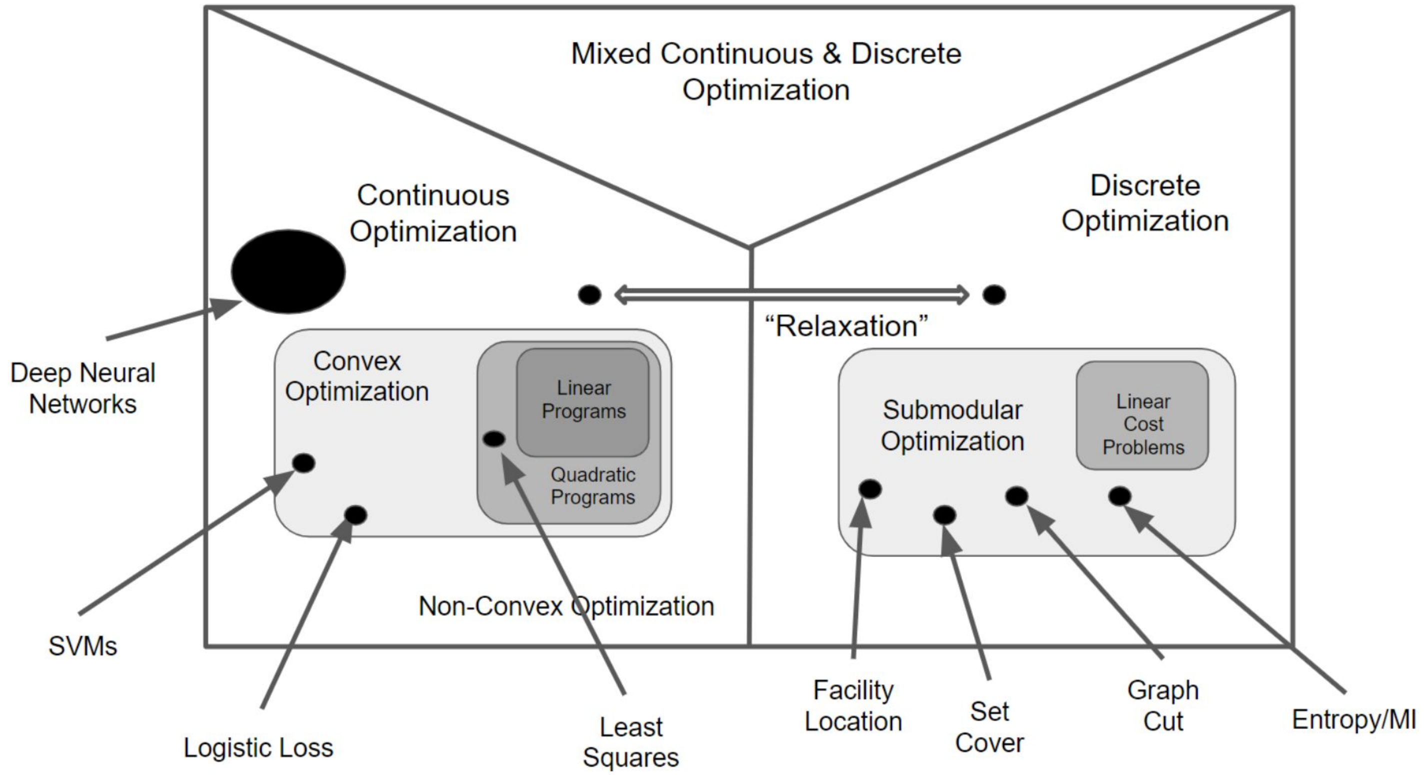
Submodular Functions: Definitions, Examples, Optimization Problems and Applications

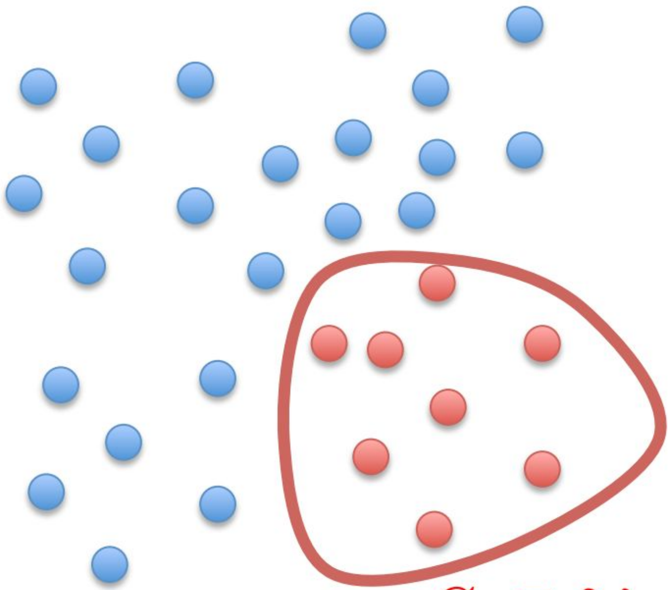
Optimization in Machine Learning

Lecture 19 & 20

Big Picture: Continuous and Discrete Optimization



Discrete Optimization

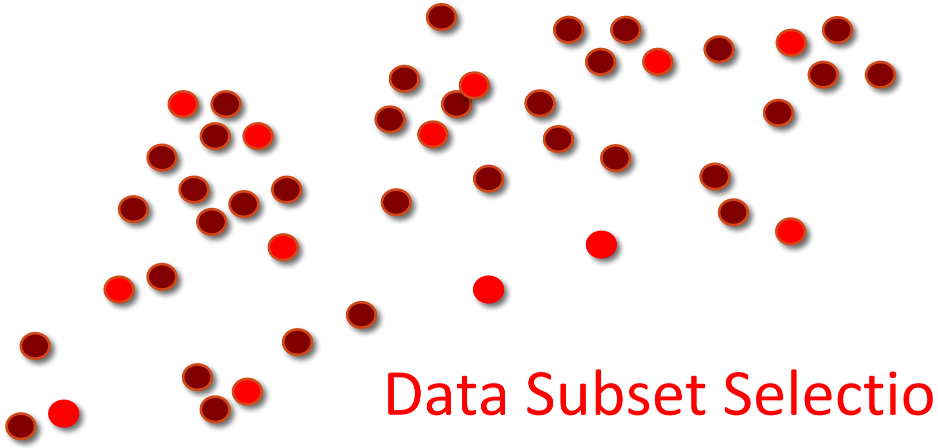


$S \subseteq \mathcal{V}$

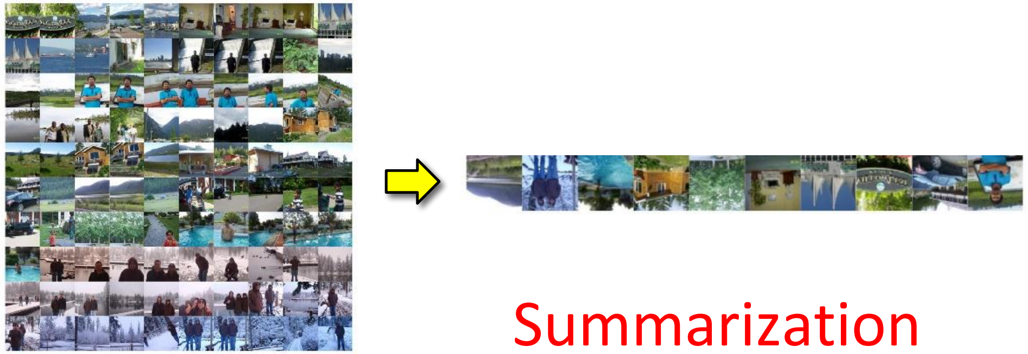
ground set $\mathcal{V}$

(scoring) function

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



Data Subset Selection



Summarization

Outline

- Discrete Optimization in Machine Learning
- **Lectures 19 & 20**
  - Definition and Intuition of Submodularity
  - Modeling Power of Submodular/Set Functions
  - Examples of Submodular Functions
  - Properties of Submodular Functions
- Lectures 21 & 22
  - Applications of Submodular Optimization
  - Optimization Algorithms for Different Function Classes and Constraints
  - Optimization Algorithms for Different Settings (Streaming/Distributed)
  - Practical Implementation tricks

Acknowledgements

Slides borrowed from several sources:

1. Submodular Optimization course at UW from Jeff Bilmes
2. Tutorial on Submodular Optimization by Stefanie Jegelka, Andreas Krause and Jeff Bilmes at ICML and NIPS
3. Some of my own tutorials at WACV, IJCAI, ECAI, ICPM etc.

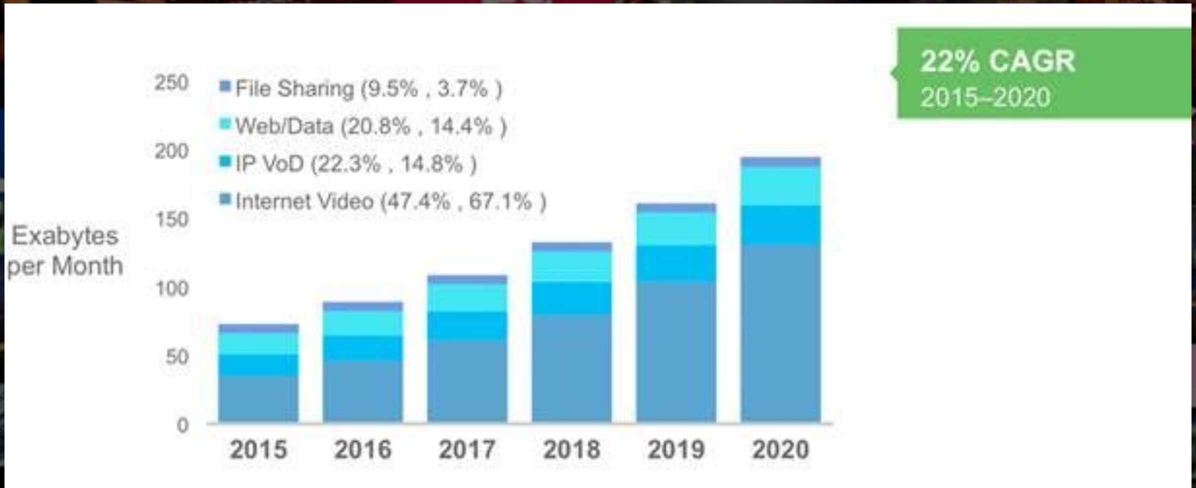
Useful Material

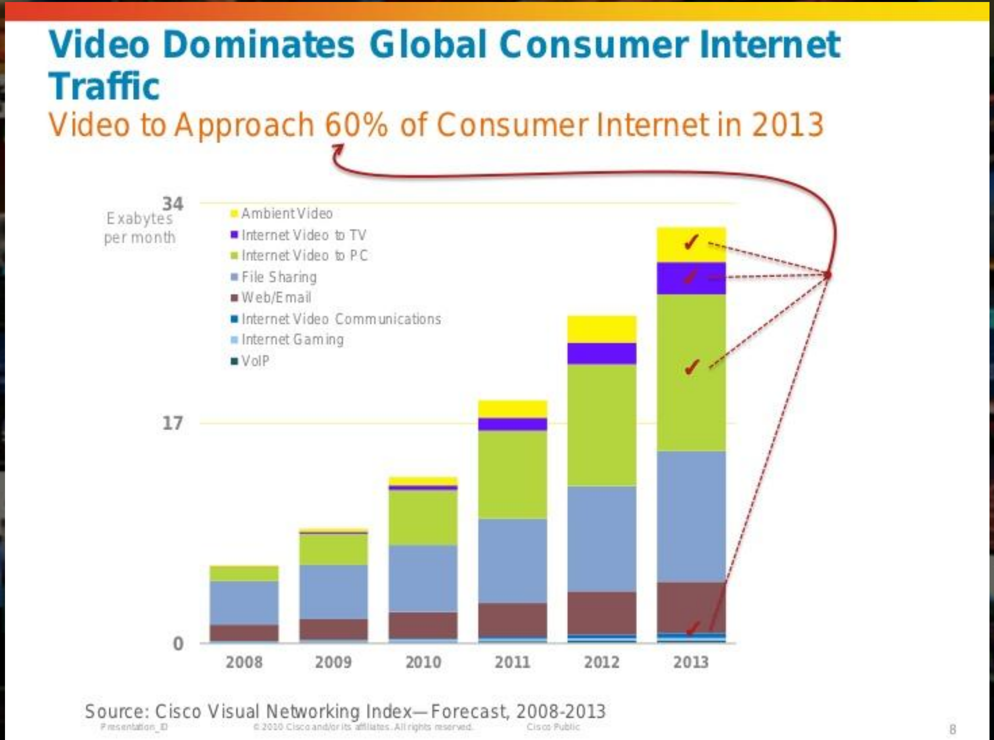
- Fujishige, “Submodular Functions and Optimization”, 2005
- Narayanan, “Submodular Functions and Electrical Networks”, 1997
- Welsh, “Matroid Theory”, 1975
- Oxley, “Matroid Theory”, 1992 (and 2011).
- Lawler, “Combinatorial Optimization: Networks and Matroids”, 1976.
- Schrijver, “Combinatorial Optimization”, 2003
- Gruenbaum, “Convex Polytopes, 2nd Ed”, 2003.
- Additional readings that will be announced here.

