Massachusetts Institute of Technology

Submodular Functions and Machine Learning

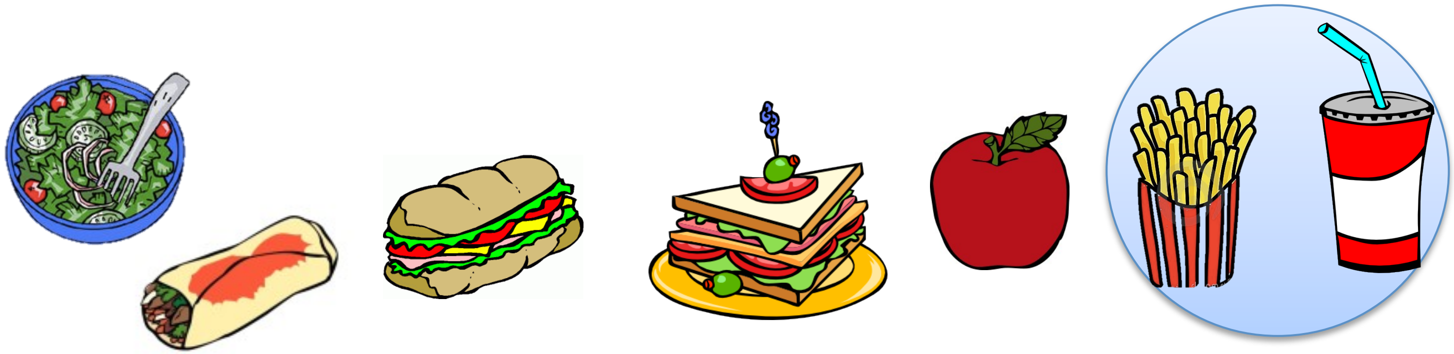
MLSS Kyoto

Stefanie Jegelka

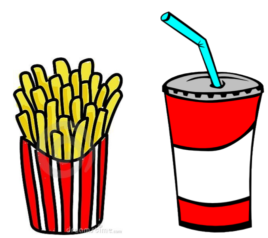
MIT

Set functions

ground set

$\mathcal{V} = \{$  $\}$

$$F : 2^{\mathcal{V}} \to \mathbb{R}$$

$F($  $) =$ cost of buying items together, or

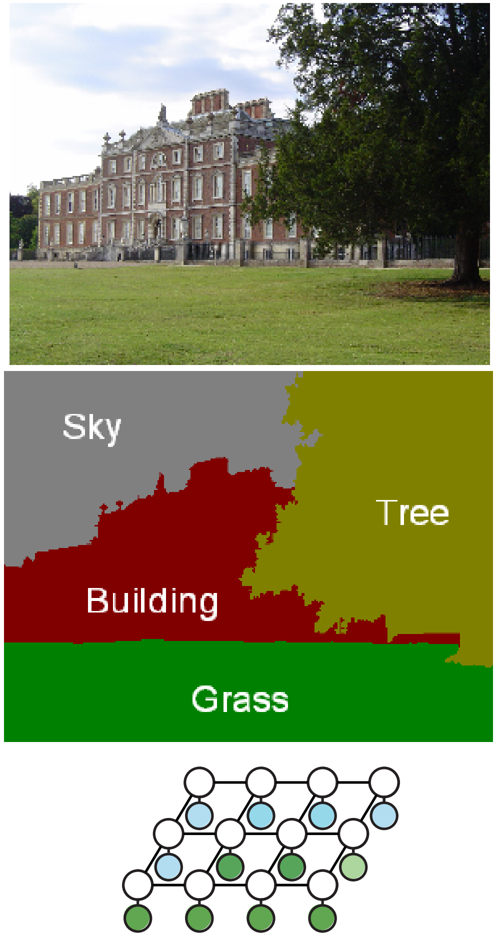
utility, or

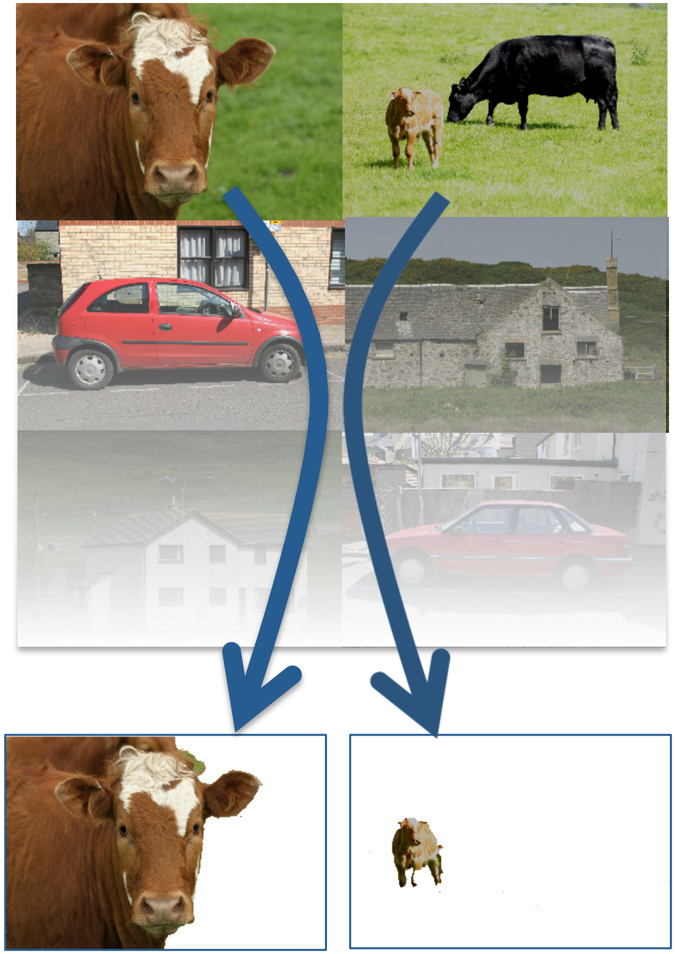
probability, ...

We will assume:

- $F(\emptyset) = 0$
- black box "oracle" to evaluate *F*

Discrete Labeling

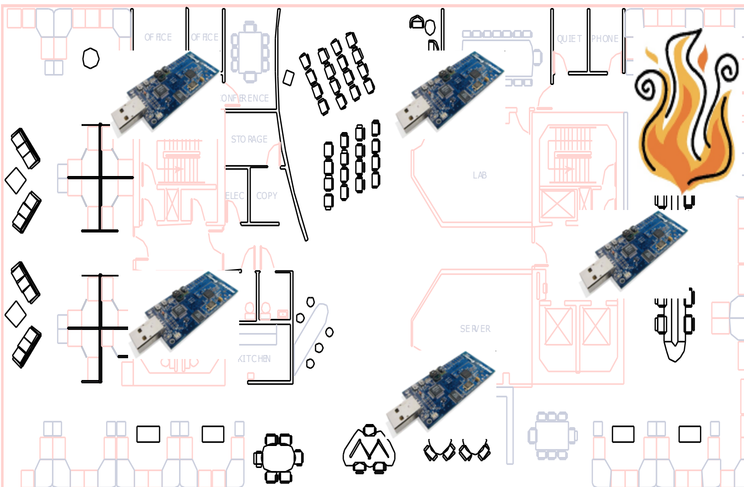




Summarization

$$F(S) = \text{relevance} + \text{diversity or coverage}$$

Informative Subsets

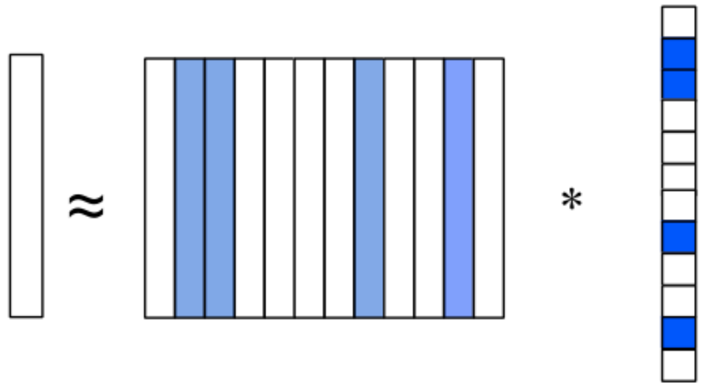


- where put sensors?
- which experiments?
- summarization

$$F(S) = \text{“information”}$$

Sparsity

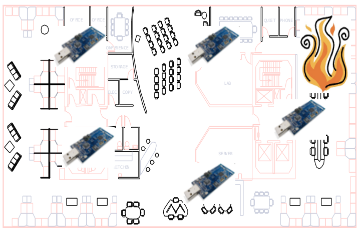
$$y = A\,{\color{red}x} + \text{noise}$$

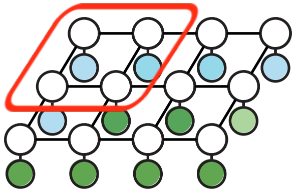


$$F(S) = \text{“penalty on support pattern”}$$

Formalization

- Formalization:
  Optimize a set function F(S) (under constraints)





- generally very hard ☹
- submodularity helps:
  efficient optimization & inference with guarantees!
  ☺

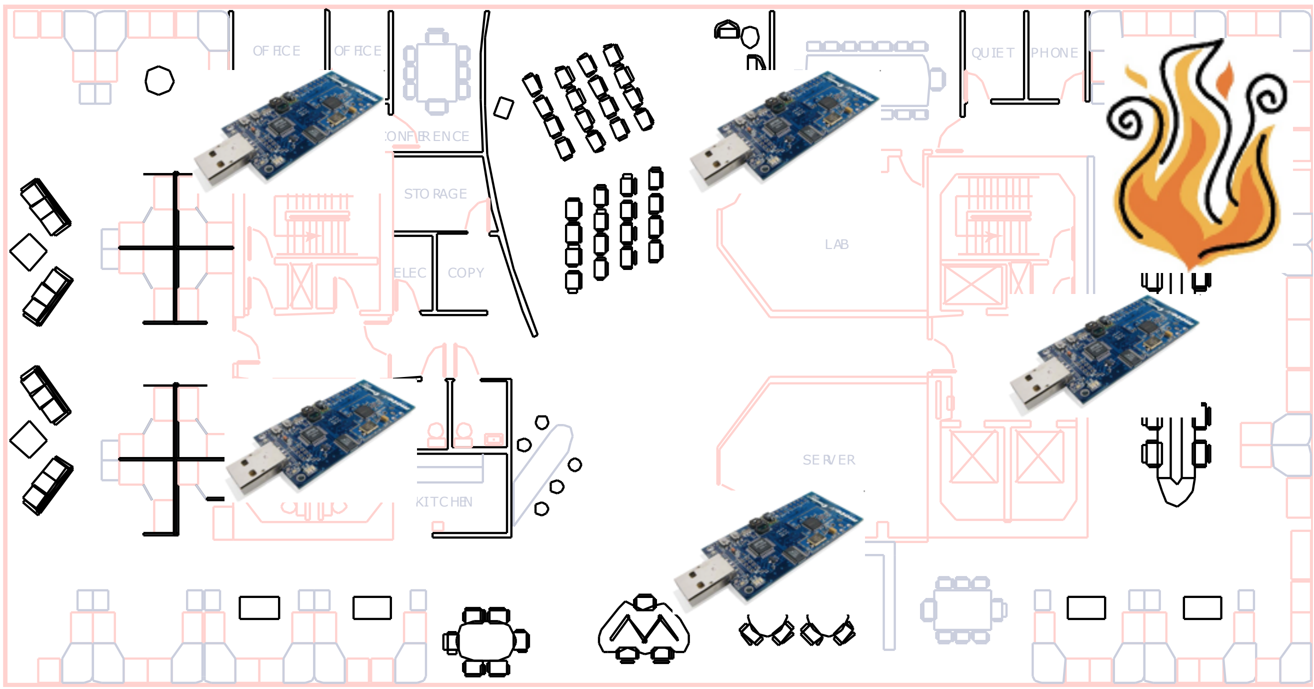
Roadmap

- Submodular set functions
  - definition & basic properties
  - links to convexity
  - special polyhedra

- Minimizing submodular functions
  coherence, regularization, convexity

- Maximizing submodular functions
  diversity, repulsion, concavity

Sensing



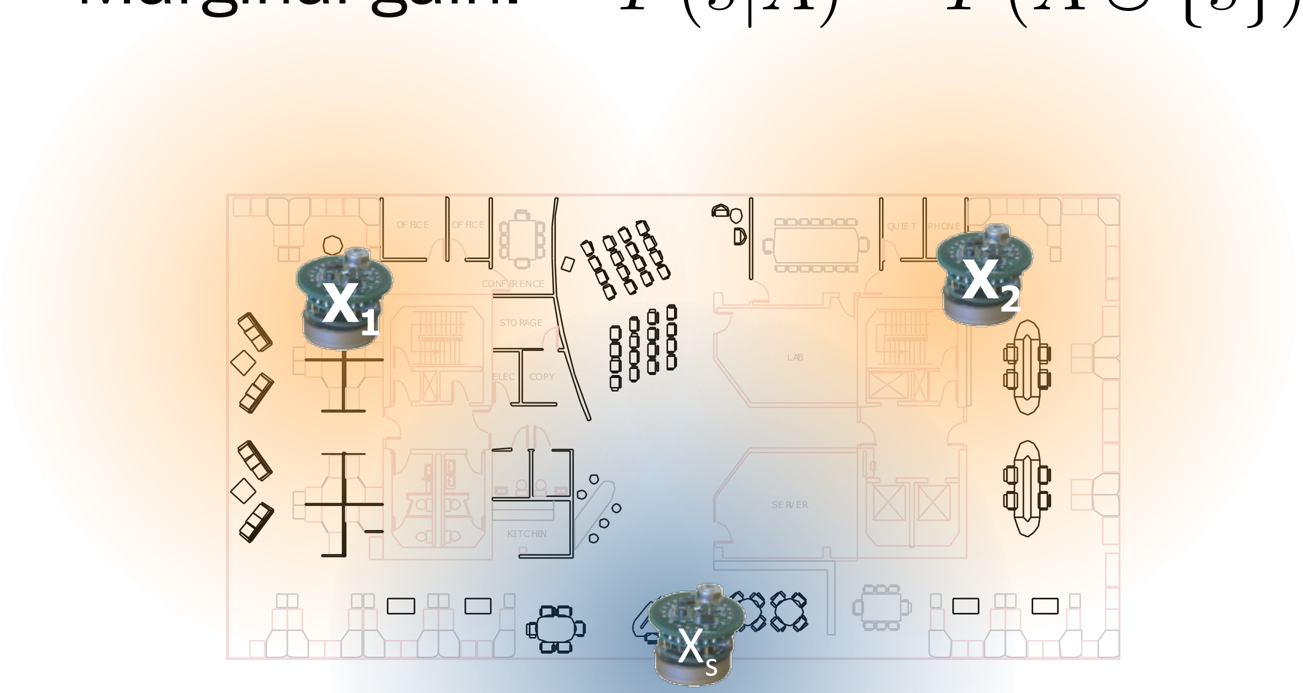
$\mathcal{V}$ = all possible locations

F(S) = information gained from locations in *S*

Marginal gain

- Given set function $F : 2^V \rightarrow \mathbb{R}$

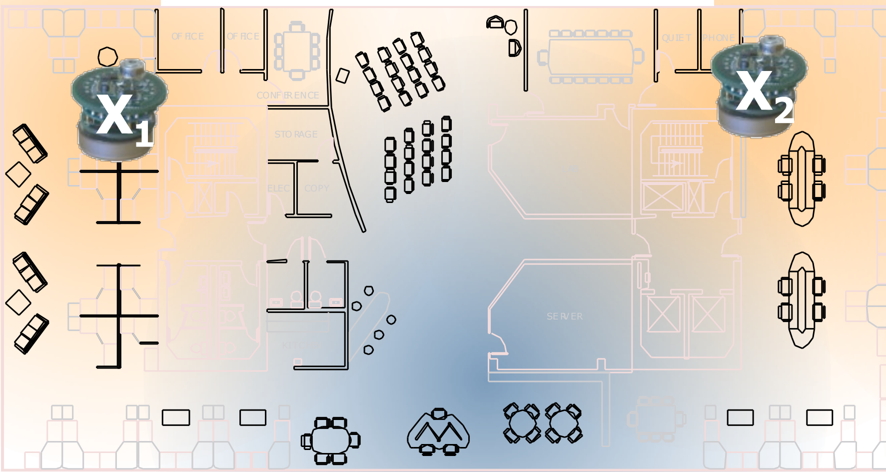
- Marginal gain: $F(s|A) = F(A \cup \{s\}) - F(A)$



