S')}{f(S|S')}.$$

Sanity check: What is $\gamma_k(\cancel{f}) = 1$ if f is submodular?

$$f(S|S') = \sum_{i=1}^{|S|} f(s_i | S' \cup \{s_1, \dots, s_{i-1}\}) \leq \sum_{i=1}^{|S|} f(s_i | S')$$

Theorem (Das-Kempe 2011)

Let $f : 2^{[n]} \to \mathbb{R}$ be a normalized, monotone, γ_k-weakly submodular set function. Choose $s_i = \arg\max_{e \in [n] \setminus S_i} f(e|S_i)$. Then

$\gamma_k \geq 1$ $\quad$ $\ell = k$

$$f(S_\ell) \geq (1 - e^{-\gamma_k}) f(S^*).$$

$$\geq (1 - e^{-1})\, f(S^*)$$

f is submodular

nxd dx1 nx1

We want to minimize $||\mathbf{A}\mathbf{x} - \mathbf{b}||^2$ over $\mathbf{x} \in \mathbb{R}^d$.

Recall from last class that the Hessian $\mathbf{H}$ of least squares regression is $2\mathbf{A}^T\mathbf{A}$ and so

dxn nxd

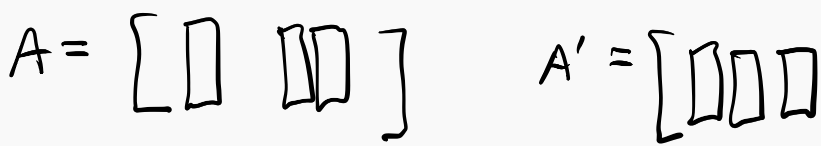
$$\alpha \mathbf{I}_{d\times d} \preceq 2\mathbf{A}^T\mathbf{A} \preceq \beta \mathbf{I}_{d\times d}$$

where we say $\mathbf{H}$ is α-strongly convex and β-smooth. In particular, we argued $\alpha = \lambda_{\min}(2\mathbf{A}^T\mathbf{A})$ and $\beta = \lambda_{\max}(2\mathbf{A}^T\mathbf{A})$.

Question: What if we can only choose k features?

$x' = \begin{bmatrix} x_3 \\ x_7 \\ x_8 \end{bmatrix}$ $\quad x = \begin{bmatrix} \cancel{x_1} \\ \cancel{x_2} \\ \vdots \\ \cancel{x_d} \end{bmatrix} \begin{matrix} 0 \\ 0 \\ \\ 0 \end{matrix}$ k non-zero

We want to minimize $||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2$ over $\mathbf{x} \in \mathbb{R}^d$ where $k << d$ entries in $\mathbf{x}$ are non-zero. Let $\mathbf{x}'$ be the 'condensed' $k \times 1$ vector and $\mathbf{A}'$ be the 'condensed' $n \times k$ matrix.



$||A'x' - b||_2^2$

$x' \in \mathbb{R}^k$

$\leq \lambda_{min}(H')$ $\qquad \leq \lambda_{max}(H')$

Then H' is α'-strongly convex and β'-smooth.

Exercise: Why is $\alpha \leq \alpha'$ and $\beta' \leq \beta$?

Theorem (Elenberg-Khanna-Dimakis-Negahban 2018)

Let $\max_{S:|S|\le k} f(S) = \max_{\mathbf{x}'} -||\mathbf{A}'\mathbf{x}' - \mathbf{b}||^2$. *Then*

$$\gamma_k \ge \frac{\alpha'}{\beta'}.$$

Corollary: Greedily choosing k features gives a $1 - e^{-\lambda_{\min}(2A'^T A')/\lambda_{\max}(2A'^T A')}$ -approximation to the optimal features.

TAKEAWAYS

- Greedy solutions often work well
- Our tools (bound progress, $(1 - 1/x)^x \leq 1/e$) are versatile
- Submodularity research is shallow (rather than deep)

Thank you!