Useful material

- Jeff's Class: https://people.ece.uw.edu/bilmes/classes/ee563_spring_2018/
- Stefanie Jegelka & Andreas Krause's 2013 ICML tutorial: http://techtalks.tv/talks/submodularity-in-machine-learning-new-directions-part-i/58125/
- Jeff's NIPS, 2013 tutorial on submodularity: http://melodi.ee.washington.edu/~bilmes/pgs/b2hd-bilmes2013-nips-tutorial.html and http://youtu.be/c4rBof38nKQ
- Andreas Krause's web page: http://submodularity.org
- Francis Bach's updated 2013 text: http://hal.archives-ouvertes.fr/docs/00/87/06/09/PDF/submodular_fot_revisd_hal.pdf
- Tom McCormick's overview paper on submodular minimization: http://people.commerce.ubc.ca/faculty/mccormick/sfmchap8a.pdf

Data Explosion





Exabytes of visual data is being created **everyday!**

Data Explosion

Challenges:

- Effectively consume big data
- Find events of interest efficiently
- Train Machine Learning Models with massive datasets

Big Data in Machine Learning

Big Data is very different from Small Data

“*our initial experiments . . . on a dataset of ten thousand images were very discouraging. However, increasing the image collection to two million yielded a qualitative leap in performance*” (Hayes & Effros 2007)

Problems with Big Data

- Plagued with redundancy
- Big Question: Can Statistical predictions and actions be made **cost effective** with the right small data?

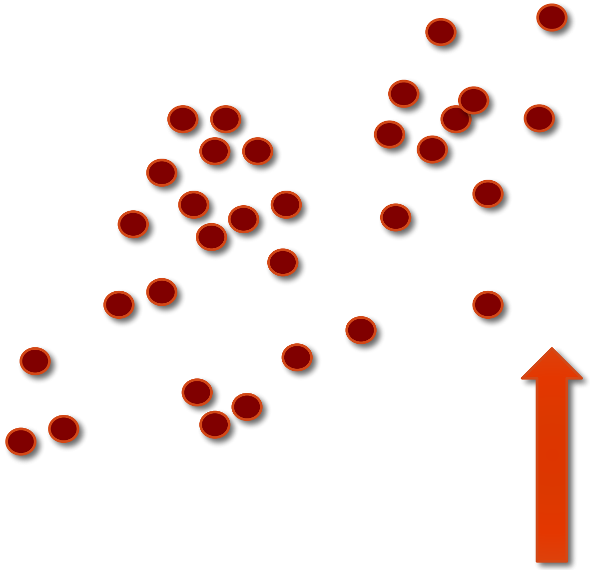
This problem is increasingly important with the advent of Data Hungry Deep Neural Networks!

Big Data in Machine Learning

Small Data

Each Data Point represents
Visual features in n-dimensional space
(for example objects or scenes in an
image)

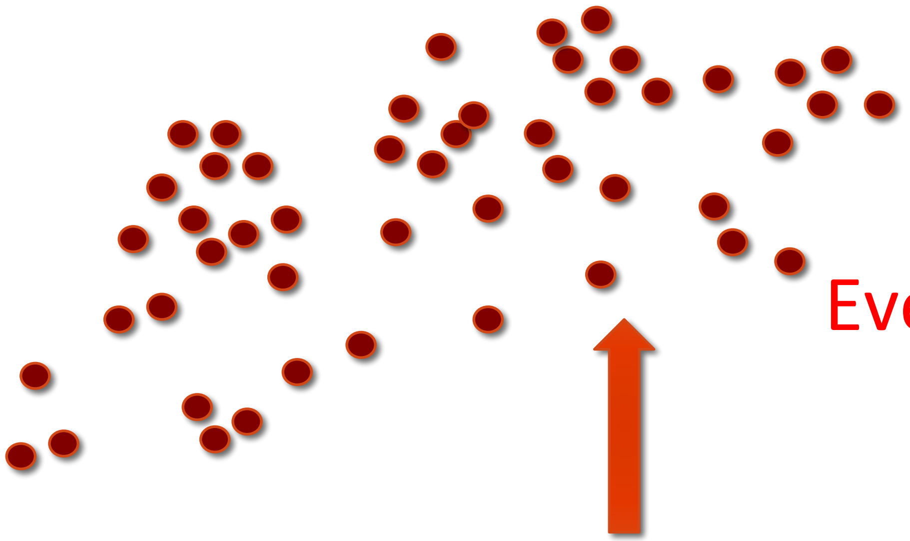
Big Data in Machine Learning



Little More Data
😊

Each Data Point represents
Visual features in n-dimensional space
(for example objects or scenes in an image)

Big Data in Machine Learning

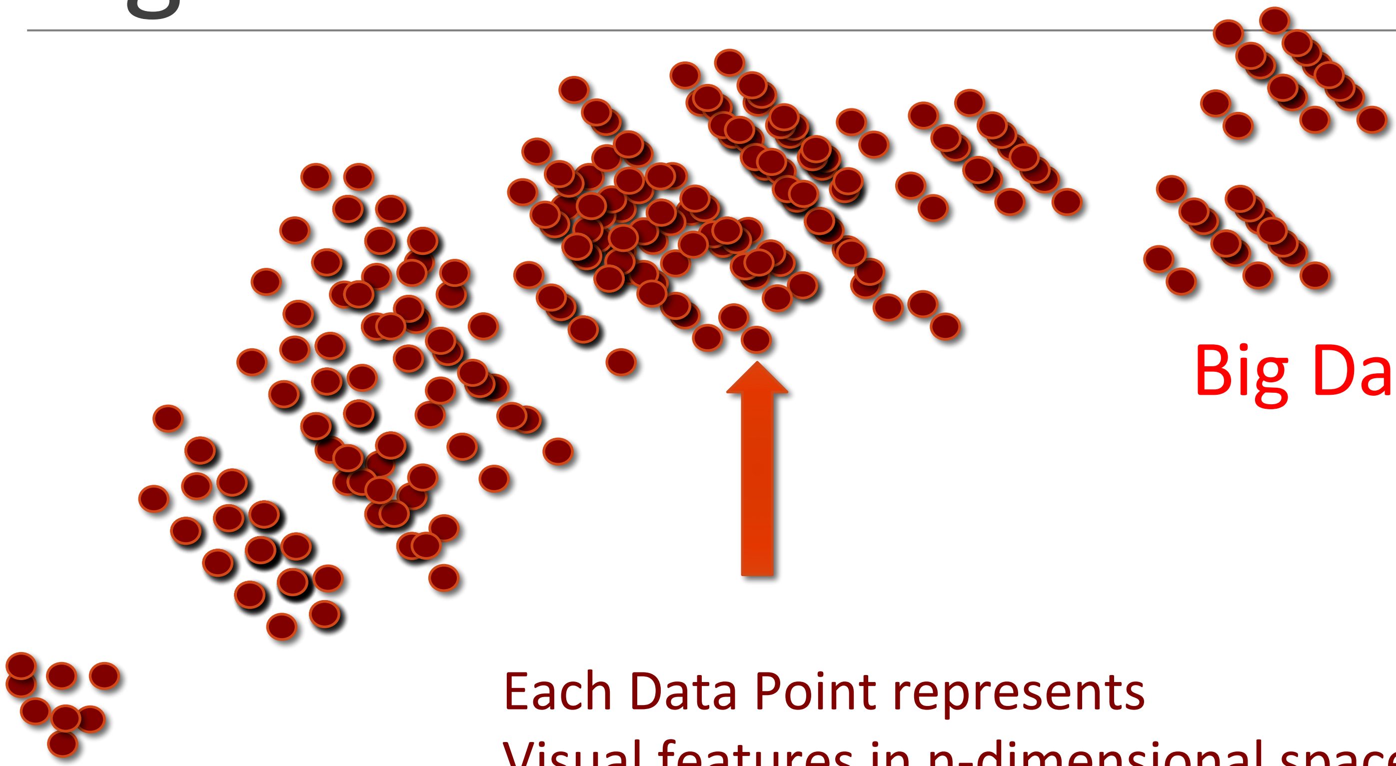


Even More Data 😊

Each Data Point represents
Visual features in n-dimensional space
(for example objects or scenes in an
image)

The more Data we get,
the better Is our
understanding of the space,
And correspondingly,
the better will be
the performance of
Statistical algorithms

Big Data in Machine Learning



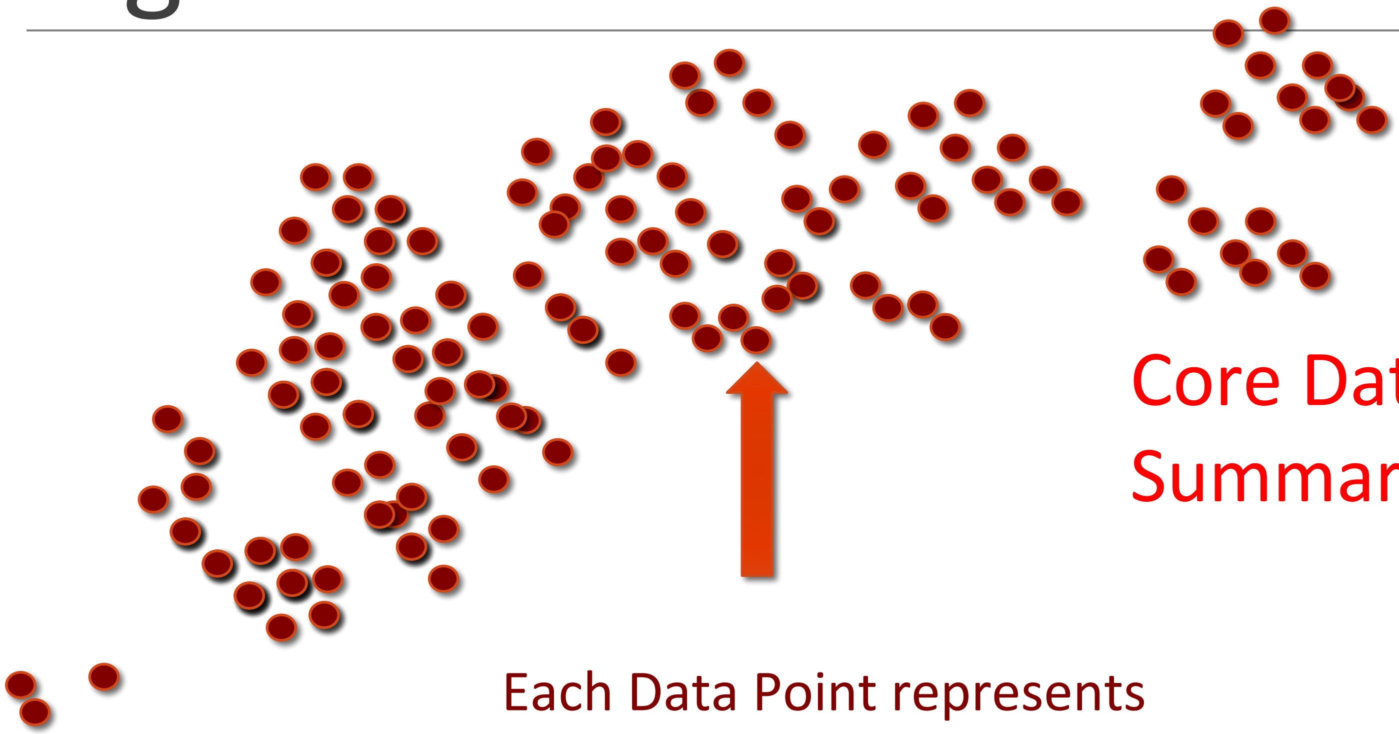
Big Data!

Each Data Point represents
Visual features in n-dimensional space
(for example objects or scenes in an
image)

Has the complete picture
of the distribution of data.

However, Clearly there is a
lot of redundancy,
And wasteful utilization
of the resources!

Big Data in Machine Learning



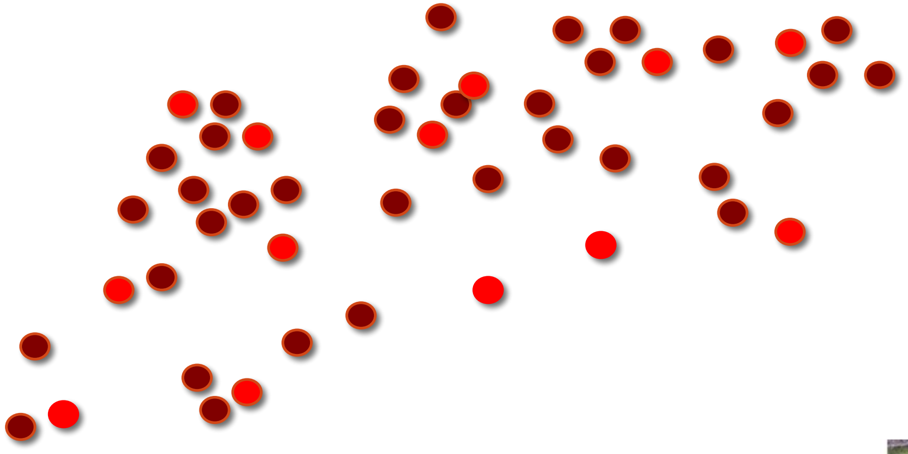
Core Data or Summary!