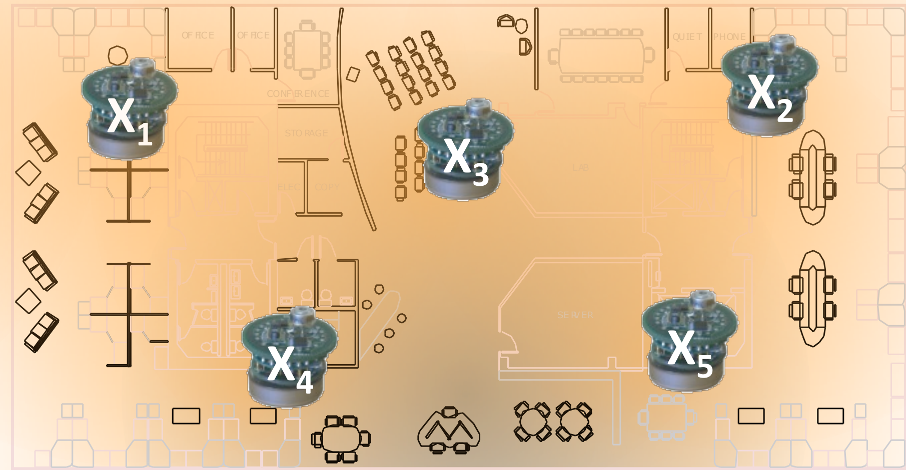
new sensor s

Diminishing marginal gains

placement A = {1,2}



placement B = {1,...,5}

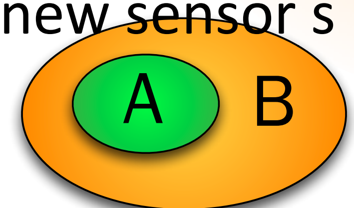


Big gain

new sensor s

small gain

+ • s



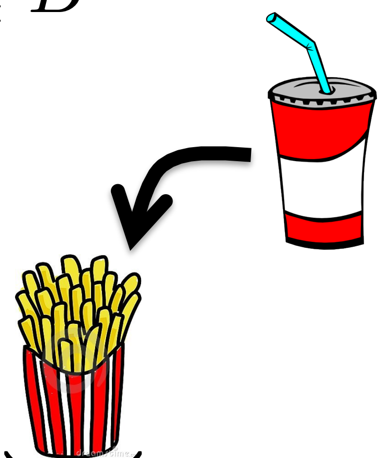
+ • s

$$A \subseteq B$$

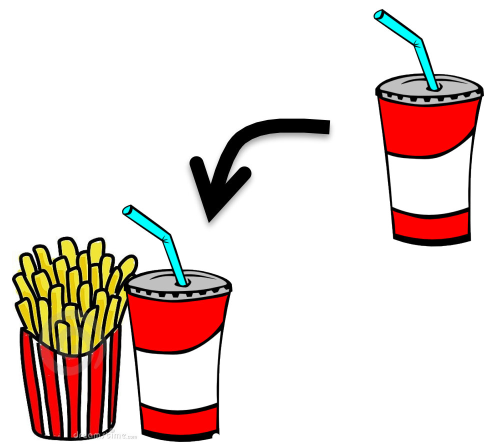
$$F(A \cup s) - F(A) \quad \geq \quad F(B \cup s) - F(B)$$

Submodularity

$A \subseteq B$



A



B

$$F(A \cup s) - F(A) \quad \geq \quad F(B \cup s) - F(B)$$

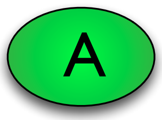
extra cost:
one drink

extra cost:
free refill ☺

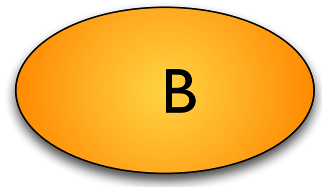
diminishing marginal costs

Submodular set functions

- Diminishing gains: for all $A \subseteq B$



+ • e

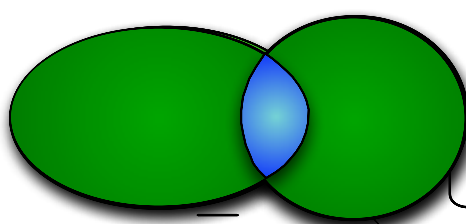


+ • e

$$F(\underline{A \cup e}) - F(A) \geq F(B \cup e) - \underline{F(B)}$$

- Union-Intersection: for all $S, T \subseteq \mathcal{V}$

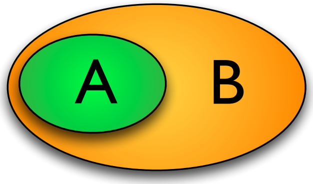
$$\underline{F(S)} + \underline{F(T)} \geq F(S \cup T) + F(S \cap T)$$



Supermodular set functions

- Submodularity: diminishing marginal gains

$$F(A \cup e) - F(A) \quad \geq \quad F(B \cup e) - F(B)$$

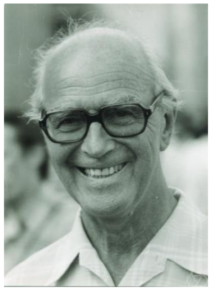


- Supermodularity: increasing marginal gains

$$F(A \cup e) - F(A) \quad \leq \quad F(B \cup e) - F(B)$$

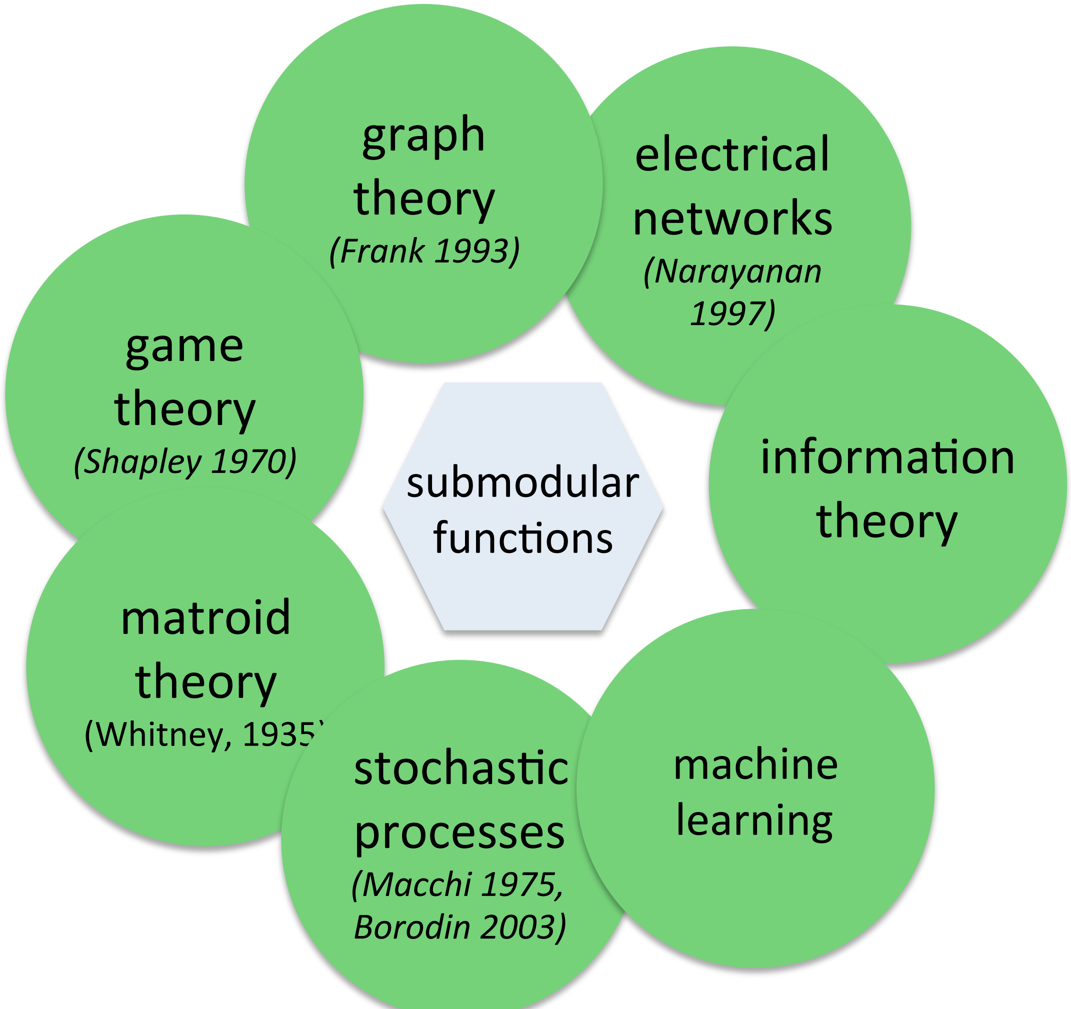
$\emptyset$

The big picture



G. Choquet

J. Edmonds



L.S. Shapley

L. Lovász

Examples

- each element e has a weight w(e)

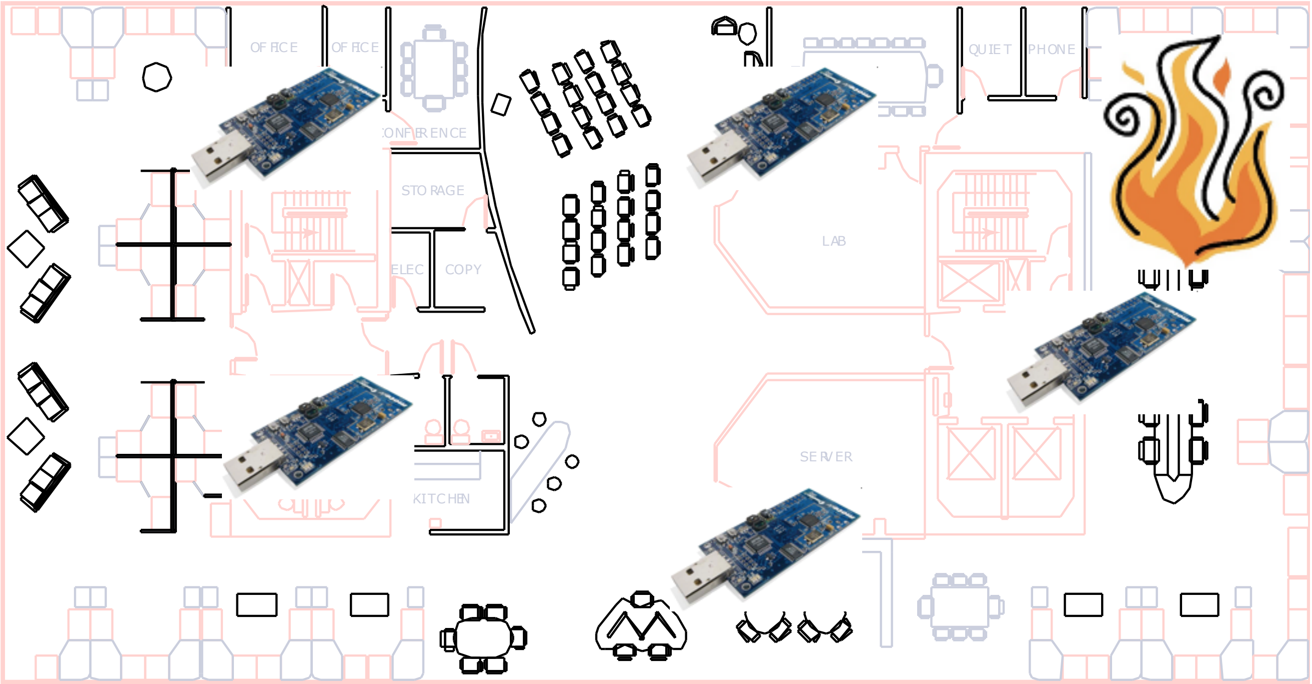
$$F(S) = \sum_{e \in S} w(e)$$

$A \subset B$

$$F(A \cup e) - F(A) = w(e) \quad = \quad F(B \cup e) - F(B) = w(e)$$

linear / modular function
always submodular!