Each Data Point represents
Visual features in n-dimensional space
(for example objects or scenes in an image)

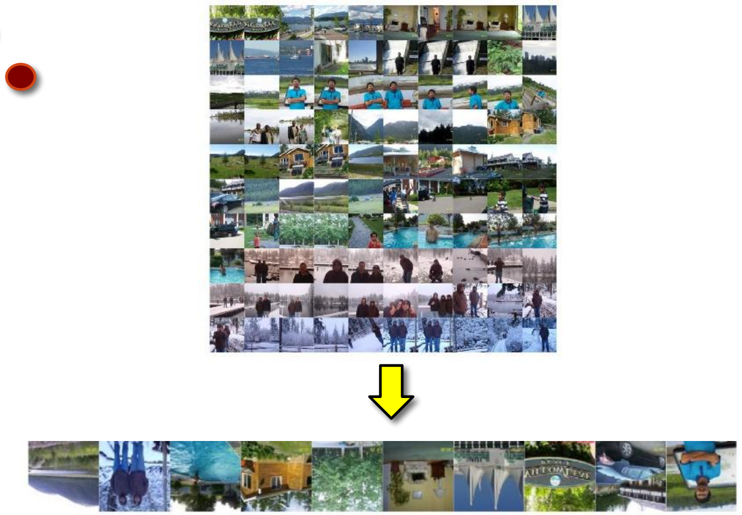
Maintains the Representation of the Data, while requiring Much lesser number Of samples

⇒ Lesser Training and Labeling Costs!

Combinatorial Subset Selection Problems



Data Subset Selection



Summarization

Image Segmentation

Subset Selection Problems everywhere!

Discrete Optimization in Machine Learning

- MAP inference in Probabilistic Models: Ising Models, DPPs
- Feature Subset Selection
- Data Partitioning
- Data Subset Selection
- Data Summarization: Text, Images, Video Summarization
- Social networks, Influence Maximization
- Natural Language Processing: words, phrases, n-grams, syntax trees, semantic structures
- Computer Vision: Image Segmentation, Image Correspondence
- Genomics and Computational Biology: cell types or assay selection, selecting peptides and proteins

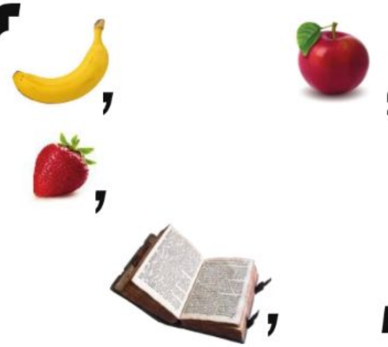
Outline

- Discrete Optimization in Machine Learning
- **Lecture 19 & 20**
  - Definition and Intuition of Submodularity
  - Modeling Power of Submodular/Set Functions
  - Examples of Submodular Functions
  - Examples of Submodular Optimization
- Lectures 21 & 22
  - Optimization Algorithms for Different Function Classes and Constraints
  - Optimization Algorithms for Different Settings (Streaming/Distributed)
  - Practical Implementation tricks

Combinatorial Subset Selection Problems

$V = \{$  $\}$

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

$A = \{$  $\}$

Choose Subset $A \subseteq V$

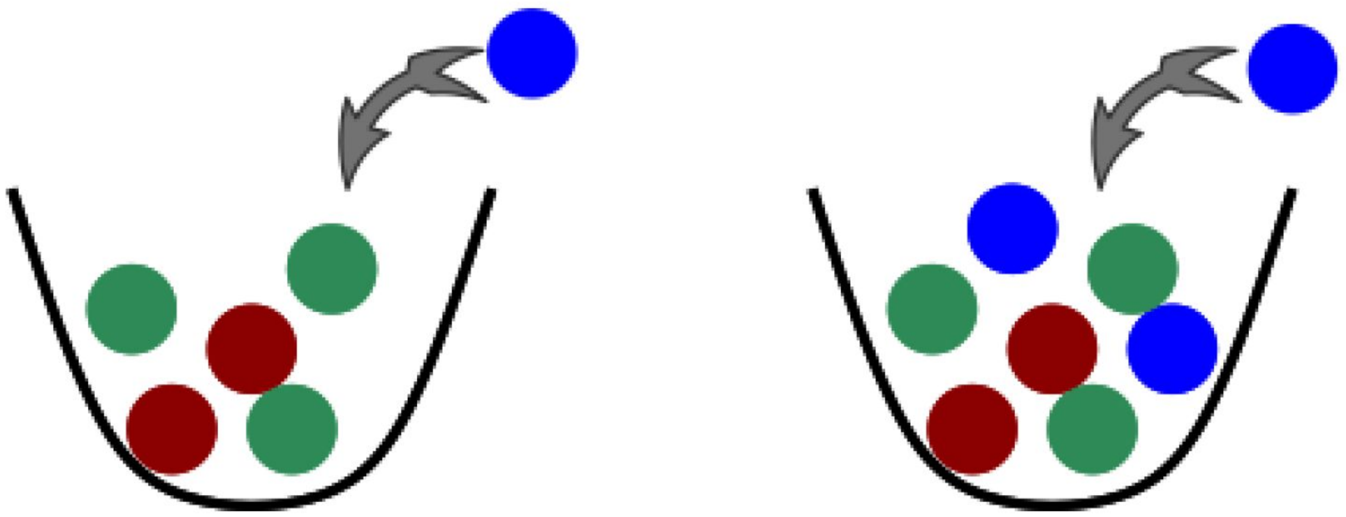
$$f(A) = 22$$

General Set function Optimization: very hard!

What if there is some special structure?

Submodular Functions

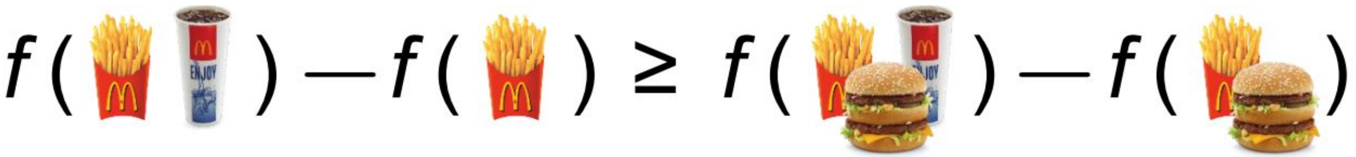
$$f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B$$



$f = \#$ of distinct colors of balls in the urn.

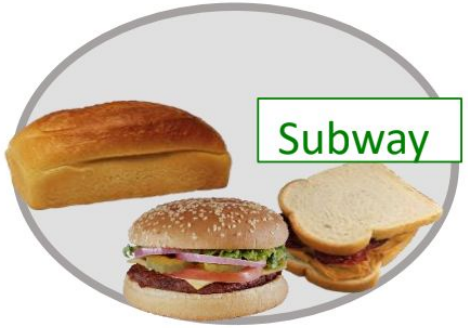
Negative of a Submodular Function is a Super-modular Function!

Submodular Functions



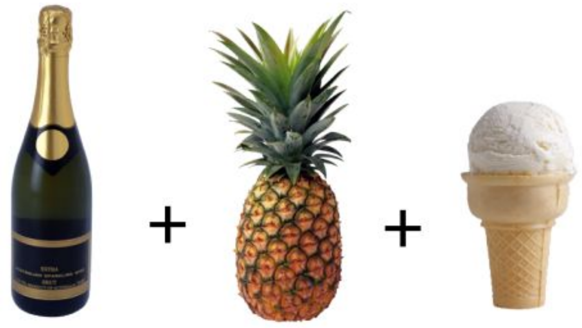
Diminishing Returns!

The more items you buy,
the more the
discount!



Modular Functions

- each element e has a weight $w(e)$



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

$$A \subset B$$

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

Modular Functions are both submodular and super-modular!

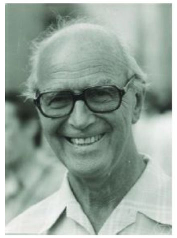
Monotone Submodular Functions

- A set function is called monotonic if

  $$A \subseteq B \subseteq V \Rightarrow F(A) \leq F(B)$$

- Examples:
  - Influence in social networks [Kempe et al KDD '03]
  - For discrete RVs, entropy $F(A) = H(X_A)$ is monotonic:
    Suppose $B = A \cup C$. Then
    $$F(B) = H(X_A, X_C) = H(X_A) + H(X_C \mid X_A) \geq H(X_A) = F(A)$$
  - Information gain: $F(A) = H(Y) - H(Y \mid X_A)$

Submodularity (almost) everywhere



THEORY OF CAPACITIES (1)
by Gustave CHOQUET (2)(3).

INTRODUCTION

This work originated from the following]
significance had been emphasized by M. Brelot
Is the interior Newtonian capacity of an ar
subset X of the space R^3 equal to the exte
capacity of X?

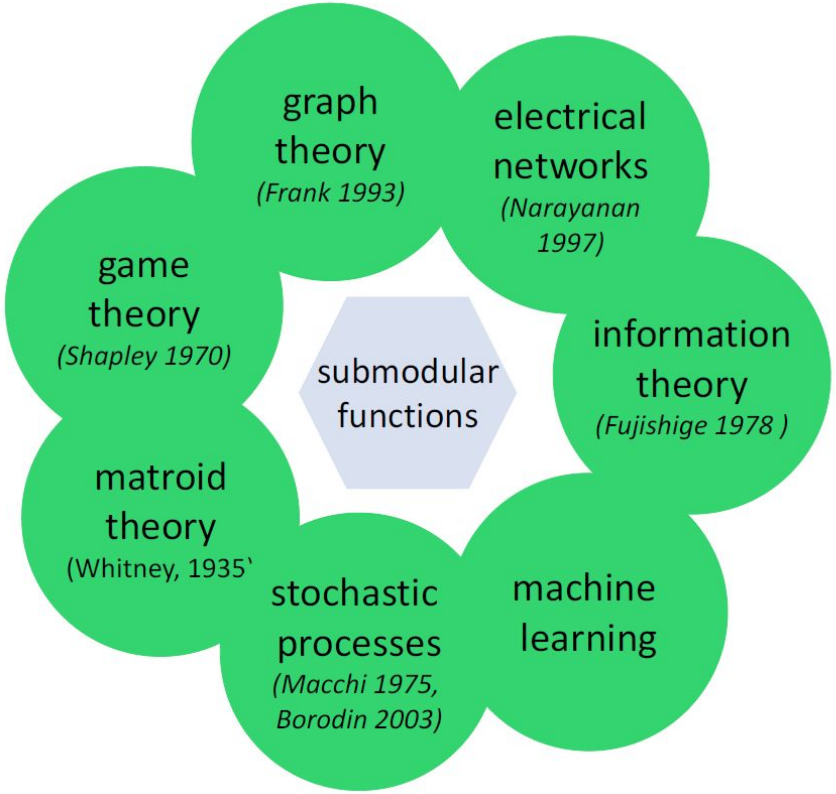
Submodular Functions, Matroids, and Certain Polyhedra*

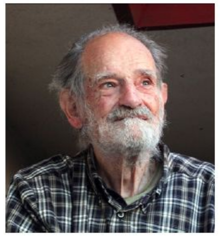
Jack Edmonds

National Bureau of Standards, Washington, D.C.,

I

The viewpoint of the subject of matroids, and related areas of lattice theory has always been, in one way or another, abstraction of algebraic dependence or equivalently, abstraction of the incidence relations in geometric representation of algebra. Often one of the main derived facts is that all bases have the sam cardinality. (See Van der Waerden, Section 33.)





Cores of Convex Games 1)

By LLOYD S. SHAPLEY 2)

ct: The core of an n-person game is the set of feasible outcomes tl
coalition of players. A convex game is defined as one that is ba
paper it is shown that the core of a convex game is not empty
structure. It is further shown that certain other cooperative solu
way to the core: The value of a convex game is the center of gravit

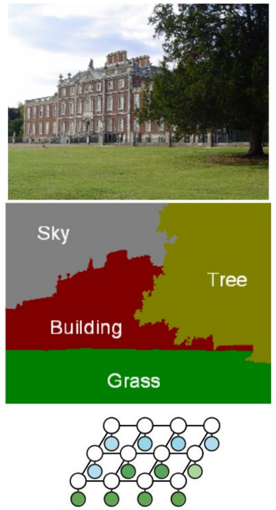
Submodular functions and convexity

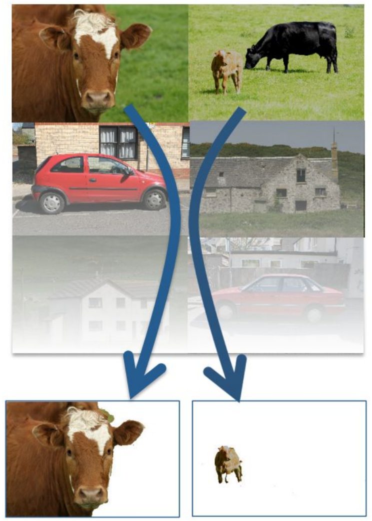
L. Lovász
Eötvös Loránd University, Department of Analysis I, Múzeu
Budapest, Hungary

0. Introduction

In "continuous" optimization convex functions play a central role. Bes
mentary tools like differentiation, various methods for finding the min
a convex function constitute the main body of nonlinear optimizat

Submodular Optimization in Machine Learning



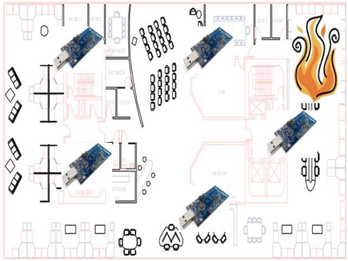


$F(S) = \text{coherence} \ + \ \text{likelihood}$

Discrete Labeling

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

Summarization



- where put sensors?
- which experiments?
- summarization

$F(S) = \text{``information''}$

Sensor Placement

Submodular Functions for Summarization

Subset

$\subset$

Superset

Submodular Functions for Summarization

Subset

$\subset$

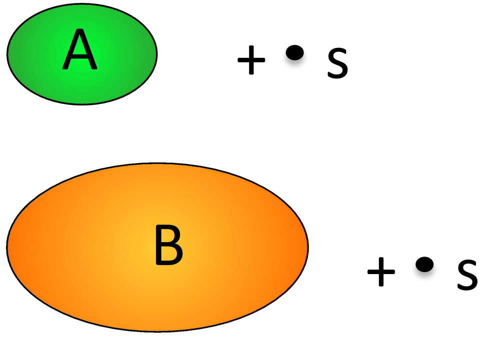
Superset

Information gain reduces with larger sets!

Two Equivalent Definitions of Submodularity

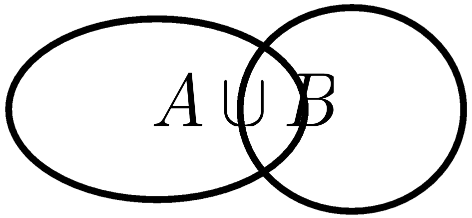
- Diminishing gains: for all $A, B \subseteq V$

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



- Union-Intersection: for all $A, B \subseteq V$

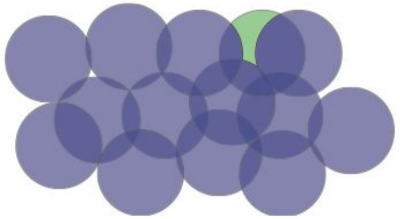
$$F(A) + F(B) \quad \geq \quad F(A \cup B) + F(A \cap B)$$



Instantiations of Set Functions

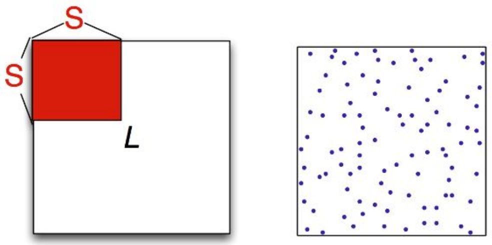
- **Representation Functions**
  - Facility Location Function (k-mediods clustering)
  - Graph Cut Family, Saturated Coverage
- **Diversity Functions**
  - Dispersion Functions (Min, Sum, Min-Sum)
  - Determinantal Point Processes
- **Coverage Functions**
  - Set Cover Function
  - Probabilistic Set Cover Function
  - Feature Based Functions
- **Importance Functions**
  - Modular Functions
- **Information Functions**
  - Mutual Information
  - Entropy
- **Discounted Cost Functions**
  - Clustered Concave over Modular Functions
  - Cooperative Costs and Saturations
- **Complexity Functions**
  - Bipartite Neighborhood Functions

Facets of Submodularity: Maximization



$$F(A) = \cup_{s \in A} \text{area}(s)$$

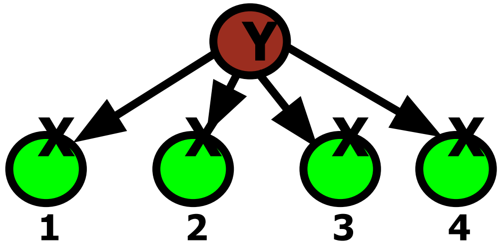
Coverage



$L \in \mathbb{R}^{p \times p}$ psd

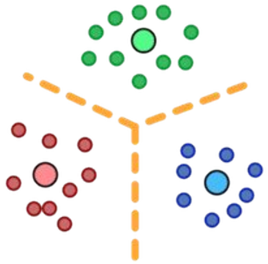
$$F(A) = \log det(L_A)$$

Diversity



$$F(A) = H(X_A)$$

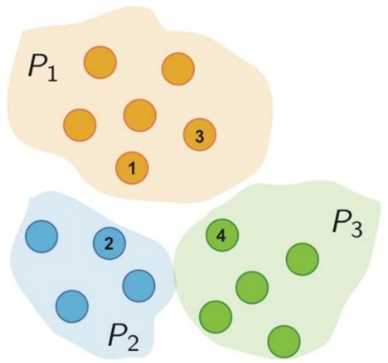
Information



$$F(A) = \sum_{i \in V} \max_{j \in A} s_{ij}$$

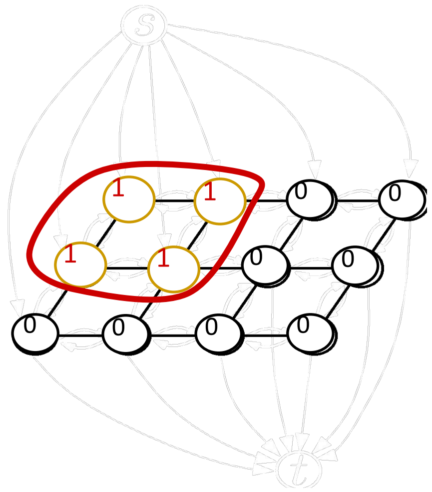
Representation

Facets of Submodularity: Minimization



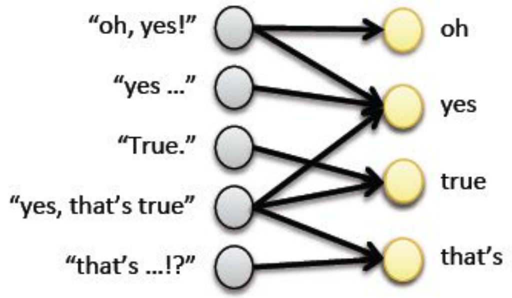
$$F(A)=\sum_{i=1}^{3}\sqrt{\sum_{j\in A\cap P_i} r_i}$$

Cooperative Costs



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

Attractive Potentials



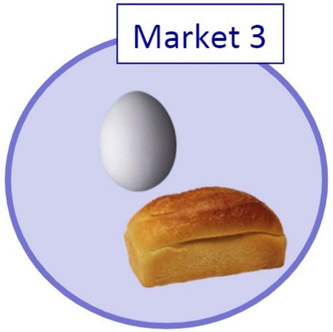
$$F(A)=\gamma(A)$$

Complexity

Co-operative Costs

breakfast??

cost:
time to shop
+ price of items

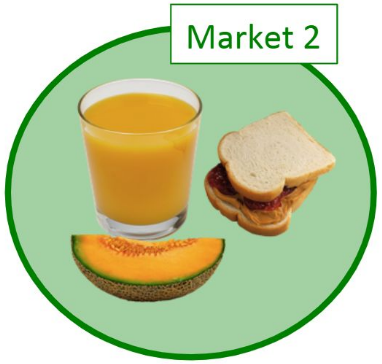


F( ) = cost() + cost( , )

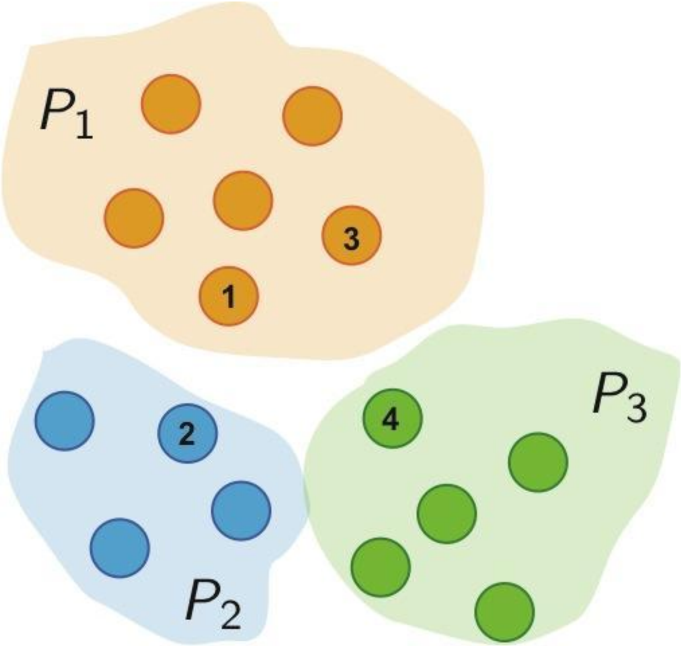
$= t_1 + 1 \quad + \quad t_2 \quad + 2$

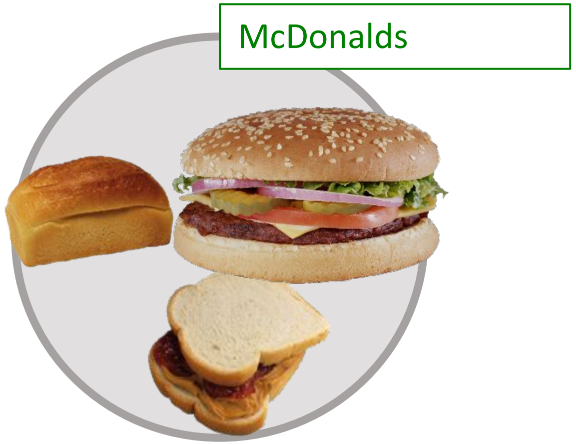
= #shops + #items

submodular?



Cooperative Costs



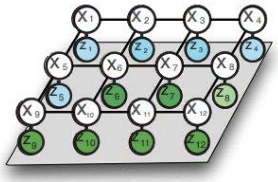


$$F(A) = \sum_{i=1}^{3} \sqrt{\sum_{j \in A \cap P_i} r_i}$$

Iyer-Bilmes 2013, Jegelka-Bilmes 2011, …

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:

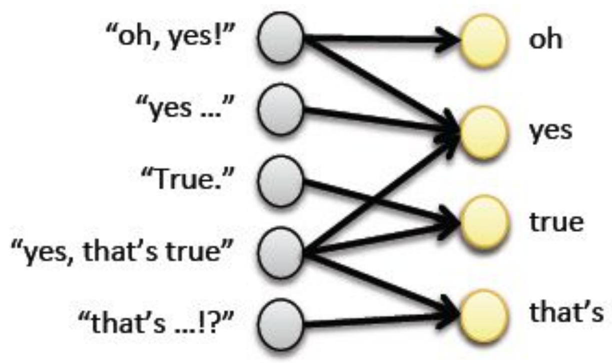
$$E_{ij}(1,0) + E_{ij}(0,1) \;\geq\; E_{ij}(0,0) + E_{ij}(1,1)$$

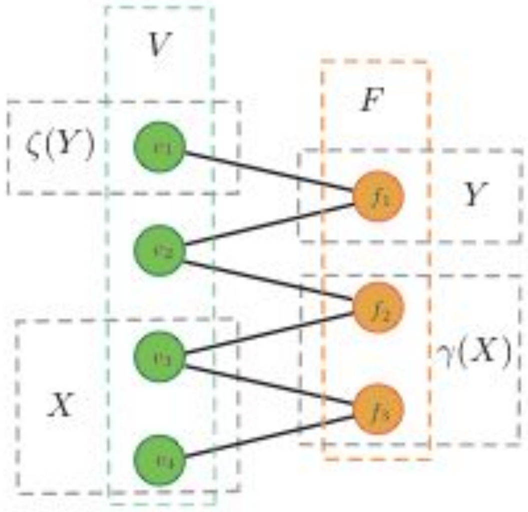
i j	i j	i j	i j
$S = \{i\}$	$T = \{j\}$	$S \cap T = \emptyset$	$S \cup T$

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

Boykov-Jolly 2001, …

Complexity Functions





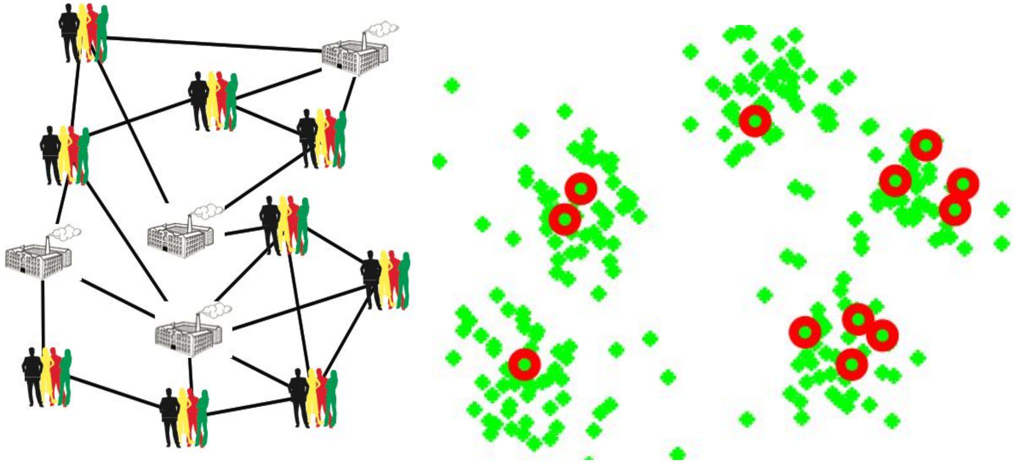
(a) Bipartite graph

Example: Selecting a limited complexity subset for quick Experimental turn around time!