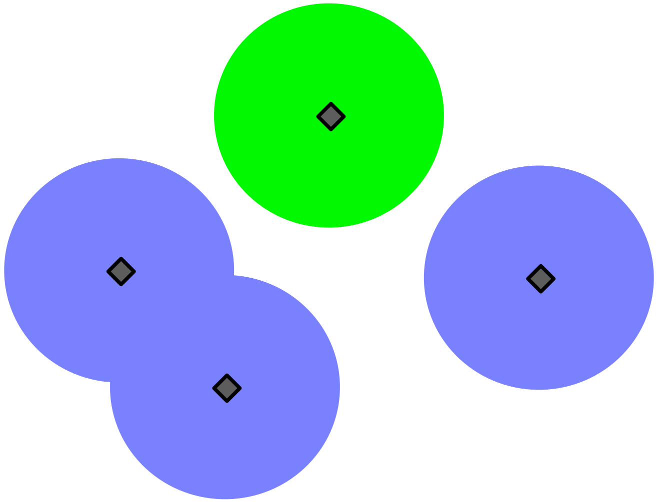
Examples

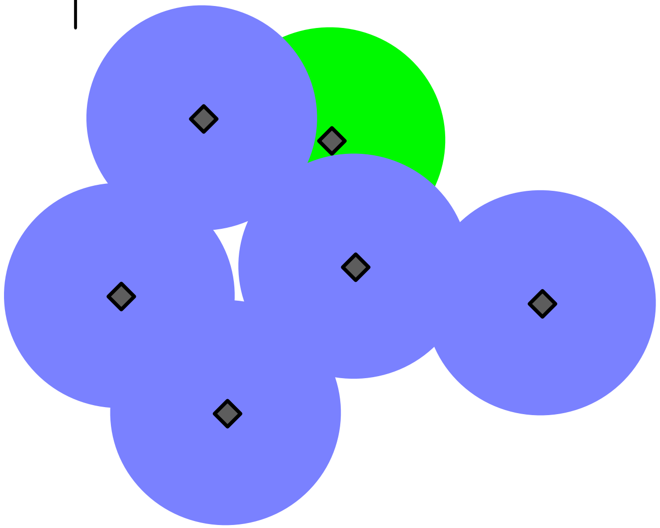


sensing:
F(S) = information gained from locations S

Example: cover

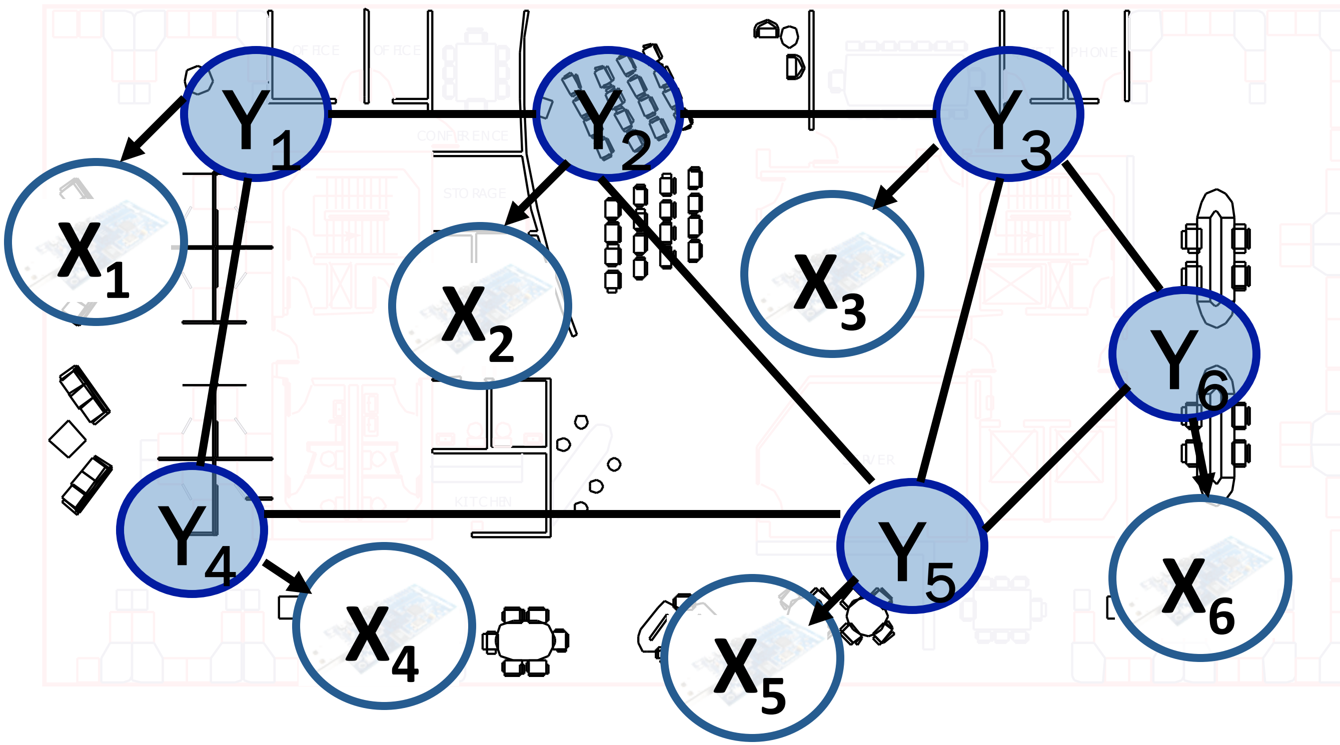
$$F(S) = \left| \bigcup_{v \in S} \mathrm{area}(v) \right|$$





$$F(A \cup v) - F(A) \geq F(B \cup v) - F(B)$$

More complex model for sensing



Y_s: temperature at location s

X_s: sensor value at location s

$X_s = Y_s + \text{noise}$

Joint probability distribution

$$P(X_1,\ldots,X_n,Y_1,\ldots,Y_n) = \underbrace{P(Y_1,\ldots,Y_n)}_{\textbf{Prior}}\ \underbrace{P(X_1,\ldots,X_n \mid Y_1,\ldots,Y_n)}_{\textbf{Likelihood}}$$

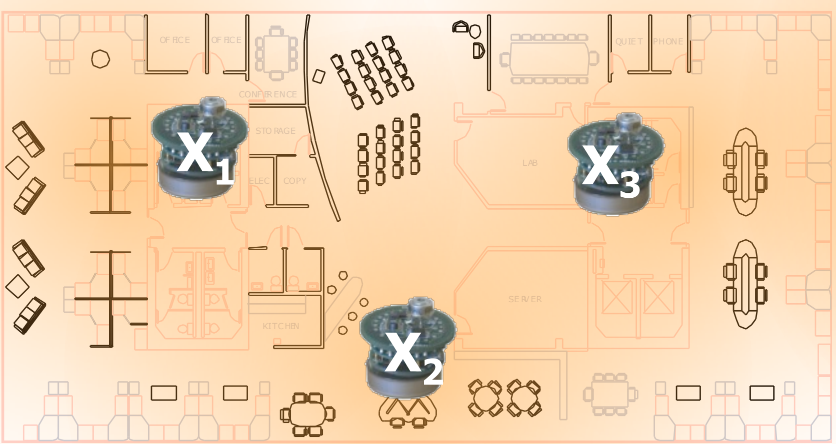
Sensor placement

Utility of having sensors at subset A of all locations

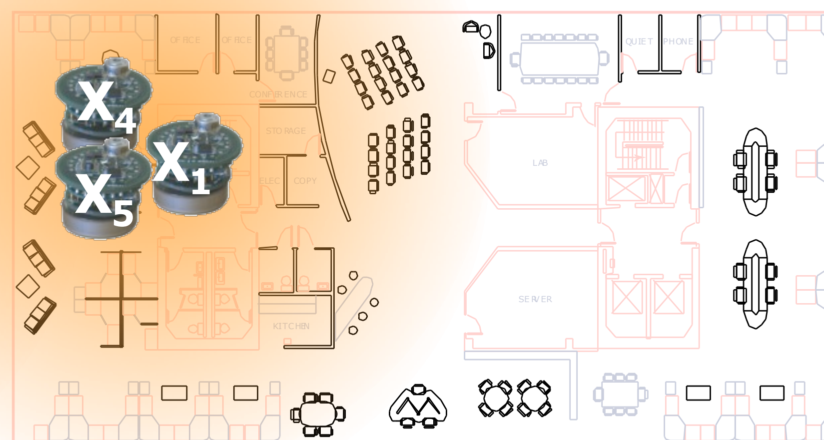
$$F(A) = H(\mathbf{Y}) - H(\mathbf{Y} \mid \mathbf{X}_A) = I(\mathbf{Y}; \mathbf{X}_A)$$

Uncertainty
about temperature Y
before sensing

Uncertainty
about temperature Y
after sensing



A={1,2,3}: High value F(A)



A={1,4,5}: Low value F(A)

Information gain

$X_1, \ldots X_n, Y_1, \ldots, Y_m$ discrete random variables

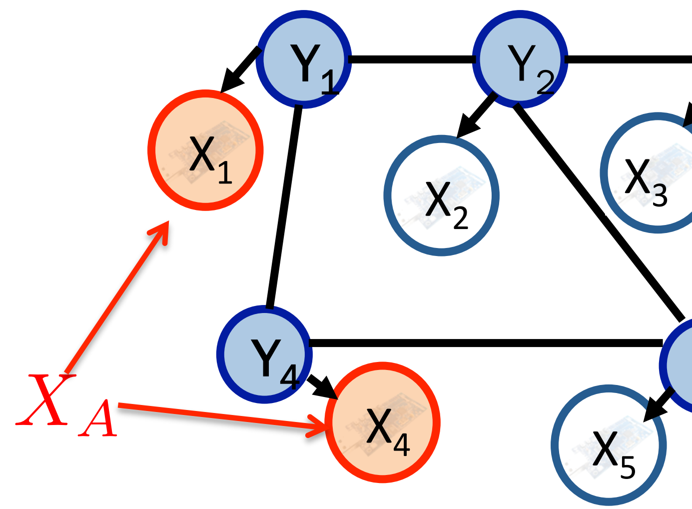
$$F(A) = I(Y; X_A) \quad = H(X_A) - H(X_A|Y)$$

modular

if all X_i, X_j conditionally independent given Y

then *F* is submodular!

(Exercise: complete the proof)



Entropy

$X_1, \ldots, X_n$ discrete random variables

$$F(S) = H(X_S) = \text{ joint entropy of variables indexed by } S$$

$$A \subset B$$

$$H(X_{A \cup e}) - H(X_A) \quad = H(X_e | X_A)$$

$$\leq H(X_e | X_B)$$ "information never hurts"

$$= H(X_{B \cup e}) - H(X_B)$$

discrete entropy is submodular!

Submodularity and independence

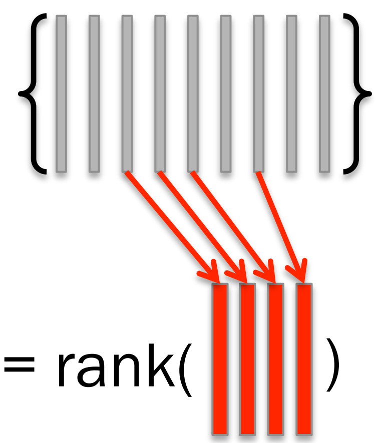
$X_1, \ldots, X_n$ discrete random variables

$X_i, i \in S$ statistically independent

⇔ H is modular/linear on S $H(X_S) = \sum_{e \in S} H(X_e)$

Similarly: linear independence

$\mathcal{V} =$



F(S) = rank( )

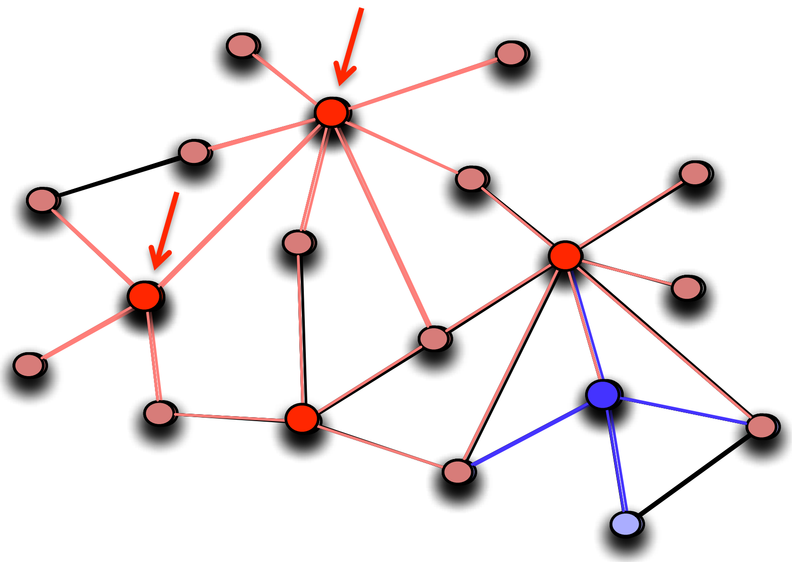
vectors in S linearly independent

⇔ F is modular/linear on S:

F(S) = |*S*|

Maximizing Influence

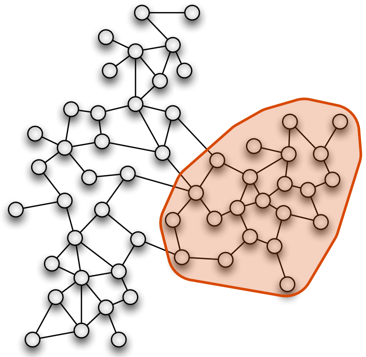
$$F(S) = \text{expected } \# \text{ infected nodes}$$



$$F(S \cup s) - F(S) \qquad \geq \qquad F(T \cup s) - F(T)$$

(Kempe, Kleinberg & Tardos 2003)

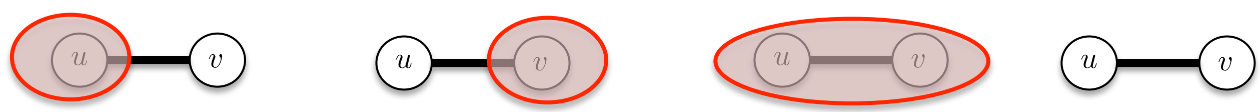
Graph cuts



$$F(S) = \sum_{u \in S, v \notin S} w_{uv}$$

- Cut for one edge:

$$F(\{u\}) \;+\; F(\{v\}) \;\geq\; F(\{u,v\}) \;+\; F(\emptyset)$$



w_{uv} w_{uv} 0 0

- cut of one edge is submodular!
- large graph: sum of edges

Useful property: sum of submodular functions is submodular

Types of submodular functions

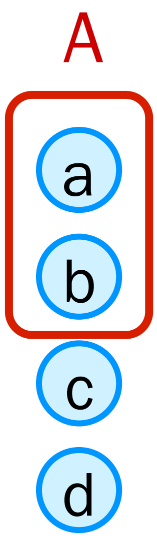
- monotone increasing and integer-valued
  - rank functions
- monotone increasing $\quad A \subseteq B \ \Rightarrow \ F(A) \leq F(B)$
  - coverage
  - entropy
  - spread
- general (non-monotone)
  - graph cuts

Sets and boolean vectors

any set function

with $|V| = n$.

$$F : 2^V \to \mathbb{R}$$



$\widehat{=}$

... is a function on binary vectors!

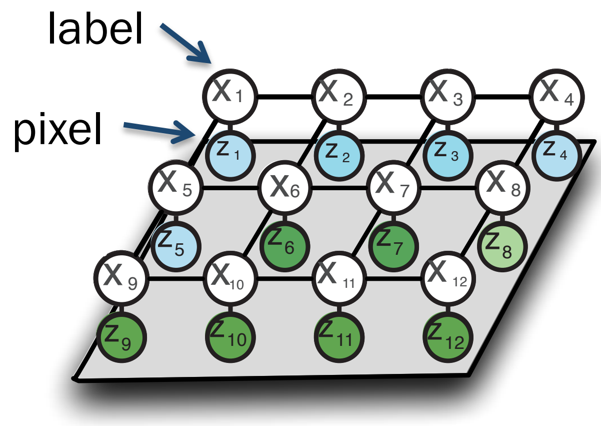
$$F : \{0,1\}^n \to \mathbb{R}$$

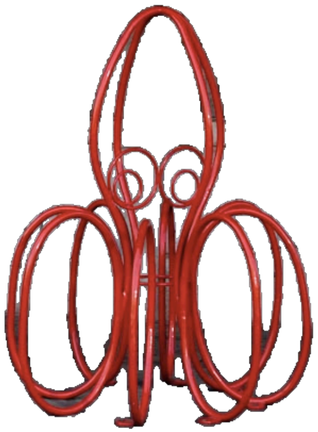
$$x = 1_A$$

$$\begin{array}{cl} \boxed{1} & a \\ \boxed{1} & b \\ \boxed{0} & c \\ \boxed{0} & d \end{array}$$

subset selection = binary labeling!