$$F(A) = \gamma(A)$$

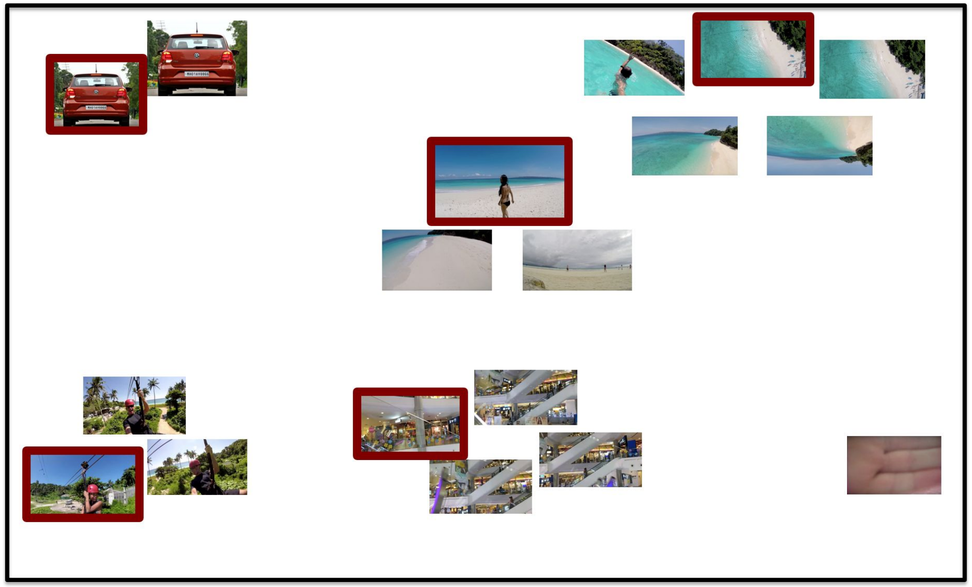
Liu-Iyer et al 2015, ...

Representation Functions



Facility Location	$\sum_{i \in V} \max_{k \in X} s_{ik}$
Saturated Coverage	$\sum_{i \in V} \min\{\sum_{j \in X} s_{ij}, \alpha_i\}$
Graph Cut	$\lambda \sum_{i \in V} \sum_{j \in X} s_{ij} - \sum_{i,j \in X} s_{ij}$

Similarity Kernel

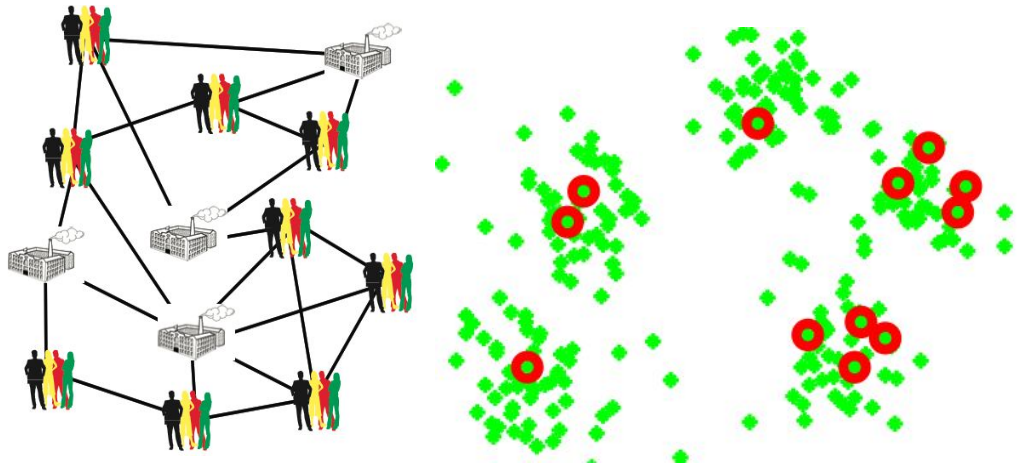


Representation Functions

Picks Centroids

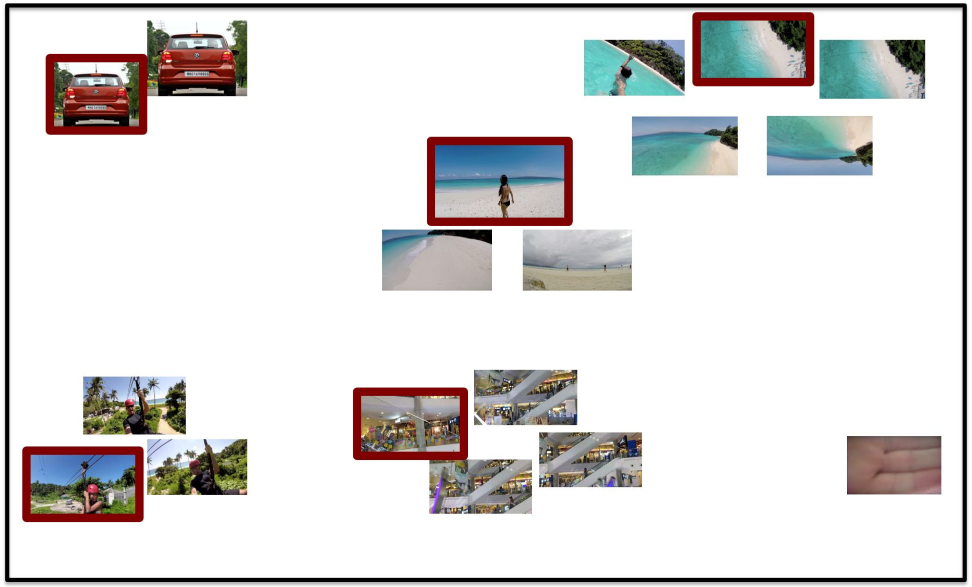
Iyer 2015, Kaushal et al 2019, Tschiatchek et al 2014, …

Representation Functions



Facility Location	$\sum_{i \in V} \max_{k \in X} s_{ik}$
Saturated Coverage	$\sum_{i \in V} \min\{\sum_{j \in X} s_{ij}, \alpha_i\}$
Graph Cut	$\lambda \sum_{i \in V} \sum_{j \in X} s_{ij} - \sum_{i,j \in X} s_{ij}$

Graph Cut is not monotone submodular when $\lambda < 2$



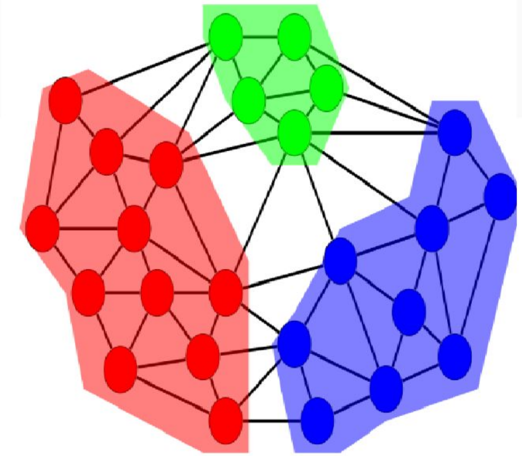
Representation Functions

Picks Centroids

Iyer 2015, Kaushal et al 2019, Tschiatchek et al 2014, …

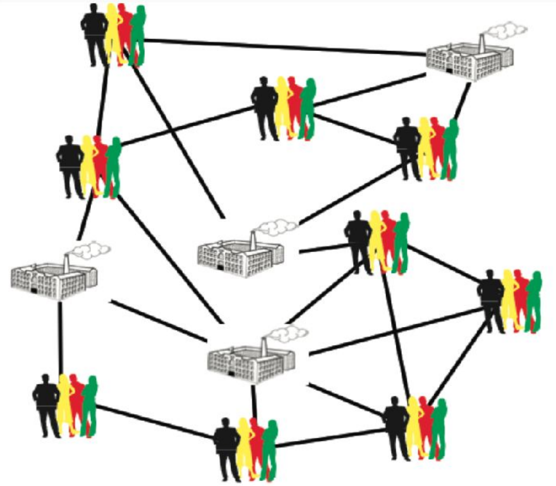
Representation Functions

Characterized by similarity kernel s_{ij} between elements i and j



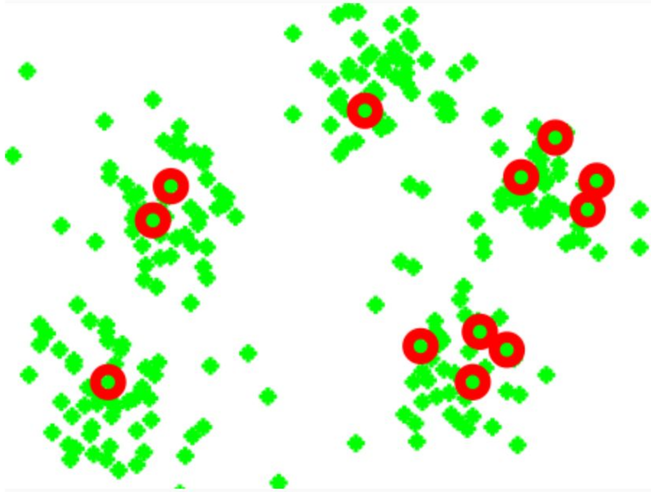
Graph Cut: Not monotone submodular when $\lambda < 2$

$$\lambda \sum_{i \in V} \sum_{j \in X} s_{ij} - \sum_{i,j \in X} s_{ij}$$



Facility Location:

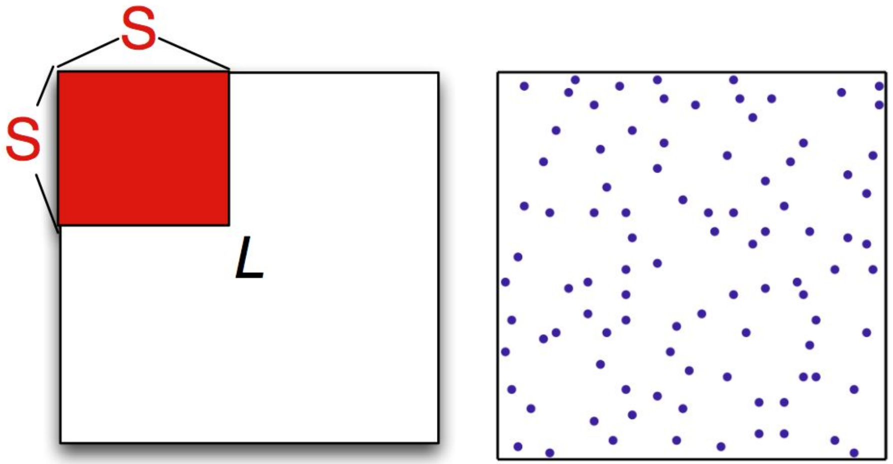
$$\sum_{i \in V} \max_{k \in X} s_{ik}$$



Saturated Coverage:

$$\sum_{i \in V} \min\{\sum_{j \in X} s_{ij}, \alpha_i\}$$

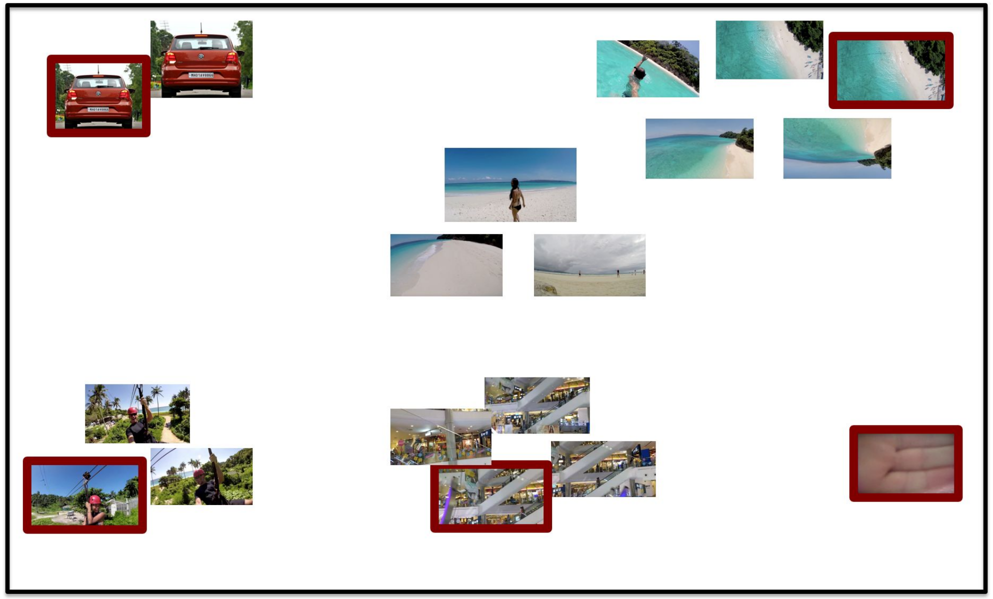
Diversity Functions: DPPs



Determinantal Point Processes

$$F(S) = \log \det(L_S)$$

Similarity Kernel



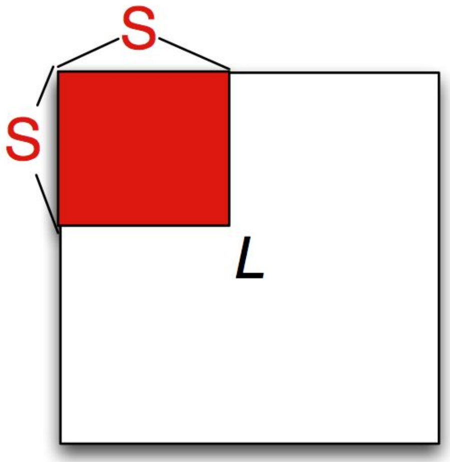
Diversity Functions

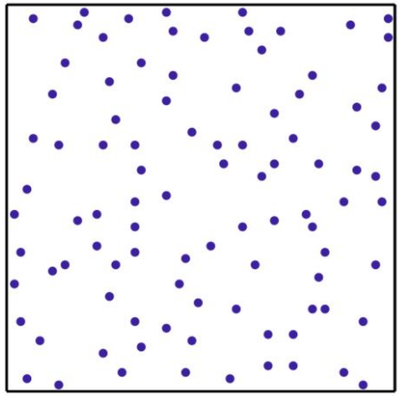
Picks items as different as possible!