Attractive potentials





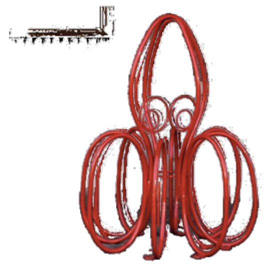
$$\max_{\mathbf{x}\in\{0,1\}^n} \quad P(\mathbf{x} \mid \mathbf{z}) \propto \exp(-E(\mathbf{x};\mathbf{z}))$$

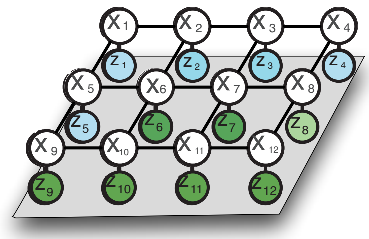
labels

pixel values

$$\Leftrightarrow \quad \min_{\mathbf{x}\in\{0,1\}^n} \quad E(\mathbf{x};\mathbf{z})$$

Attractive potentials



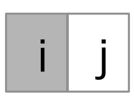


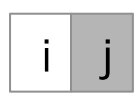
$$P(\mathbf{x} \mid \mathbf{z}) \propto \exp(-E(\mathbf{x};\mathbf{z}))$$

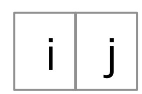
$$E(\mathbf{x};\mathbf{z}) = \sum_i E_i(x_i) + \sum_{ij} E_{ij}(x_i, x_j)$$

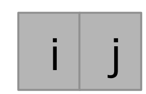
spatial coherence:

$$E_{ij}(1,0) + E_{ij}(0,1) \;\geq\; E_{ij}(0,0) + E_{ij}(1,1)$$





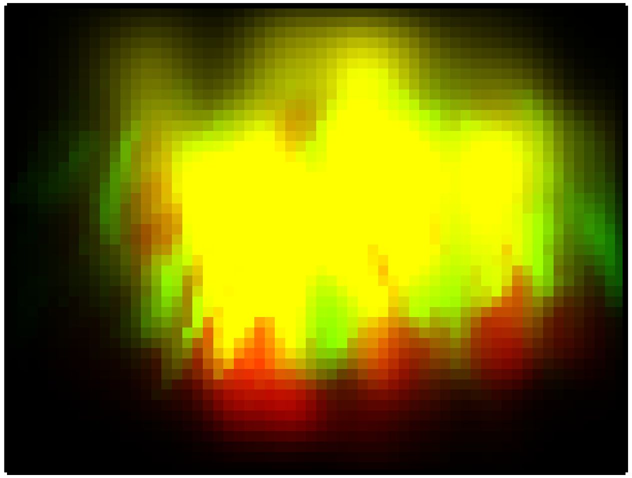




$S = \{i\}$ $\quad T = \{j\}$ $\quad S \cap T = \emptyset$ $\quad S \cup T$

$$\boxed{F(S) \;+\; F(T) \;\geq\; F(S \cup T) \;+\; F(S \cap T)}$$

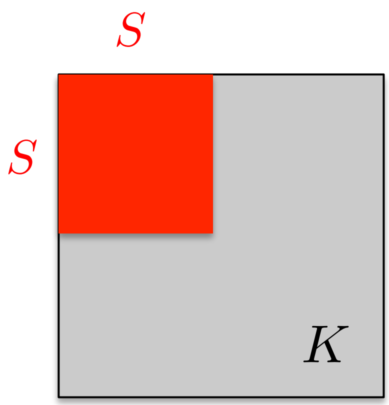
Diversity priors



$$P(S \mid \text{data}) \propto P(S)\, P(\text{data} \mid S)$$

"spread out"

Determinantal point processes



- normalized similarity matrix K
- sample *Y*:

$$P(S \subseteq Y) = \det(K_S)$$

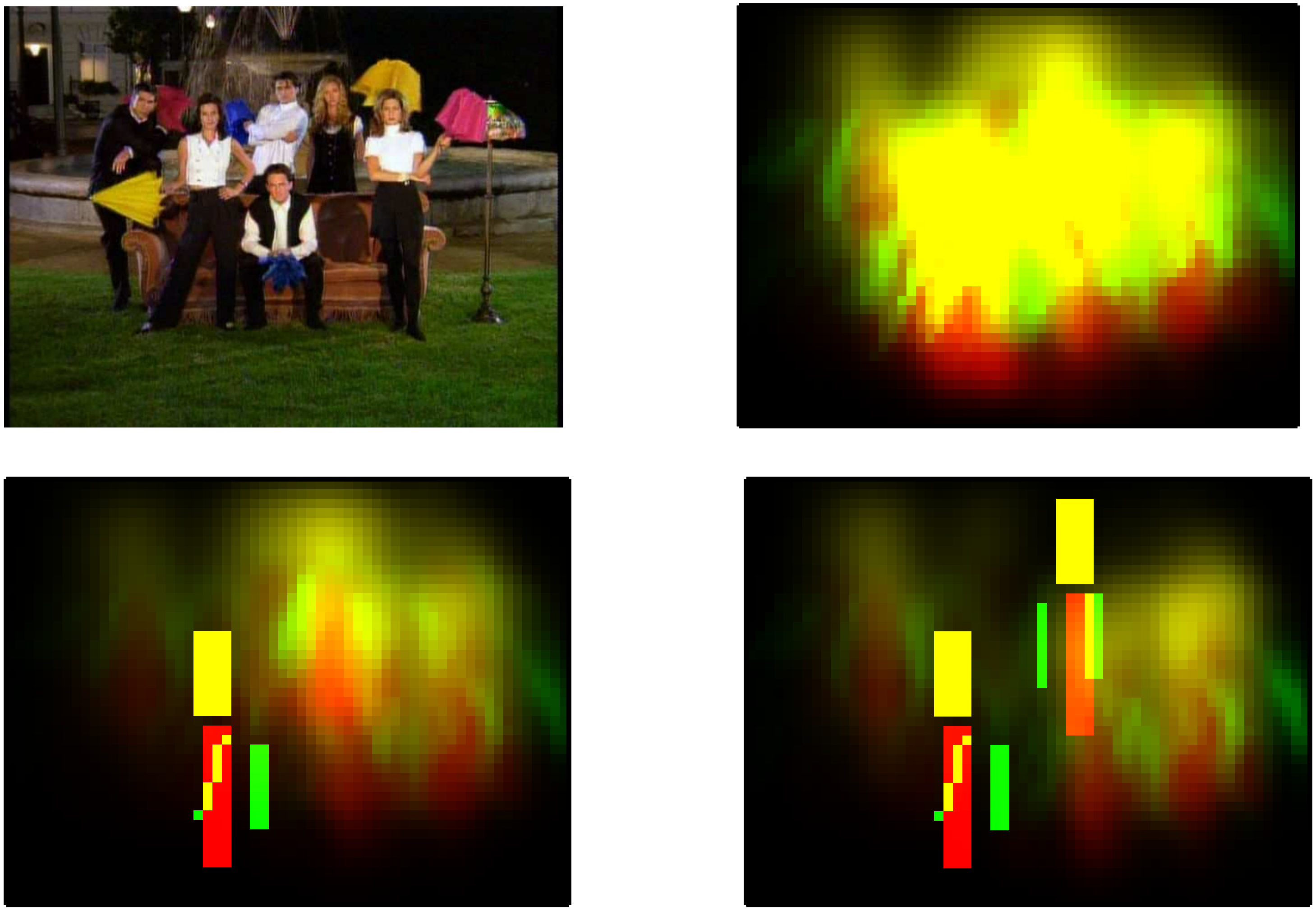
$$P(e_i \in Y) = K_{ii}$$

$$P(e_i, e_j \in Y) = K_{ii}K_{jj} - K_{ij}^2$$
$$= P(e_i \in Y)P(e_j \in Y) - K_{ij}^2$$

repulsion

$F(S) = \log\det(K_S)$ is submodular

Diversity priors



(Kulesza & Taskar 10)

Submodularity and machine learning

distributions over labels, sets
log-submodular/
supermodular probability
e.g. “attractive” graphical models,
determinantal point processes

submodularity
in machine
learning!

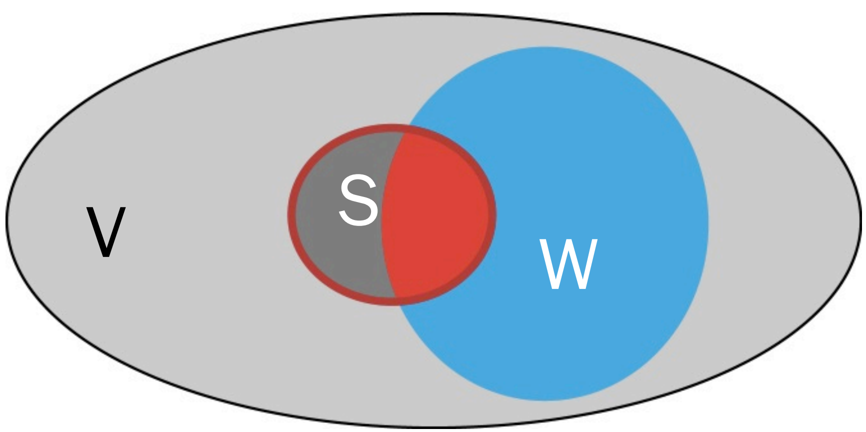
diffusion processes,
covering, rank,
connectivity,
entropy,
economies of scale,
summarization, …
submodular
phenomena

(convex) regularization
submodularity: “discrete
convexity”
e.g. combinatorial sparse estimation

Closedness properties

$F(S)$ submodular on V. The following are submodular:

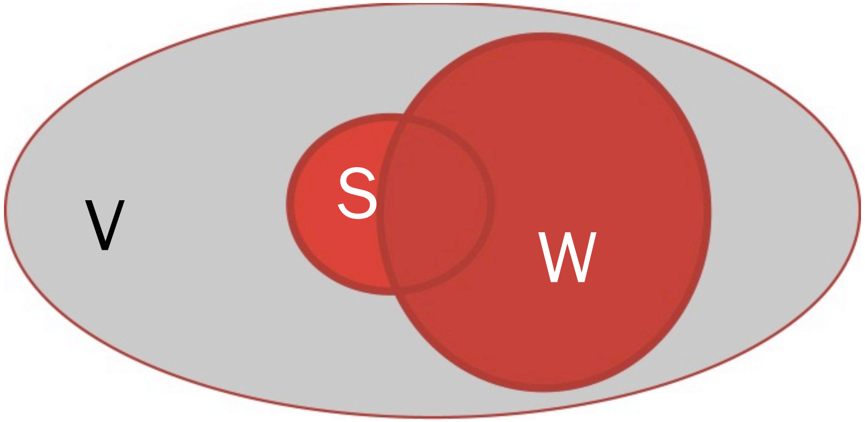
- Restriction: $F'(S) = F(S \cap W)$



Closedness properties

$F(S)$ submodular on V. The following are submodular:

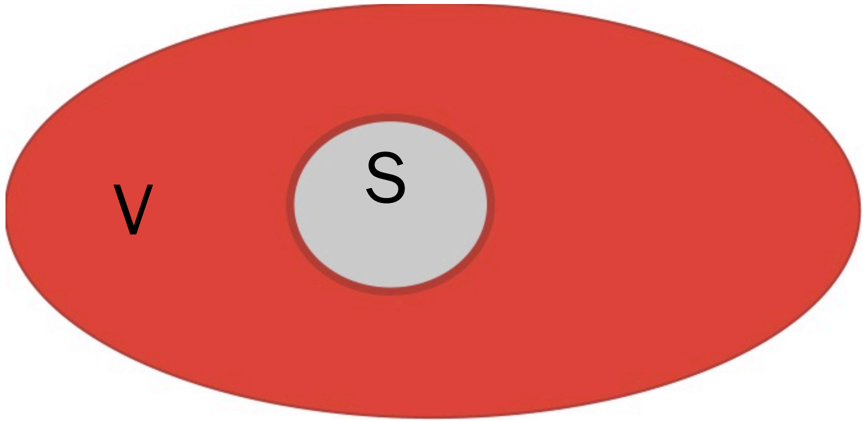
- Restriction: $F'(S) = F(S \cap W)$
- Conditioning: $F'(S) = F(S \cup W)$



Closedness properties

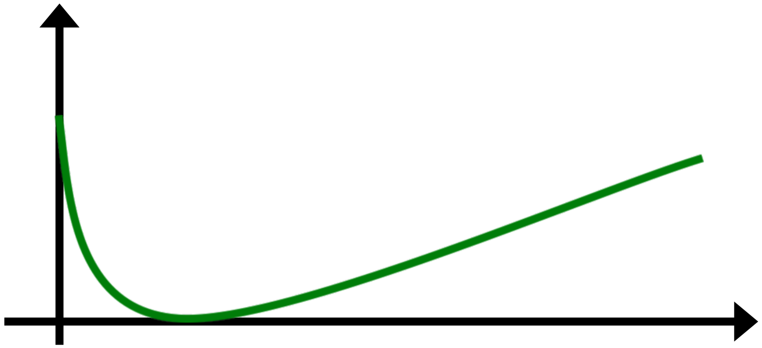
$F(S)$ submodular on V. The following are submodular:

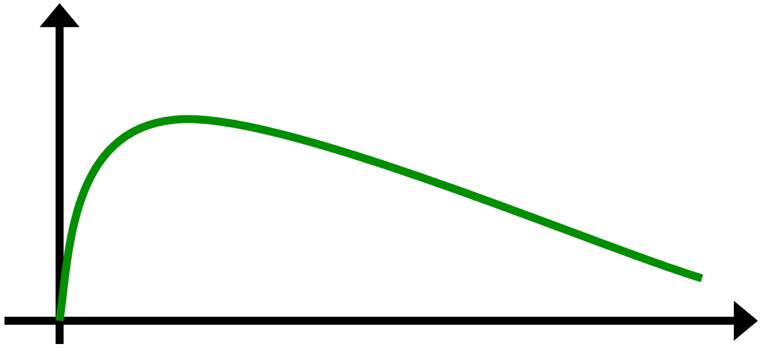
- Restriction: $F'(S) = F(S \cap W)$
- Conditioning: $F'(S) = F(S \cup W)$
- Reflection: $F'(S) = F(V \setminus S)$



Submodularity ...

discrete convexity ....





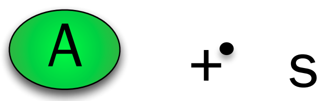
... or concavity?

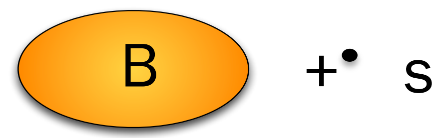
Concave aspects

- submodularity:

$$A \subseteq B, \quad s \notin B:$$

$$F(A \cup s) - F(A) \quad \geq \quad F(B \cup s) - F(B)$$

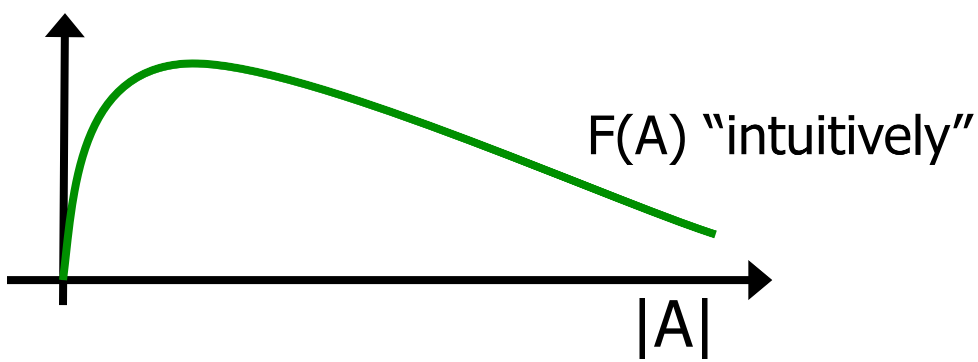




- concavity:

$$a \leq b, \quad s > 0:$$

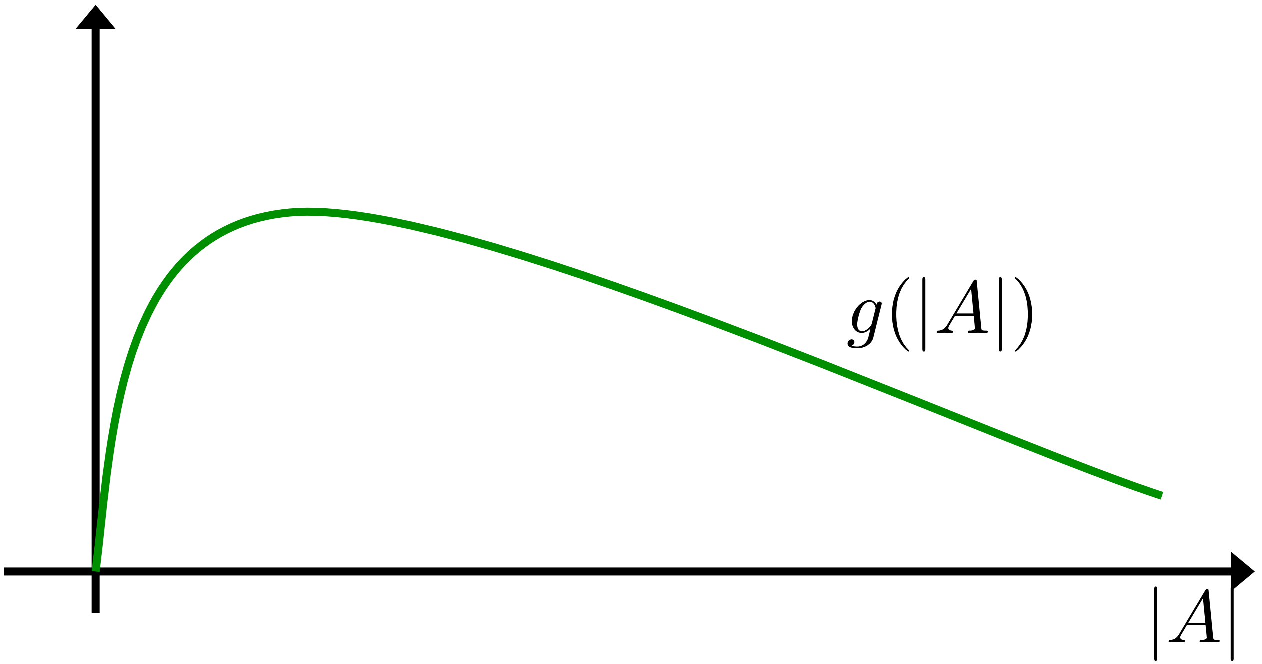
$$f(a+s) - f(a) \quad \geq \quad f(b+s) - f(b)$$



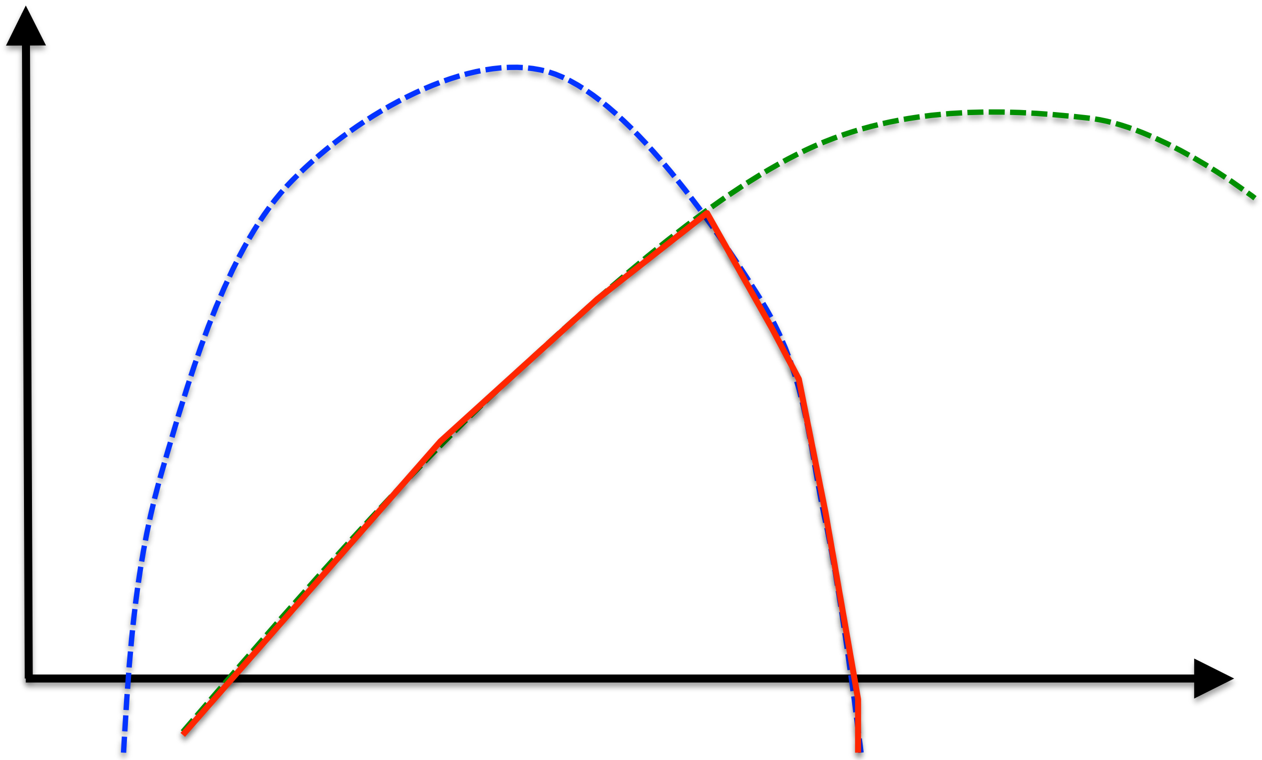
Submodularity and concavity

- suppose $g : \mathbb{N} \to \mathbb{R}$ and $F(A) = g(|A|)$

$F(A)$ submodular *if and only if ...* g is concave



Minimum of concave functions



Minimum of submodular functions

$$F(A) = \min\{\ F_1(A), F_2(A)\ \}$$ submodular?

$$\overset{0}{F(A)} + \overset{0}{F(B)} \ \geq\ \overset{1}{F(A \cup B)} + \overset{0}{F(A \cap B)}\ ?$$

		$F_1(A)$	$F_2(A)$
$A \cap B$	{}	0	0
A	{a}	1	0
B	{b}	0	1
$A \cup B$	{a,b}	1	1

$A \cap B$

A

B

$A \cup B$

min(F_1,F_2) not submodular in general!

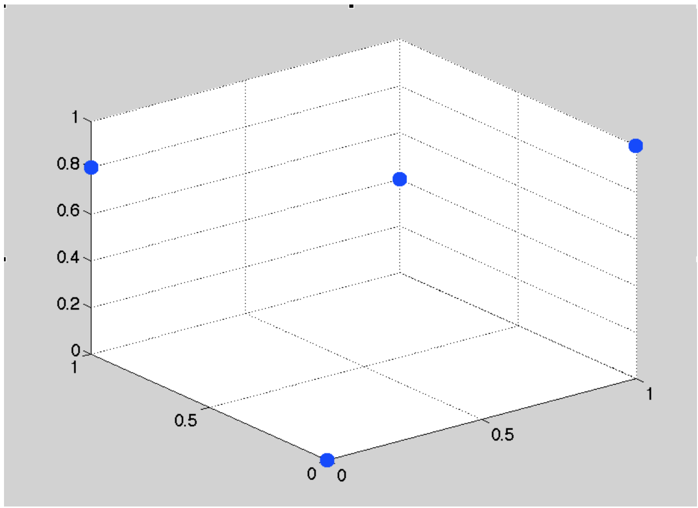
Convex functions *(Lovász, 1983)*

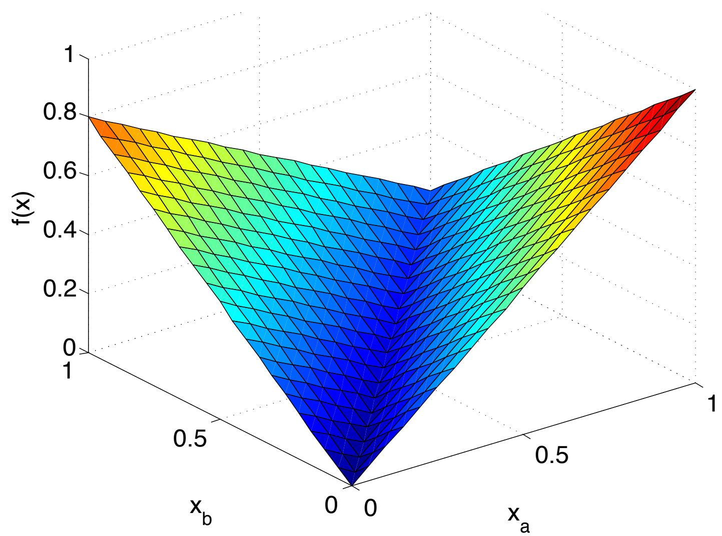
- "occur in many models in economy, engineering and other sciences", "often the only nontrivial property that can be stated in general"
- preserved under many operations and transformations: larger effective range of results
- sufficient structure for a "mathematically beautiful and practically useful theory"
- efficient minimization

"It is less apparent, but we claim and hope to prove to a certain extent, that a similar role is played in discrete optimization by *submodular set-functions*" [...]
they share the above four properties.

Convex aspects

- convex extension
- duality results
- poly-time minimization

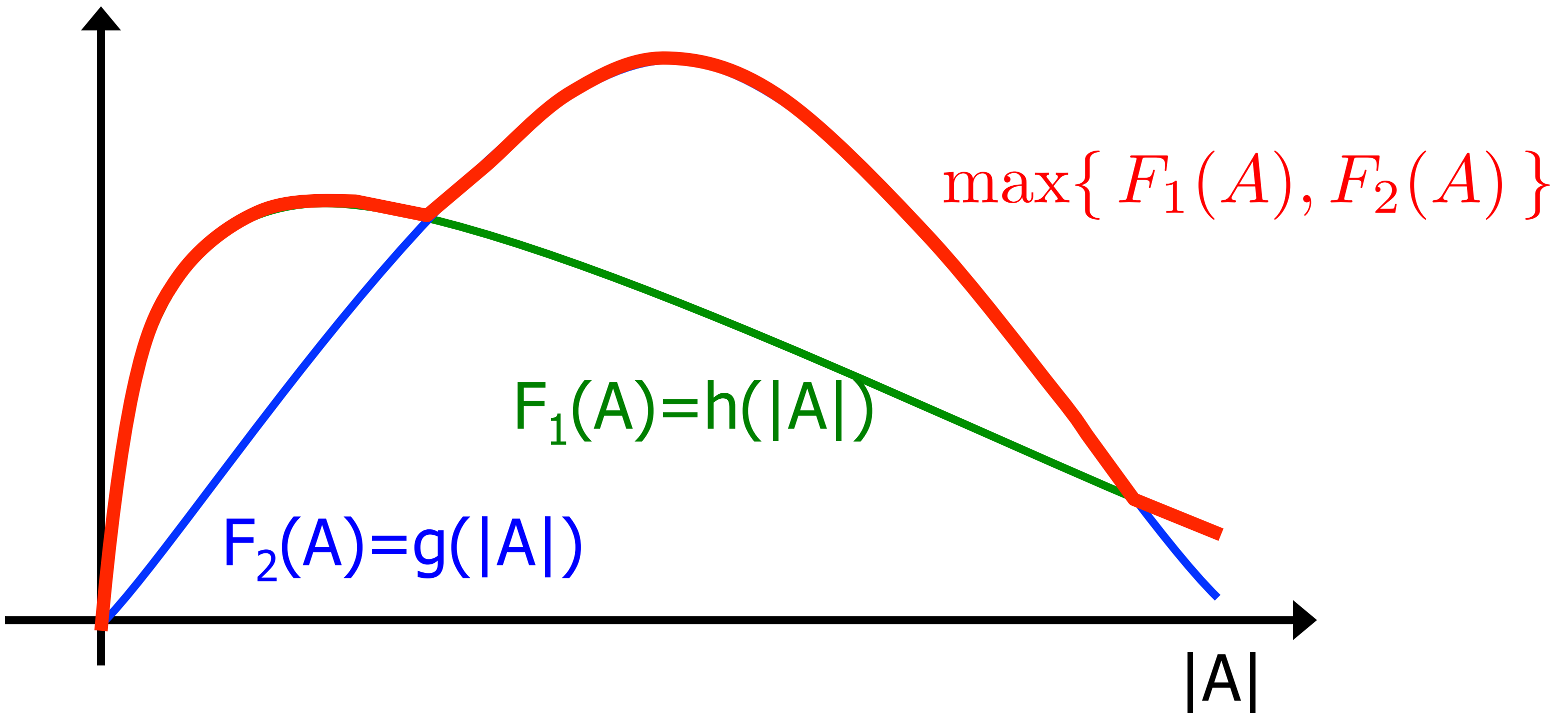




Maximum of submodular functions

- $F_1(A), F_2(A)$ submodular. What about

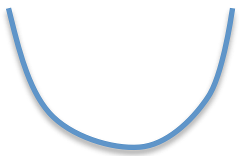
$$F(A) = \max\{\, F_1(A), F_2(A) \,\}\ ?$$

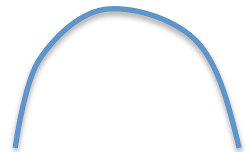


$\max\{\, F_1, F_2 \,\}$ not submodular in general!

Roadmap

convex …





… and concave aspects!

- Submodular set functions
  - definition & basic properties
  - links to convexity
  - special polyhedra
- Minimizing submodular functions
- Maximizing submodular functions

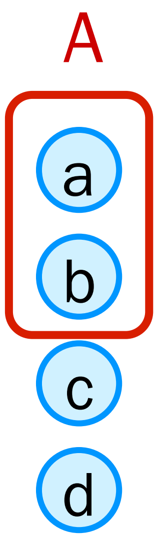
Submodularity and convexity

any set function

with $|V| = n$.