Kulesza-Taskar 2012,

...

Diversity Functions: DPPs

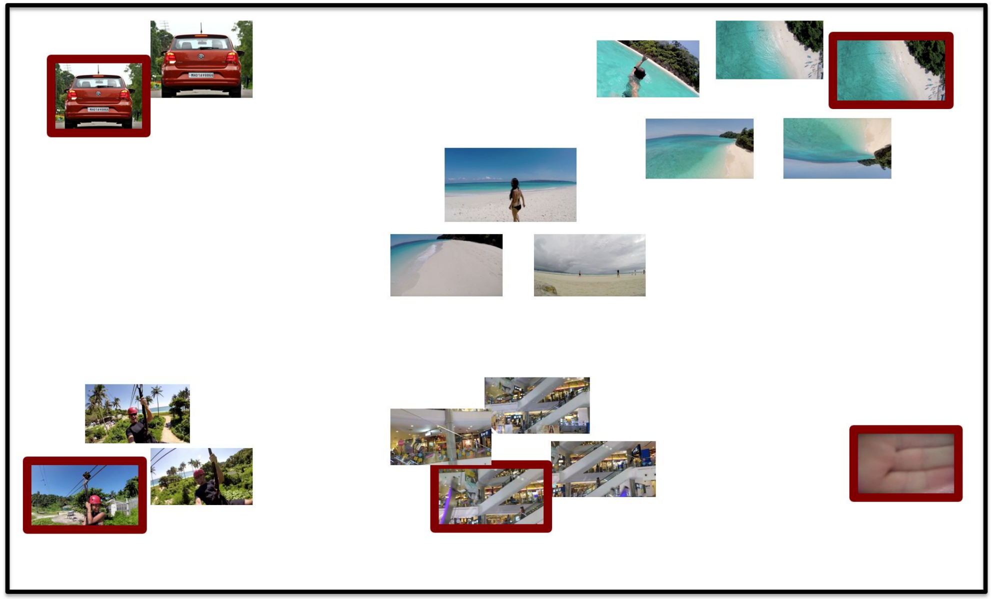




Determinantal Point Processes

$$F(S) = \log\det(L_S)$$

Log-Det Function is Non-Monotone Submodular!



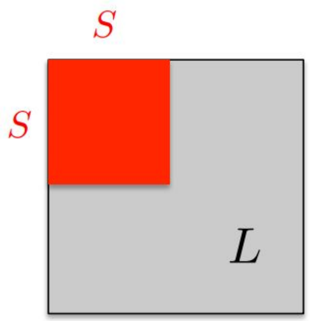
Diversity Functions

Picks items as different as possible!

Kulesza-Taskar 2012,

...

Determinantal Point Processes



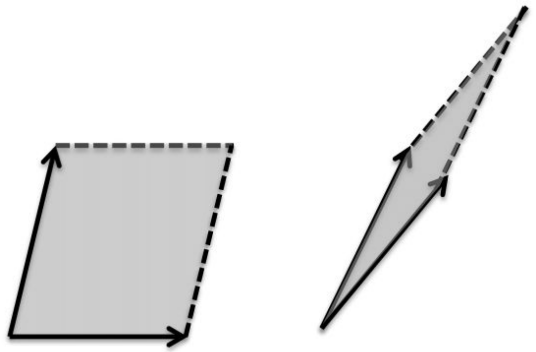
- similarity matrix L

$$L_{ij} = x_i^\top x_j$$

- sample set *Y*:

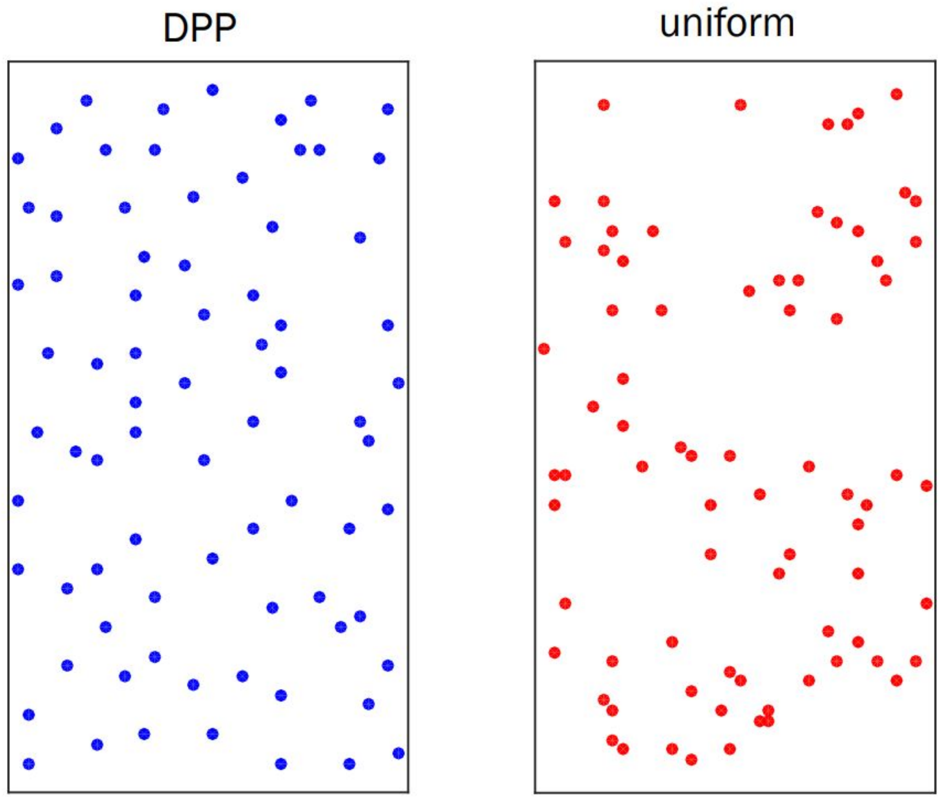
$$P(Y = S) \propto \det(L_S)$$

$$= \mathrm{Vol}(\{x_i\}_{i \in S})^2$$

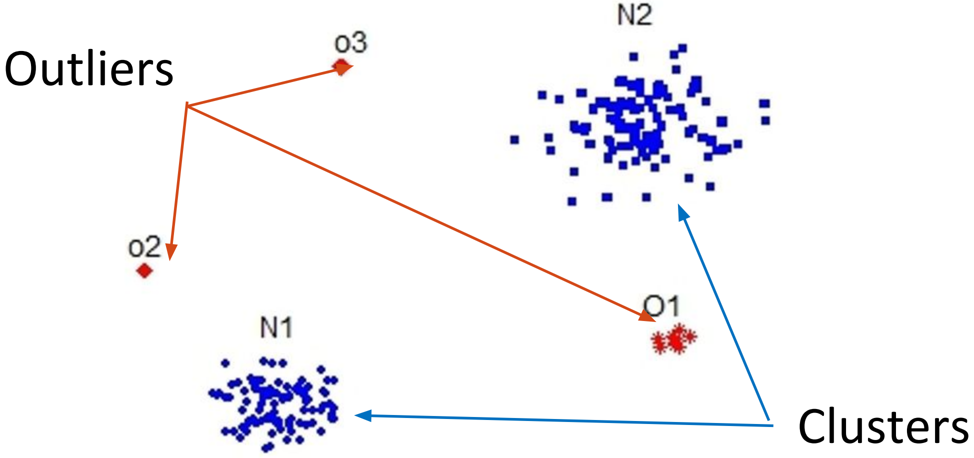


$$F(S) = \log\det(K_S)$$

is submodular!

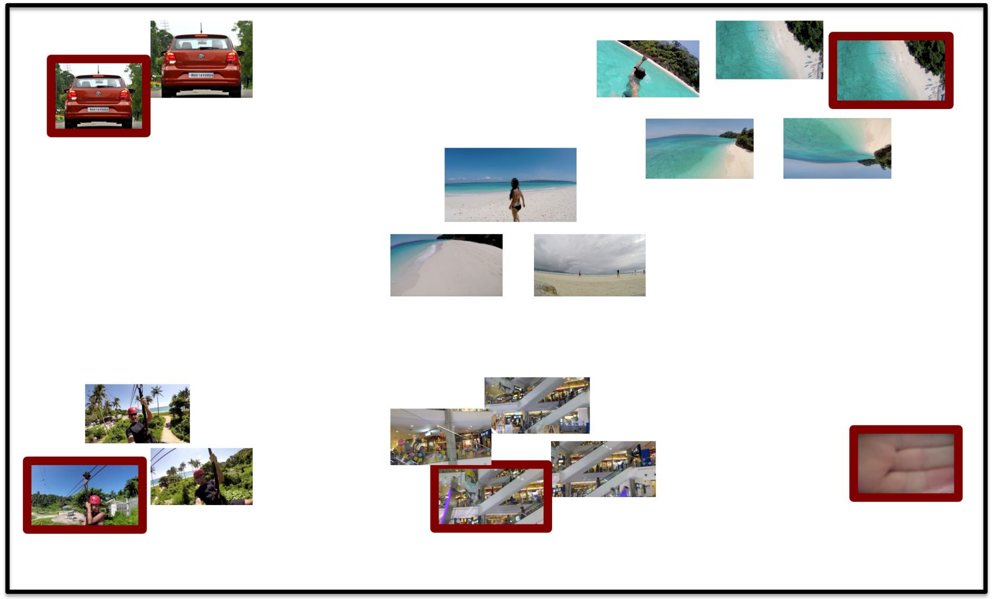


Diversity Functions: Dispersion



Dispersion Min	$\min_{k,l \in X, k \neq l} d_{kl}$
Dispersion Sum	$\sum_{k,l \in X} d_{kl}$
Dispersion Min-Sum	$\sum_{k \in X} \min_{l \in X} d_{kl}$

Distance Measure

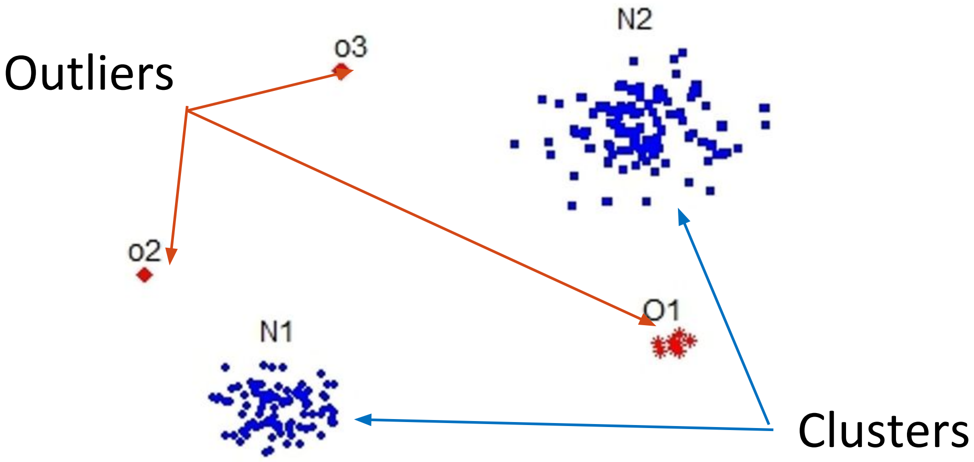


Diversity Functions

Picks items as different as possible!