... is a function on binary vectors!

$$F : 2^V \to \mathbb{R} \qquad\qquad F : \{0,1\}^n \to \mathbb{R}$$



$$\hat{=}$$

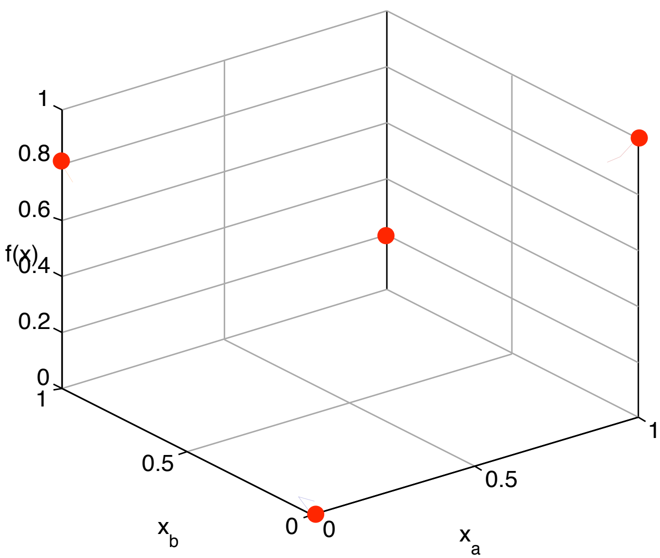
$$x = 1_A$$

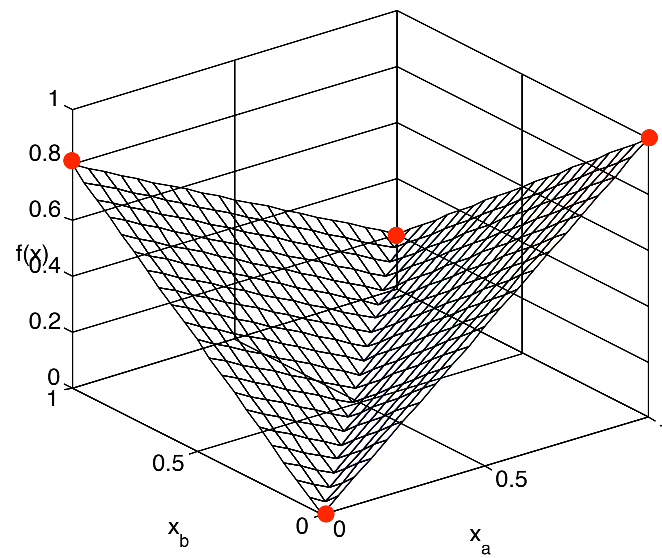
$$\begin{array}{cc} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} & \begin{array}{c} a \\ b \\ c \\ d \end{array} \end{array}$$

subset selection = binary labeling!

Relaxation: idea

$$\min_{x \in \{0,1\}^n} F(x) \longrightarrow \min_{x \in [0,1]^n} f(x)$$





A relaxation (extension)

have

$$F : \{0,1\}^n \to \mathbb{R}$$

want: extension

$$f : \mathbb{R}^n_+ \to \mathbb{R}$$

$$\underset{x}{\begin{bmatrix} 0.5 \\ 1.0 \\ 0.5 \\ 0.2 \\ 0.2 \end{bmatrix}} = (1.0 - 0.5) \overset{S_1}{\begin{bmatrix} 0 \\ 1.0 \\ 0 \\ 0 \\ 0 \end{bmatrix}} + (0.5 - 0.2) \overset{S_2}{\begin{bmatrix} 1.0 \\ 1.0 \\ 1.0 \\ 0 \\ 0 \end{bmatrix}} + (0.2) \overset{S_3}{\begin{bmatrix} 1.0 \\ 1.0 \\ 1.0 \\ 1.0 \\ 1.0 \end{bmatrix}}$$

$$x = \sum_{i=1}^{k} \alpha_i \mathbf{1}_{S_i}$$

$$\boxed{f(x) = \sum_{i=1}^{k} \alpha_i F(S_i)}$$

Examples

$$f(x) = \sum_{i=1}^{k} \alpha_i \, F(S_i)$$

$$\begin{bmatrix} 0.5 \\ 1.0 \end{bmatrix} = 0.5 \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix} + 0.5 \begin{bmatrix} 0 \\ 1.0 \end{bmatrix}$$

1.0 - 0.5

- truncation

$$F(S) = \max\{|S|, 1\}$$

$$f(x) = 0.5 + 0.5 \quad = \max_i \; x_i$$

- cut function

$$F(S) = \begin{cases} 1 & \text{if } S = \{1\}, \{2\} \\ 0 & \text{if } S = \emptyset, \{1, 2\} \end{cases}$$

$$f(x) = 0.5 \cdot 0 + (1 - 0.5) \cdot 1$$

$$= |x_1 - x_2|$$

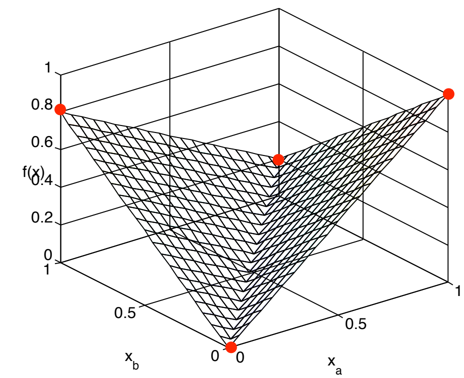
"total variation"!

Alternative characterization

$$f(x) = \sum_{i=1}^{k} \alpha_i F(S_i)$$

if *F* is submodular, this is equivalent to:

$$\boxed{f(x) = \max_{y \in \mathcal{B}_F} y^\top x}$$



Theorem *(Lovász, 1983)*
Lovász extension is convex $\Leftrightarrow$ F is submodular.

Theorem *(Lovász, 1983)*
Lovász extension is convex $\Leftrightarrow$ F is submodular.

$$f(x) = \sum_{i=1}^{k} \alpha_i \, F(S_i)$$

$\Leftarrow$ If F is submodular, then f is the max of linear functions.

$\Rightarrow$ Let $A, B \subseteq \mathcal{V}$.

$$\underset{1_A}{\begin{bmatrix}1\\1\\0\\0\end{bmatrix}} + \underset{1_B}{\begin{bmatrix}0\\1\\1\\0\end{bmatrix}} = \begin{bmatrix}1\\2\\1\\0\end{bmatrix} = \underset{1_{A\cup B}}{\begin{bmatrix}1\\1\\1\\0\end{bmatrix}} + \underset{1_{A\cap B}}{\begin{bmatrix}0\\1\\0\\0\end{bmatrix}}$$

convexity and positive homogeneity implies:

$$f(1_A + 1_B) \quad \leq \quad f(1_A) + f(1_B)$$

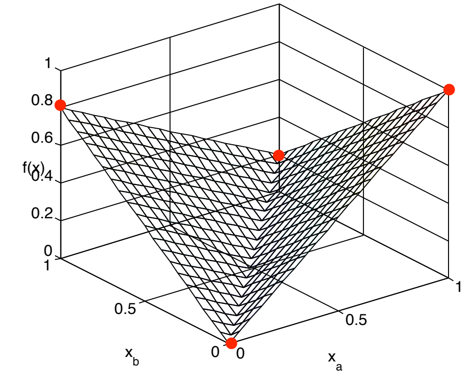
Exercise: this implies that F is submodular

Alternative characterization

$$f(x) = \sum_{i=1}^{k} \alpha_i F(S_i)$$

if *F* is submodular, this is equivalent to:

$$\boxed{f(x) = \max_{y \in \mathcal{B}_F} y^\top x}$$



Theorem *(Lovász, 1983)*
Lovász extension is convex $\Leftrightarrow$ F is submodular.

Submodular polyhedra

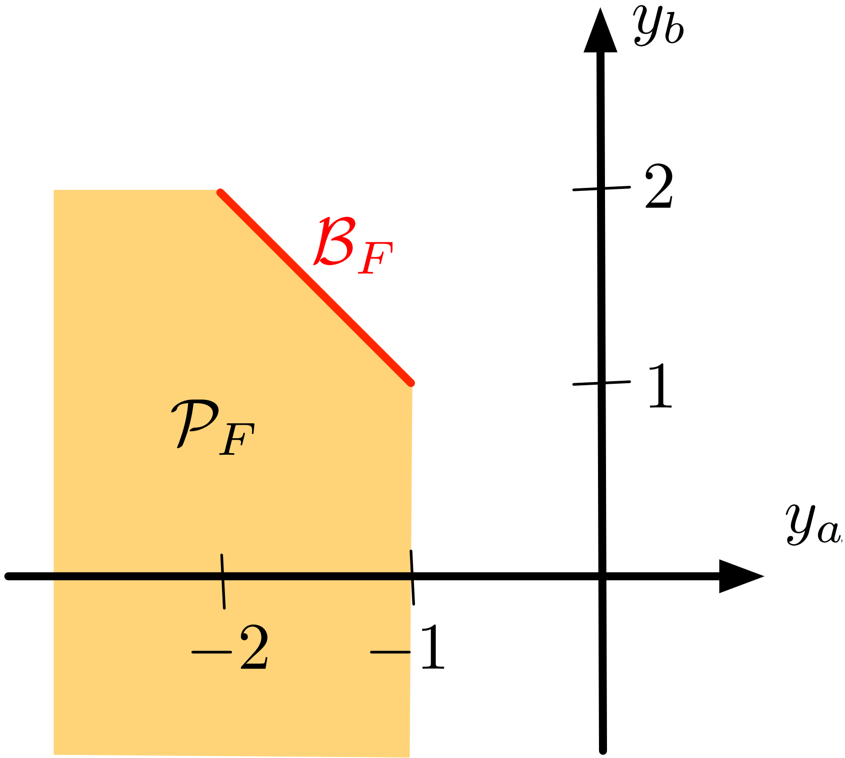
submodular polyhedron:

$$\mathcal{P}_F = \{y \in \mathbb{R}^n \mid y(A) \leq F(A) \text{ for all } A \subseteq \mathcal{V}\}$$

$$y(A) = \sum_{a \in A} y_a$$

Base polytope

$$\mathcal{B}_F = \{y \in \mathcal{P}_F \mid y(\mathcal{V}) = F(\mathcal{V})\}$$

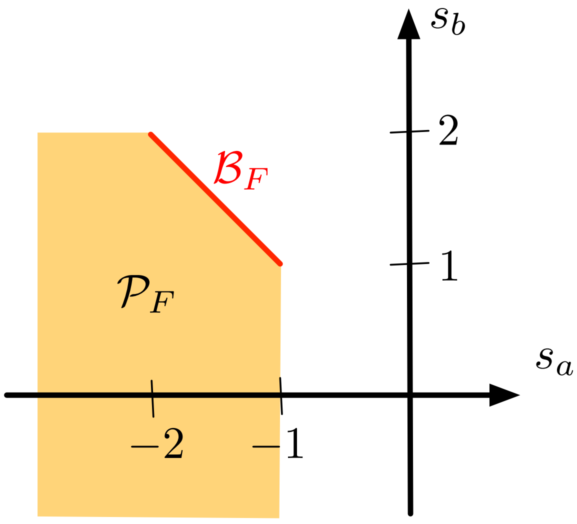


A	$F(A)$
$\emptyset$	0
a	-1
b	2
$\{a, b\}$	0

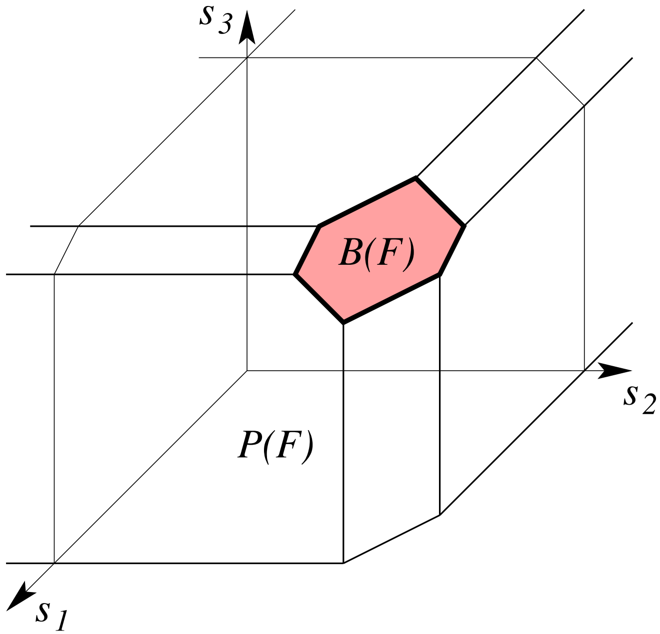
Base polytopes

Base polytope $\mathcal{B}_F = \{y \in \mathcal{P}_F \mid y(\mathcal{V}) = F(\mathcal{V})\}$

2D (2 elements)



3D (3 elements)

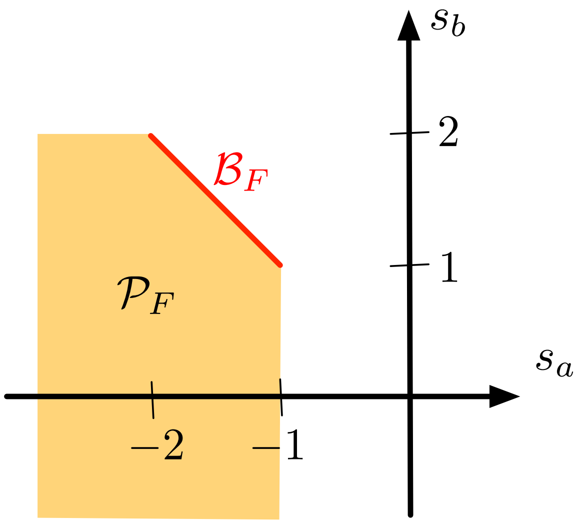


Base polytope

$$\mathcal{B}_F = \Big\{ y \in \mathbb{R}^n \mid \sum_{a \in S} y_a \leq F(S) \quad \text{for all } S \subseteq \mathcal{V}$$
$$y(\mathcal{V}) = F(\mathcal{V}) \Big\}$$

exponentially
many constraints!

$$\boxed{f(x) = \max_{y \in \mathcal{B}_F} \; y^\top x}$$



Edmonds 1970: “magic”
compute argmax in *O(n log n)* ☺

basis for submodular minimization!