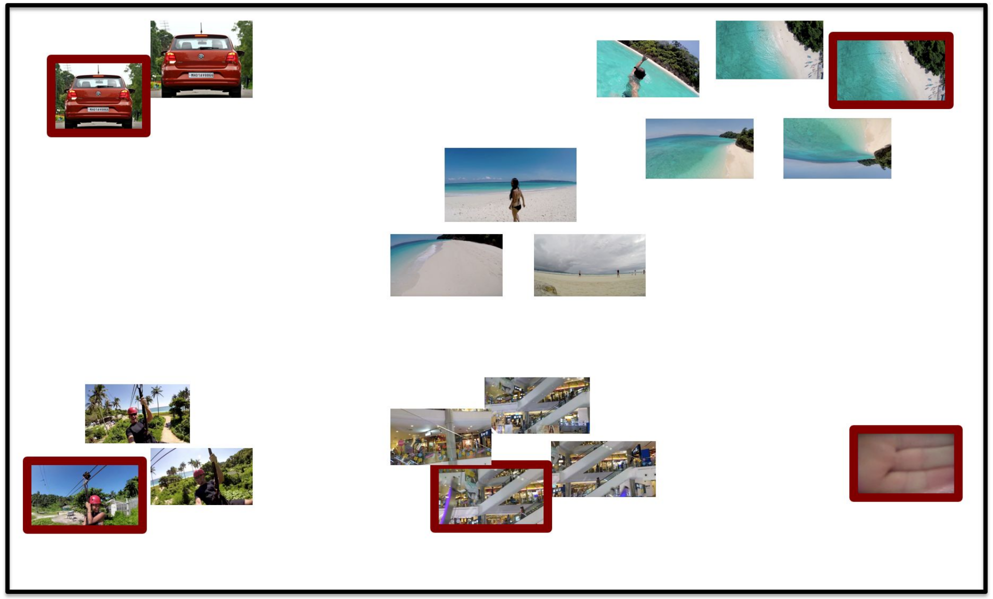
Dasgupta et al 2013, Chakraborty et al 2015

Diversity Functions: Dispersion



Dispersion Min	$\min_{k,l \in X, k \neq l} d_{kl}$
Dispersion Sum	$\sum_{k,l \in X} d_{kl}$
Dispersion Min-Sum	$\sum_{k \in X} \min_{l \in X} d_{kl}$

Dispersion Sum and Dispersion Min Not Submodular!

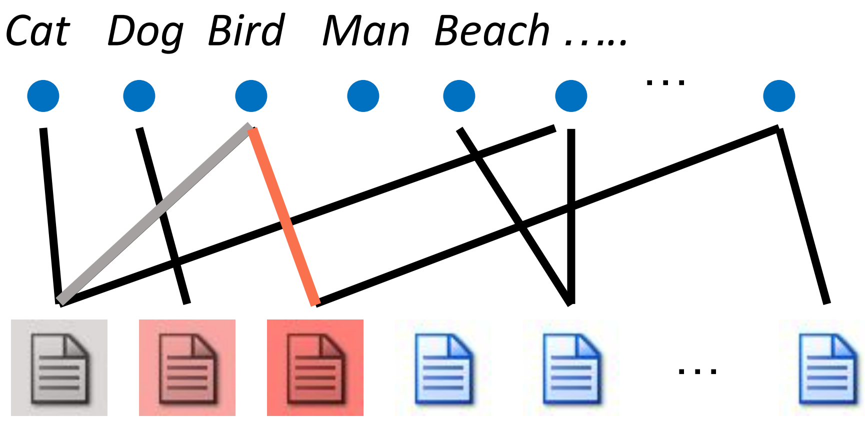


Diversity Functions

Picks items as different as possible!

Dasgupta et al 2013, Chakraborty et al 2015

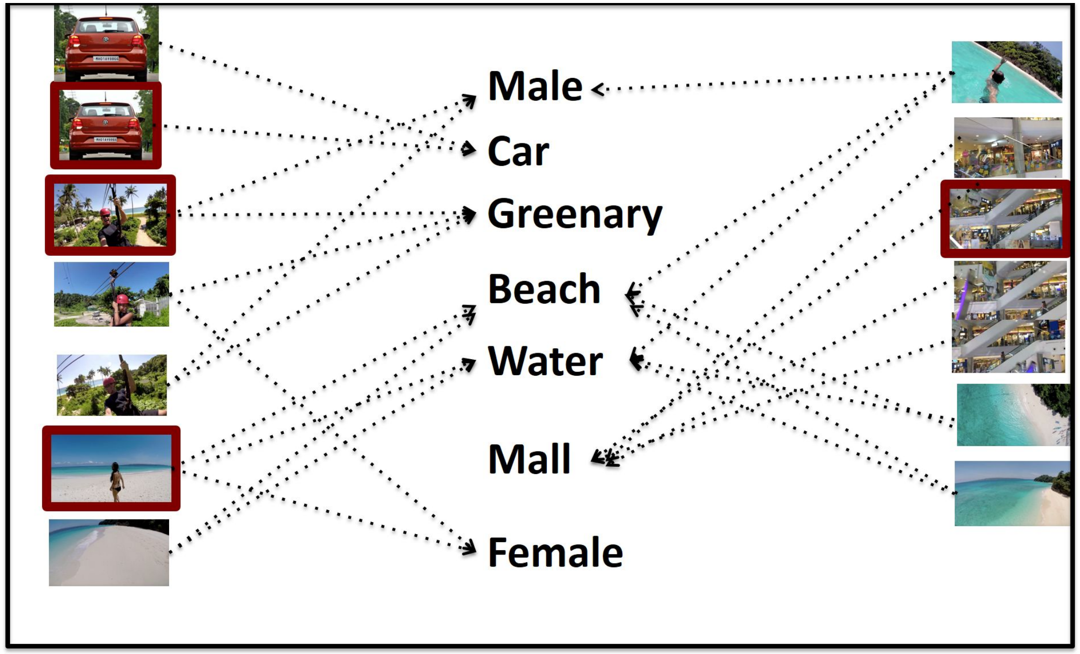
Coverage Functions



Set Cover Function

$$f(X) = w(\cup_{i \in X} U_i),$$

Concepts Covered by Instance i



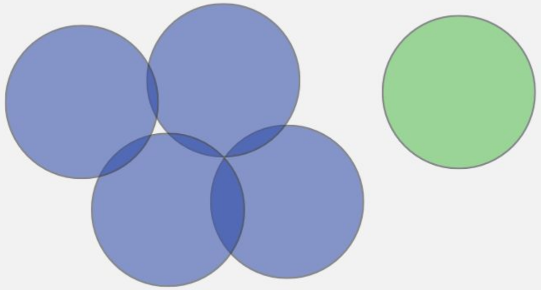
Coverage Functions

Select instances which "cover" all concepts

Wolsey et al 1982,

…

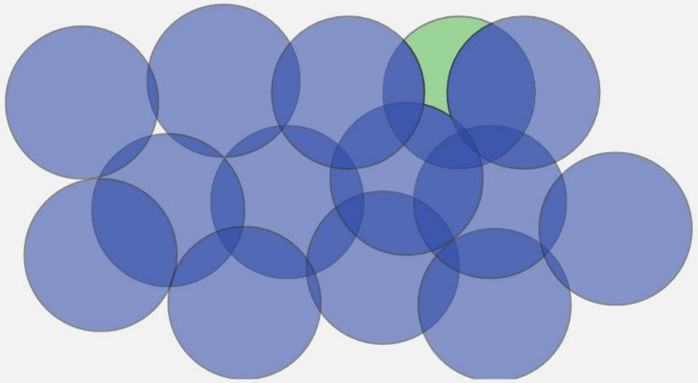
Why is Set Cover Submodular?



Gain (value) of v in context of A:

$$f(A \cup \{v\}) - f(A) = f(\{v\})$$

We get full value $f(\{v\})$ in this case since the area of v has no overlap with that of A.

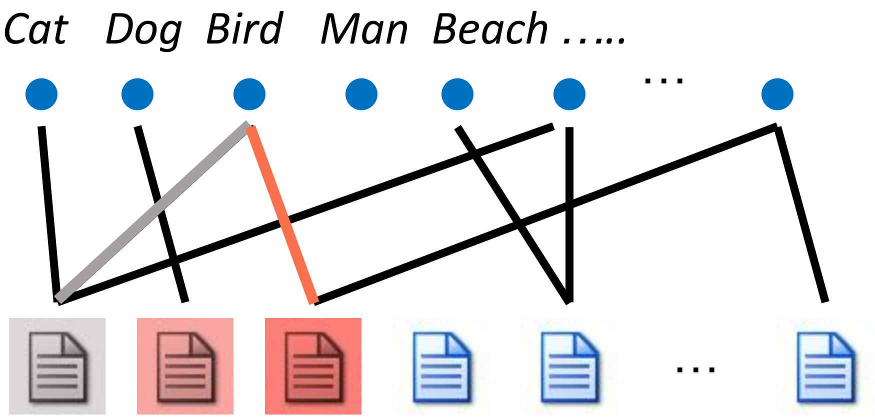


Incremental value of v in the context of $B \supset A$.

$$\begin{aligned} f(B \cup \{v\}) - f(B) &< f(\{v\}) \\ &= f(A \cup \{v\}) - f(A) \end{aligned}$$

So benefit of v in the context of A is greater than the benefit of v in the context of $B \supseteq A$.