Optimization over base polytope

$$\mathcal{B}_F = \Big\{ y \in \mathbb{R}^n \mid \sum_{a \in S} y_a \le F(S) \qquad y(\mathcal{V}) = F(\mathcal{V}) \Big\}$$

$$\boxed{f(x) = \max_{y \in \mathcal{B}_F} \; y^\top x}$$

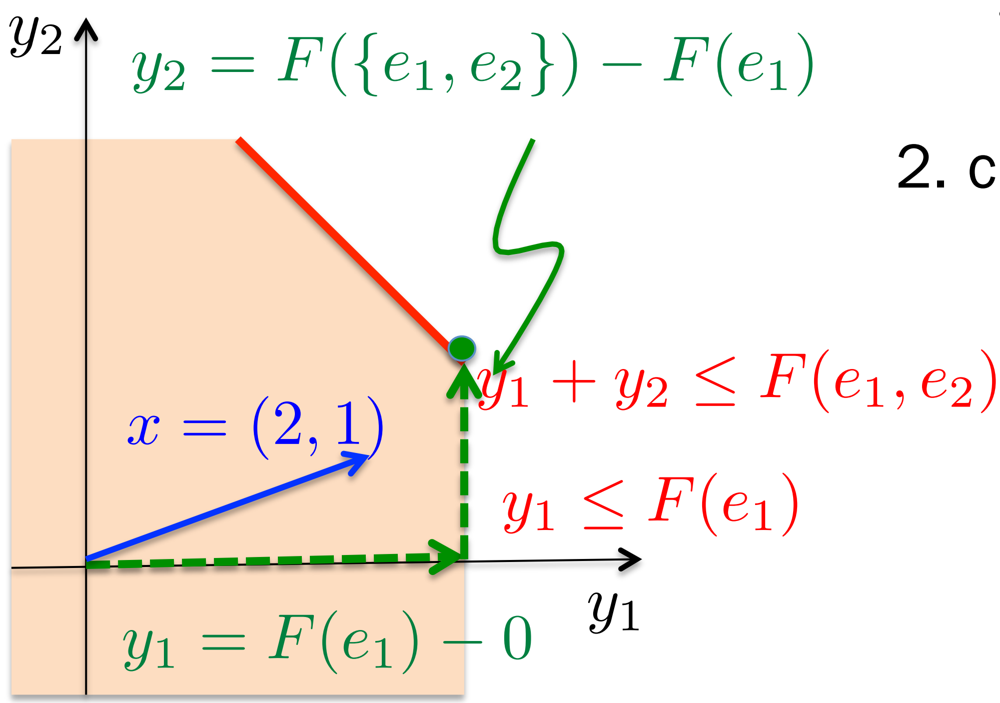
Edmonds' greedy algorithm:

1. sort

$$x_{\pi(1)} \ge x_{\pi(2)} \ge \dots \ge x_{\pi(n)}$$

2. chain of sets

$$S_0 = \emptyset,$$
$$S_1 = \{\, \pi(1) \,\} \dots$$
$$S_i = \{\, \pi(1), \dots, \pi(i) \,\}$$



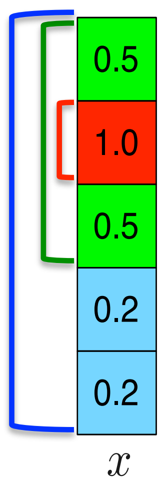
Base polytope

$$f(x) = \sum_{i=1}^{k} \alpha_i \, F(S_i)$$

$$f(x) = \max_{y \in \mathcal{B}_F} \; y^\top x$$

Remarks:

- chain of sets same as before



Edmonds' greedy algorithm:

1. sort

$$x_{\pi(1)} \geq x_{\pi(2)} \geq \cdots \geq x_{\pi(n)}$$

2. chain of sets

$$S_0 = \emptyset, \; S_i = \{\pi(1), \ldots, \pi(i)\}$$

3. assign values

$$y_{\pi(i)} = F(S_i) - F(S_{i-1})$$

Re-computing our examples

$$\underset{x}{\begin{bmatrix}0.5\\1.0\end{bmatrix}} = 0.5\begin{bmatrix}1.0\\1.0\end{bmatrix} + 0.5\begin{bmatrix}0\\1.0\end{bmatrix}$$

$$F(S) = \max\{|S|, 1\}$$

sort: $x_2 \geq x_1 \quad \Rightarrow \quad S_1 = \{2\},\ S_2 = \{2, 1\}$

$$y_2 = F(2) = 1$$

$$y_1 = F(2,1) - F(2) = 1 - 1 = 0$$

$$f(x) = y^\top x = 1 \cdot x_1 + 0 \cdot x_2 = \max_i x_i$$

in general: $F(S_i) - F(S_{i-1}) > 0$ only for i=1!

$$\Rightarrow \quad f(x) = \max_i x_i$$

Re-computing our examples

$$x = \begin{bmatrix} 0.5 \\ 1.0 \end{bmatrix} = 0.5 \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix} + 0.5 \begin{bmatrix} 0 \\ 1.0 \end{bmatrix}$$

$$F(S) = \begin{cases} 1 & \text{if } S = \{1\},\ \{2\} \\ 0 & \text{if } S = \emptyset,\ \{1,2\} \end{cases}$$

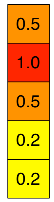
sort: $x_2 \geq x_1 \quad \Rightarrow \quad S_1 = \{2\},\ S_2 = \{2,1\}$

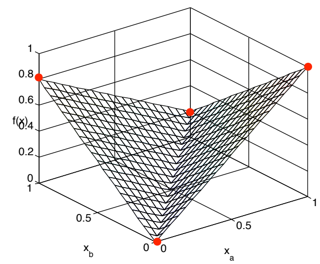
$$y_2 = F(2) = 1$$

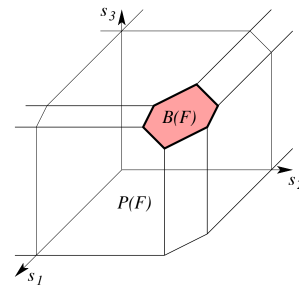
$$y_1 = F(2,1) - F(2) = 0 - 1 = -1$$

$$f(x) = y^\top x = -0.5 + 1 = |x_1 - x_2|$$

Convex relaxation







$$\min_{S \subseteq \mathcal{V}} F(S) = \min_{x \in \{0,1\}^n} F(x) \longrightarrow \min_{x \in [0,1]^n} f(x)$$

1. relaxation: convex optimization (non-smooth)

2. relaxation is exact!
   ➔ submodular minimization in polynomial time!
   (Grötschel, Lovász, Schrijver 1981)

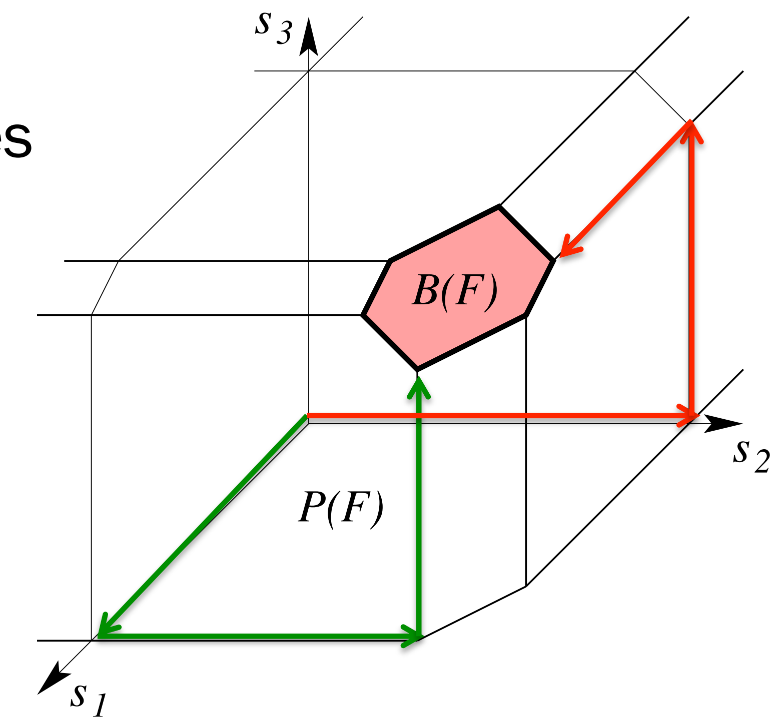
Amazing base polytopes

- linear optimization = Edmonds' greedy algorithm: each vertex determined by a permutation!

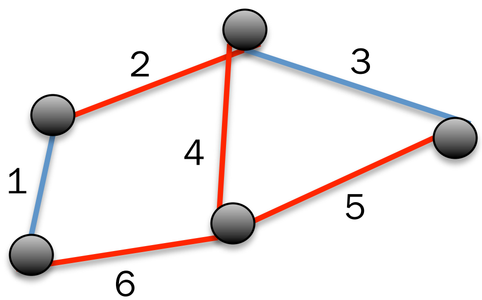
$$f(x) = \max_{y \in \mathcal{B}_F} y^\top x$$

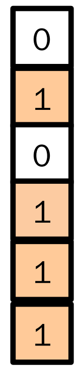
Base polytopes almost everywhere:

- cores of games *(Shapley)*
- information theory: achievable rates for lossless coding of correlated sources *(Slepian-Wolf, Cover, Fujishige)*
- matroids



Base polytopes and spanning trees





ground set: all edges

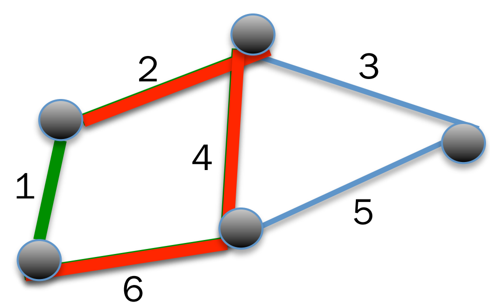
indicator vectors 1_T of all spanning trees T

$\mathcal{B} =$ convex hull of all tree indicator vectors

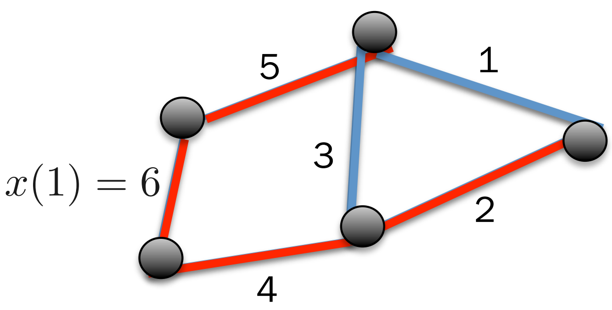
is a base polytope

What is the submodular function?

$$F(S) = \text{size of largest tree within } S$$
$$= \max\{|T| : T \text{ is a tree and } T \subseteq S\}$$



Greedy algorithm for trees



$$F(S) = \text{size of largest tree within } S$$

$$\max_{y \in \mathcal{B}} \; y^\top x$$

vector x: weight for each edge

greedy algorithm: y indicator vector of a tree

- go through edges in order of their weight

$$e_1 \quad e_2 \quad e_3 \quad e_4 \quad e_5 \quad e_6$$

- if edge does not complete a cycle:

$$F(S_i) - F(S_{i-1}) = 1 \qquad y_i = 1 \qquad \text{“pick i”}$$

if it does:

$$F(S_i) - F(S_{i-1}) = 0 \qquad y_i = 0 \qquad \text{“don’t pick i”}$$

finds the maximum weight spanning tree (Kruskal’s algorithm)

General: Matroids

Matroid $\mathcal{M} = (\mathcal{V}, \mathcal{I})$

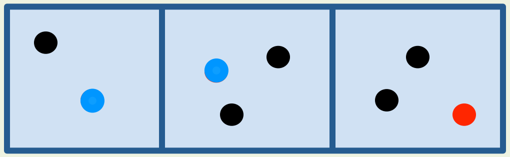
- ground set $\mathcal{V}$
- family of independent sets $\mathcal{I}$

Matroids (semi-formally)

S is independent (= feasible) if …

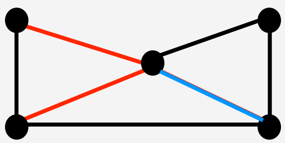
… $|S| \leq k$

Uniform matroid



… S contains at most one element from each square

Partition matroid



… S contains no cycles

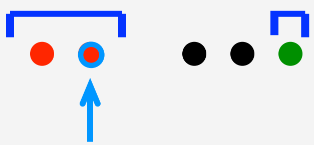
Graphic matroid

matroid properties:

- S independent ➔ $T \subseteq S$ also independent

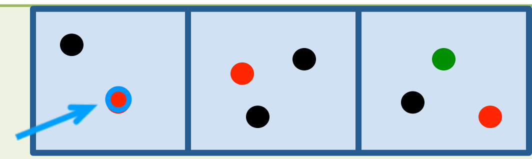
Matroids

S is independent (=feasible) if …



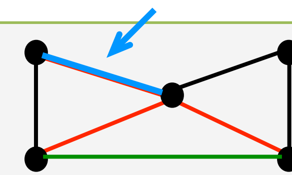
… $|S| \leq k$

Uniform matroid



… S contains at most one element from each group

Partition matroid



… S contains no cycles

Graphic matroid

- S independent ➔ $T \subseteq S$ also independent
- Exchange property: S, U independent, $|S| > |U|$
  ➔ some $e \in S$ can be added to U: $U \cup e$ independent

General: Matroids

Matroid $\mathcal{M} = (\mathcal{V}, \mathcal{I})$

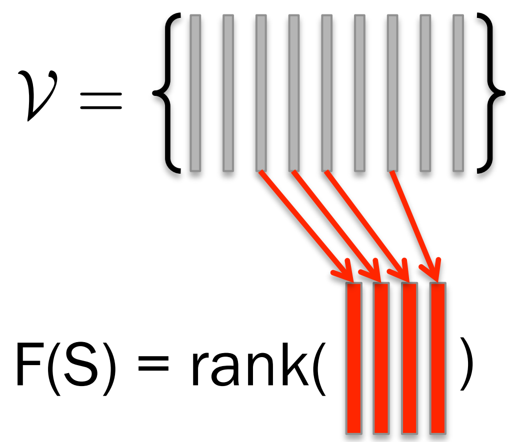
- ground set $\mathcal{V}$
- family of independent sets $\mathcal{I}$

- rank function:

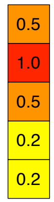
$$F(S) = \max\{\, |T| : T \subseteq S \text{ and } T \in \mathcal{I}\}$$

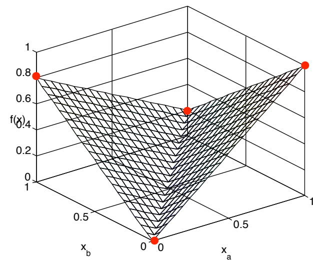
always submodular and increasing

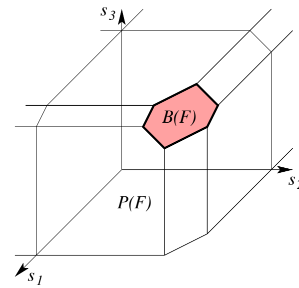
- another special case: matrix rank



Convex relaxation







$$\min_{S\subseteq\mathcal{V}} F(S) = \min_{x\in\{0,1\}^n} F(x) = \min_{x\in[0,1]^n} f(x)$$

1. relaxation: convex optimization (non-smooth)

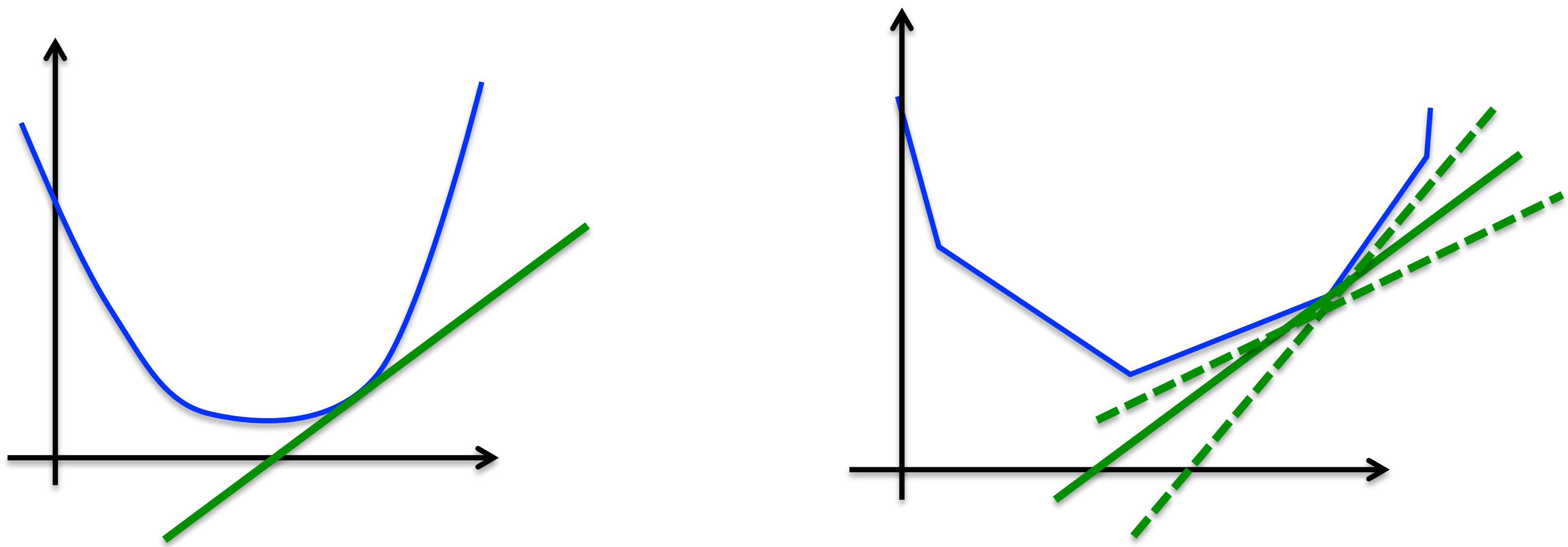
2. relaxation is exact!
   ➔ submodular minimization in polynomial time!
   (Grötschel, Lovász, Schrijver 1981)

Minimizing the Lovász extension

$$\min_{x \in [0,1]^n} \quad f(x)$$

- subgradient method
- combinatorial algorithms: dual

Subgradients



$$\text{subgradient:} \quad g_x \in \arg\max_y \; y^\top x \quad \text{s.t. } y \in \mathcal{B}_F$$

Projected subgradient method

$$\min_{x \in [0,1]^n} \quad f(x)$$

$$x^0 = 0$$

for $t = 0, \ldots$ **do**

find $g^t \in \partial f(x^t)$

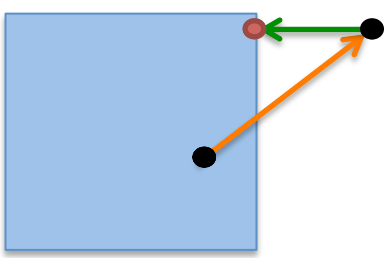
$$x^{t+1} = \Pi_{[0,1]^n}(x^t + \alpha_t g^t)$$

end for

$$g^t = \arg\min_{y \in \mathcal{B}_F} \quad y^\top x^t$$

greedy algorithm ☺

$$\Pi_{[0,1]^n}(y) = \arg\min_{z \in [0,1]^n} \|y - z\|^2$$



Convergence

Theorem

Let $D = \sqrt{n}$ and $L = \max_{g \in \mathcal{B}_F} \|g\| \leq 3 \max_S |F(S)|$.
With step size $\alpha_t = \frac{D}{L\sqrt{t}}$, the error decreases as

$$\min_{\tau \leq t} f(x^\tau) - f(x^*) \leq \frac{4DL}{\sqrt{t}}$$

- D: diameter of $[0,1]^n$
  L: Lipschitz constant
- for error $\leq \epsilon$ need $O(\frac{1}{\epsilon^2})$ iterations

Submodular minimization

convex optimization

- ellipsoid method
  (Grötschel-Lovasz-Schrijver 81)
- subgradient method
- minimum-norm point / Fujishige-Wolfe algorithm
- …

combinatorial methods

- first polynomial-time:
  (Schrijver 00, Iwata-Fleischer-Fujishige 01)
- currently fastest:

  $O(n^4 T + n^5 \log M)$ *(Iwata 03)*

  $O(n^6 + n^5 T)$ *(Orlin 09)*

Latest result: $O(n^2 T \log nM + n^3 \log^c nM)$

$O(n^3 T \log^2 n + n^4 \log^c n)$ *(Lee-Sidford-Wong 15)*

Convex duality

$$\min_{S \subseteq \mathcal{V}} F(S) = \min_{x \in [0,1]^n} f(x)$$

$$= \min_{x \in [0,1]^n} \max_{y \in \mathcal{B}_F} y^\top x$$

$$= \max_{y \in \mathcal{B}_F} \min_{x \in [0,1]^n} x^\top y \qquad = \max_{y \in \mathcal{B}_F} \Big(\sum_{i=1}^{n} \min\{y_i, 0\} \Big)$$

Optimality conditions: (S^*, y^*) optimal primal-dual pair if

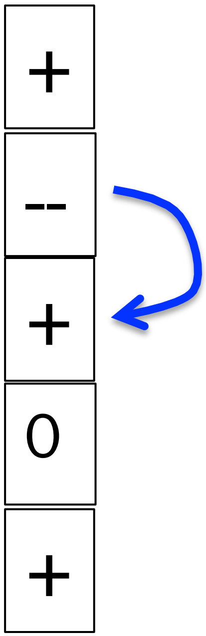
1. $y^* \in \mathcal{B}_F$
2. $\{y^* < 0\} \subseteq S^* \subseteq \{y^* \leq 0\}$
3. $y^*(S^*) = F(S^*)$

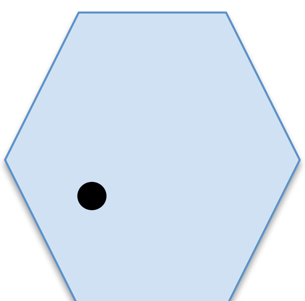
Combinatorial algorithms

- solve $$\max_{y \in \mathcal{B}_F} \Big(\sum_{i=1}^{n} \min\{y_i, 0\} \Big)$$

- remove "negative mass"
- challenges:
  - need to stay in polytope: $$\sum_{i \in S} y_i \leq F(S)$$
  - cannot test feasibility

  ➔ network flow algorithms





Submodular minimization

convex optimization

- ellipsoid method
  (Grötschel-Lovasz-Schrijver 81)
- subgradient method
- minimum-norm point / Fujishige-Wolfe algorithm
- …

combinatorial methods

- first polynomial-time:
  (Schrijver 00, Iwata-Fleischer-Fujishige 01)
- currently fastest:

$$O(n^4 T + n^5 \log M)$$ *(Iwata 03)*

$$O(n^6 + n^5 T)$$ *(Orlin 09)*

Latest result: $O(n^2 T \log nM + n^3 \log^c nM)$

$$O(n^3 T \log^2 n + n^4 \log^c n)$$ *(Lee-Sidford-Wong 15)*

Some fun ☺

1. Complete the proof that
   Lovasz extension convex ➔ set function is submodular
   (slide 51)

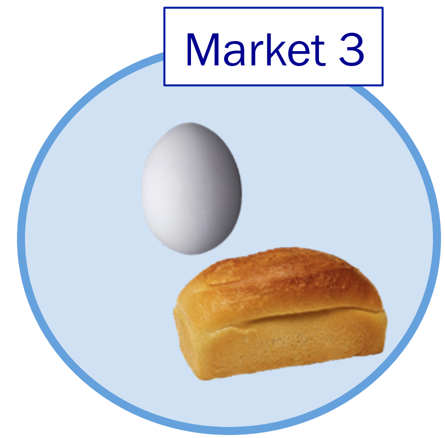
2. Submodular oder not? Let F be increasing and
   submodular, and define
   $$G(S) = \min\{ F(S), c \}$$
   for a constant c.

slides and pointers to literature:
people.csail.mit.edu/stefje/mlss

Example: costs

breakfast??

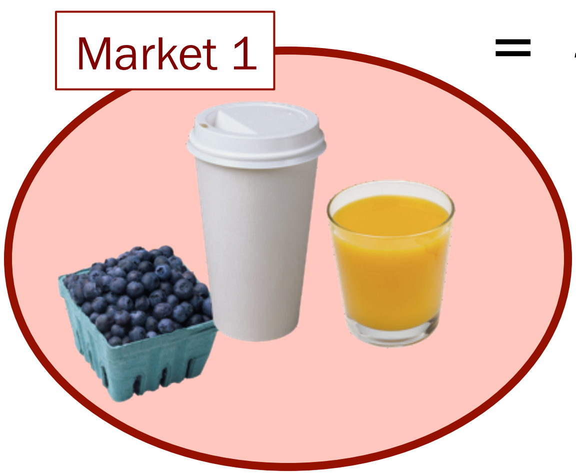
cost:
time to shop
+ price of items



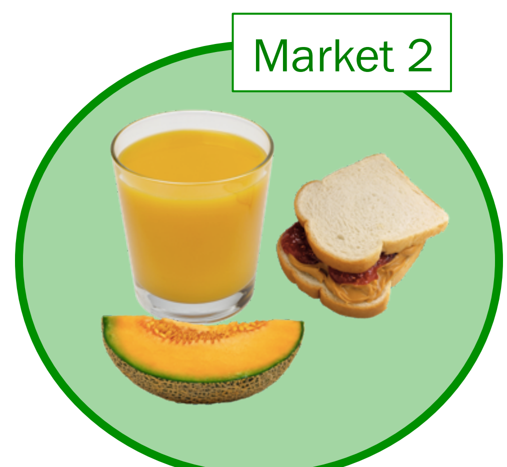
F() = cost() + cost( , )

$= t_1 + 1 + t_2 + 2$

= #shops + #items

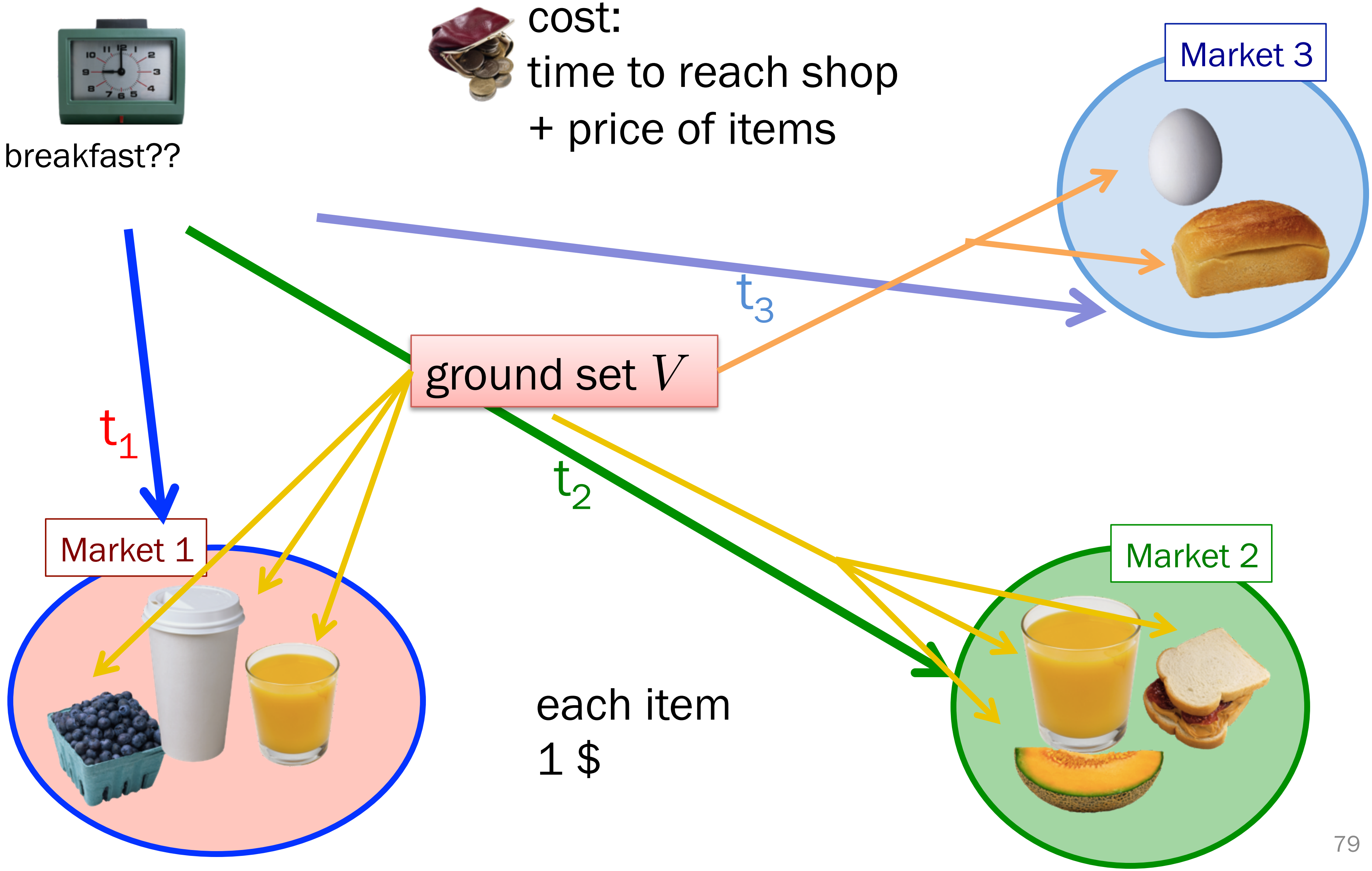


submodular?



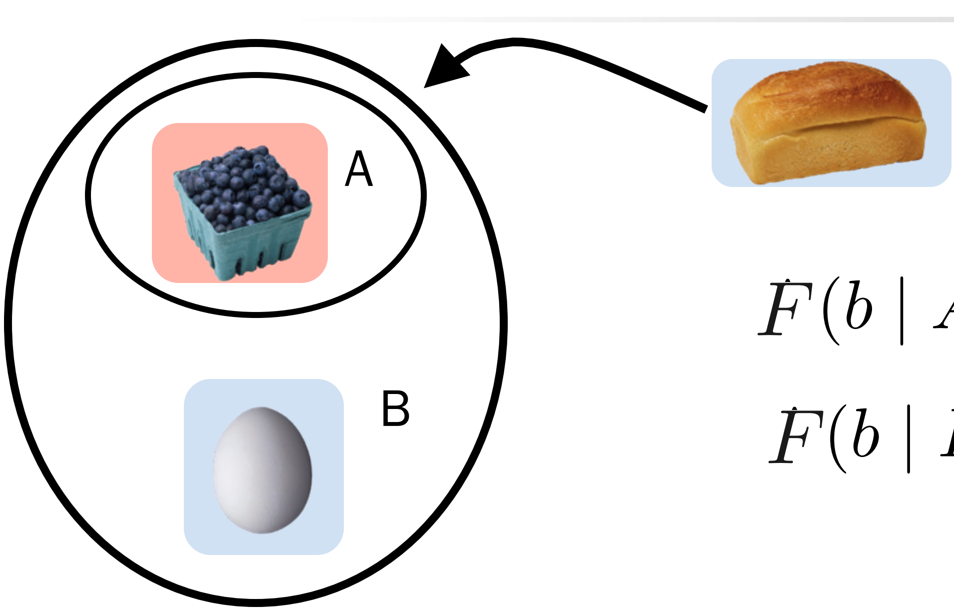
Example: costs

cost:
time to reach shop
+ price of items



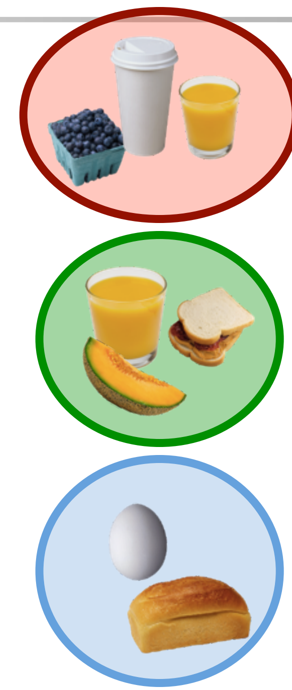
each item
1 $

Shared fixed costs



$$F(b \mid A) = 1 + t_3$$

$$F(b \mid B) = 1$$



marginal cost: #new shops + #new items

decreasing ➔ cost is submodular!

- shops: shared fixed cost
- economies of scale