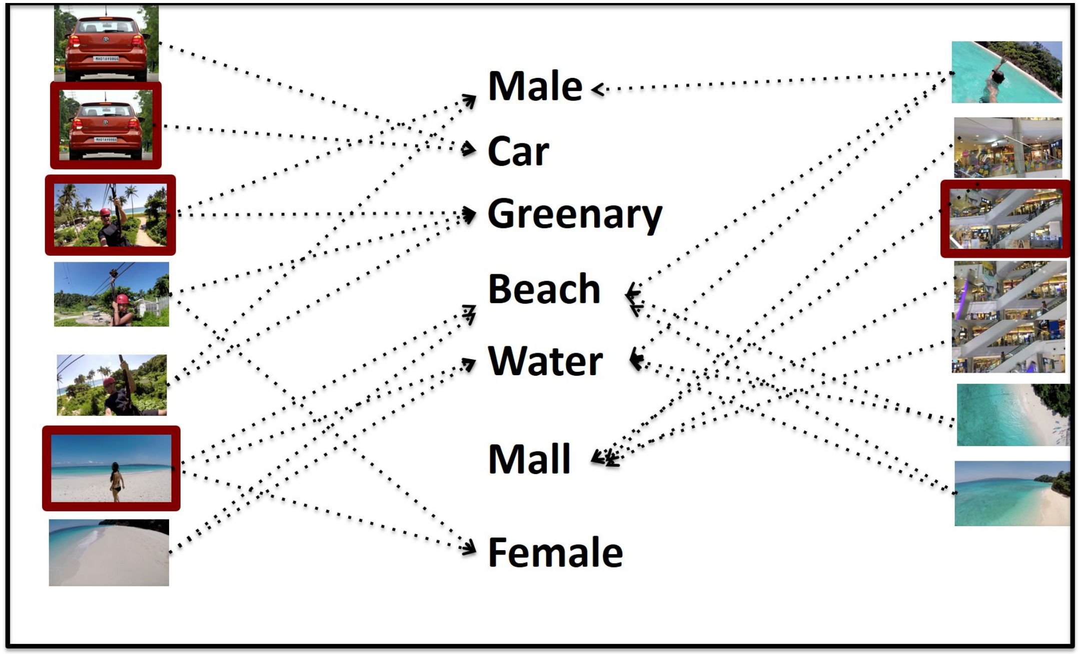
Coverage Functions



Probabilistic Set Cover Function

$$f(X) = \sum_{i \in \mathcal{U}} w_i [1 - \prod_{i \in X} (1 - p_{ij})].$$

Probability that Image i covers concept j

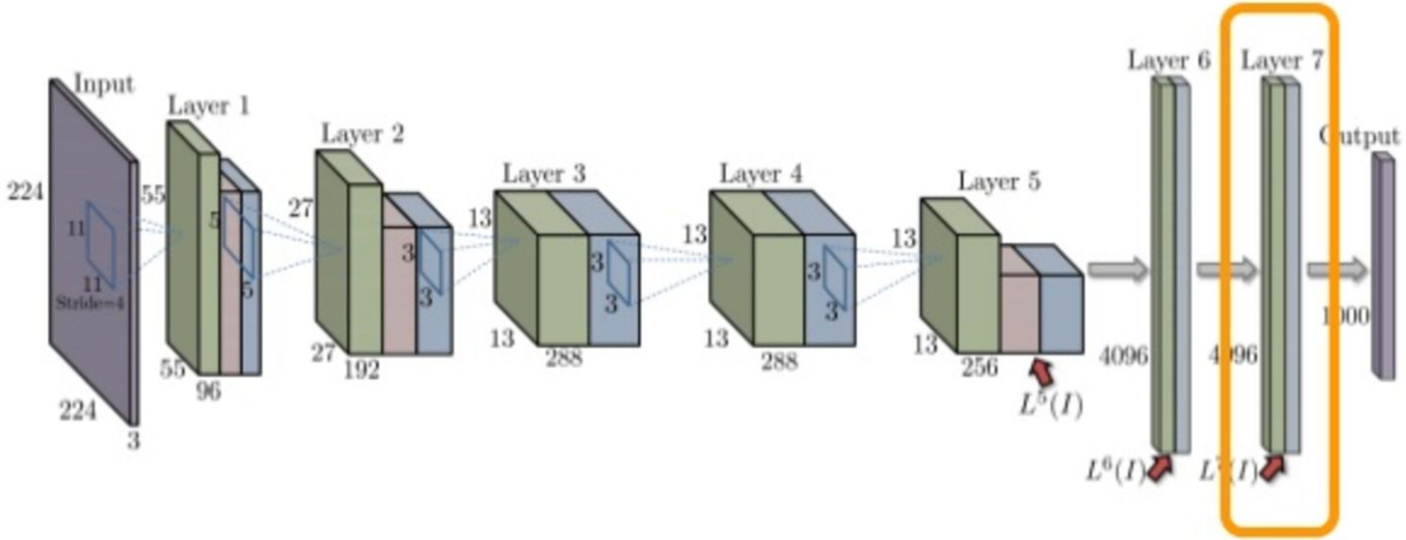


Coverage Functions

Allow for Probability of covering concepts

El-Arini & Guestrin 2013, …

Feature Based Functions



Achieve Uniformity in Feature Coverage

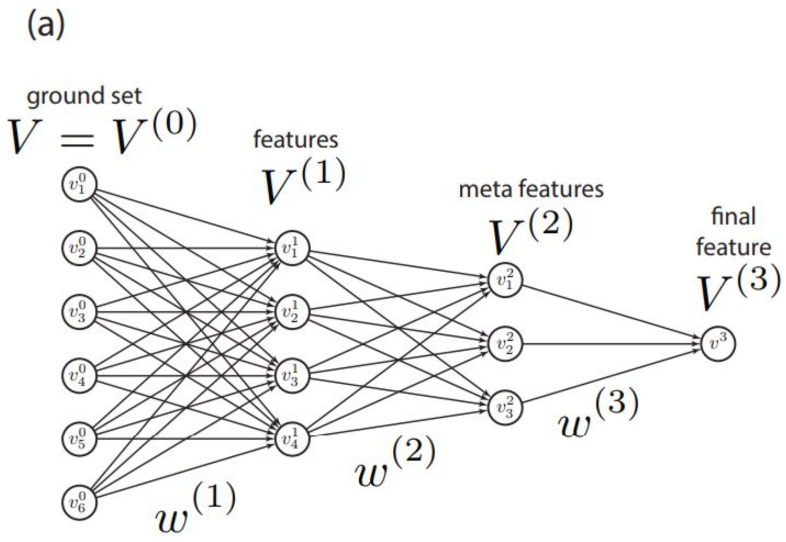
Feature Based Functions

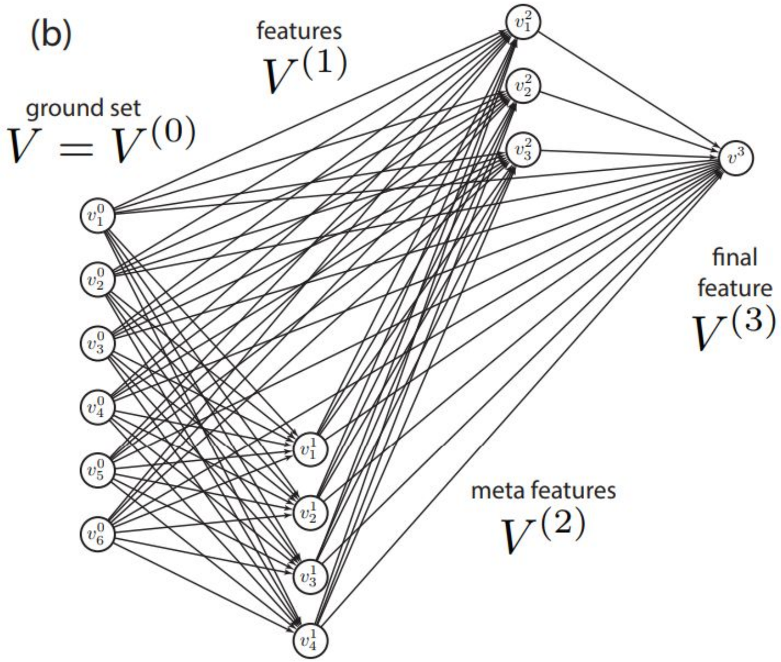
$$f_{\text{fea}}(S) = \sum_{u \in \mathcal{U}} g(m_u(S)).$$

Total Contribution of Feature u in the Set of Images S

Wei-Iyer et al 2014
…

Nested Feature Based Functions

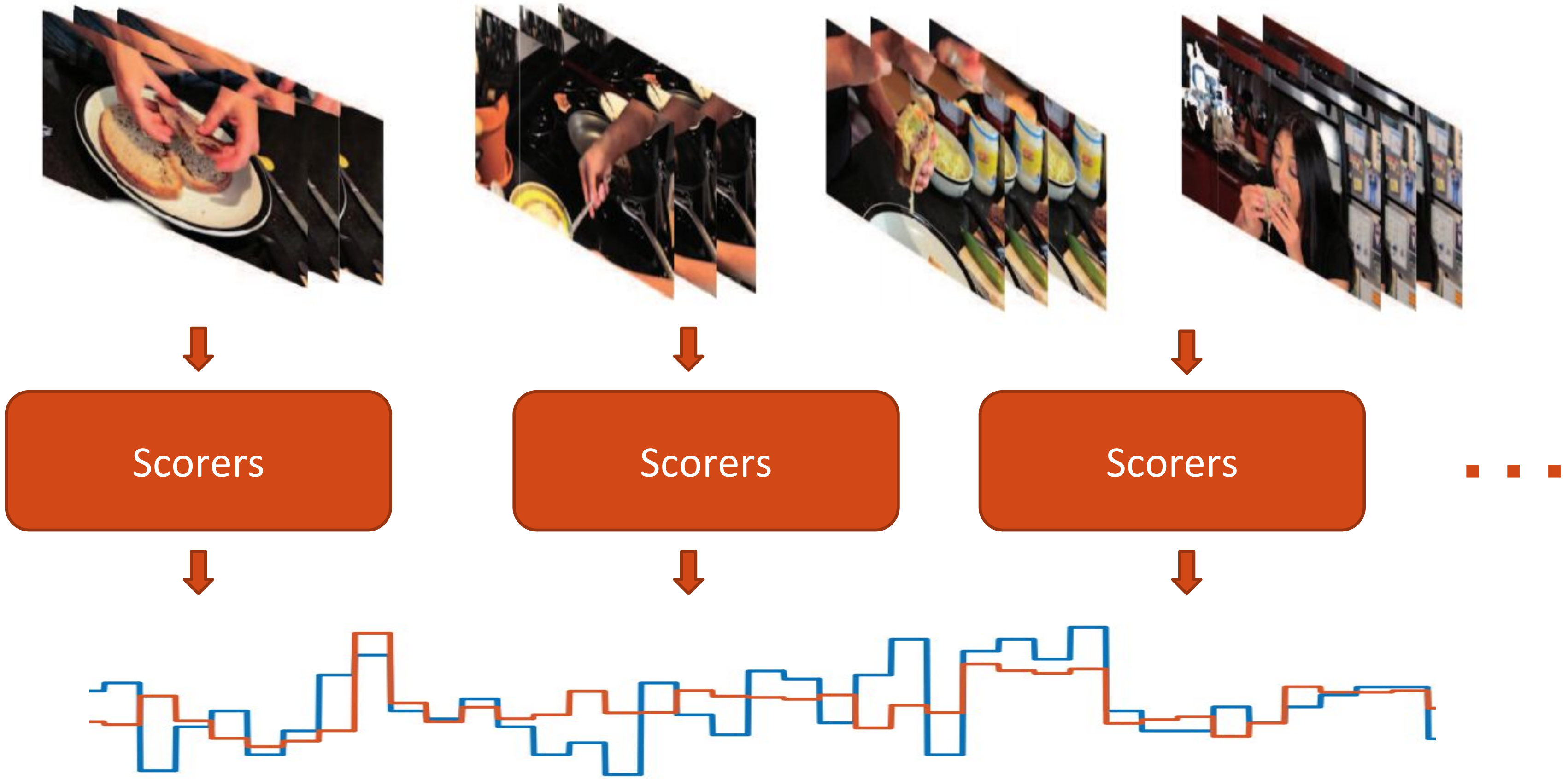




$$\bar{f}(A) = \phi_{v^K}\left(\sum_{v^{K-1}\in V^{(K-1)}} w^{(K)}_{v^K}(v^{K-1})\phi_{v^{K-1}}\left(\ldots\sum_{v^2\in V^{(2)}} w^{(3)}_{v^3}(v^2)\phi_{v^2}\left(\sum_{v^1\in V^{(1)}} w^{(2)}_{v^2}(v^1)\phi_{v^1}\left(\sum_{a\in A} w^{(1)}_{v^1}(a)\right)\right)\right)\right)$$

Bilmes-Bai 2015…

Importance Functions



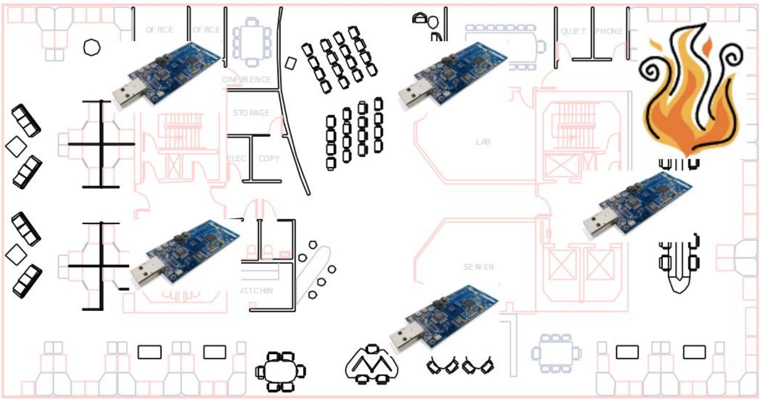
Importance Terms => Modular Functions

Information Functions

$X_1, \ldots, X_n$ discrete random variables: $X_e \in \{1, \ldots, m\}$

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

$$H(X_e) = \sum_{x \in \{1,\ldots,m\}} P(X_e = x) \log P(X_e = x)$$



$$\boxed{A \subset B, e \notin B \qquad F(A \cup e) - F(A) \geq F(B \cup e) - F(B)??}$$

$$\begin{aligned} H(X_{A \cup e}) - H(X_A) &= H(X_e | X_A) \\ &\leq H(X_e | X_B) \quad \text{“information never hurts”} \\ &= H(X_{B \cup e}) - H(X_B) \end{aligned}$$

discrete entropy is submodular!

Entropy
Mutual Information
Information Gain
…

Krause et al 2008,
…

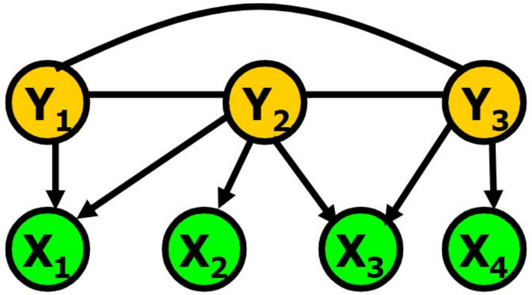
Information Gain as a Submodular Function

$Y_1,\ldots,Y_m$, X_1, …, X_n discrete RVs

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

- F(A) is NOT always submodular

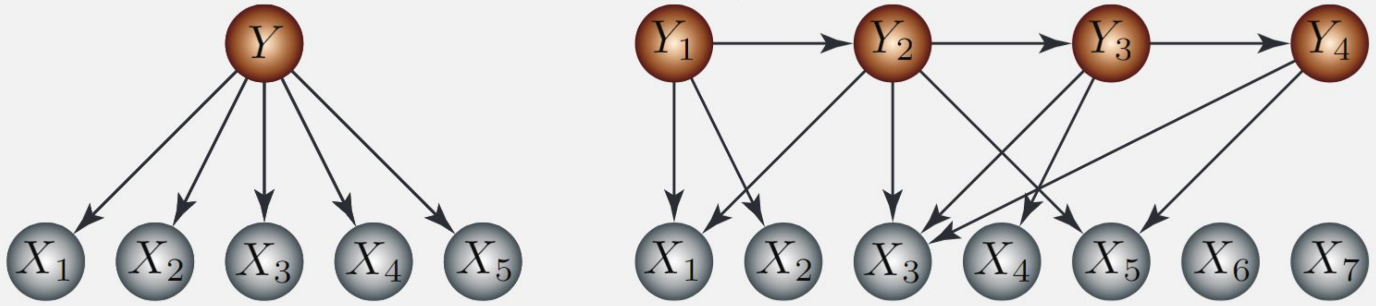
If X_i are all conditionally independent given Y, then F(A) is submodular! [Krause & Guestrin `05]



Proof:
"information never hurts"

Information Gain: Feature Selection

- Naïve Bayes property: $X_A \perp\!\!\!\perp X_B | Y$ for all A, B.



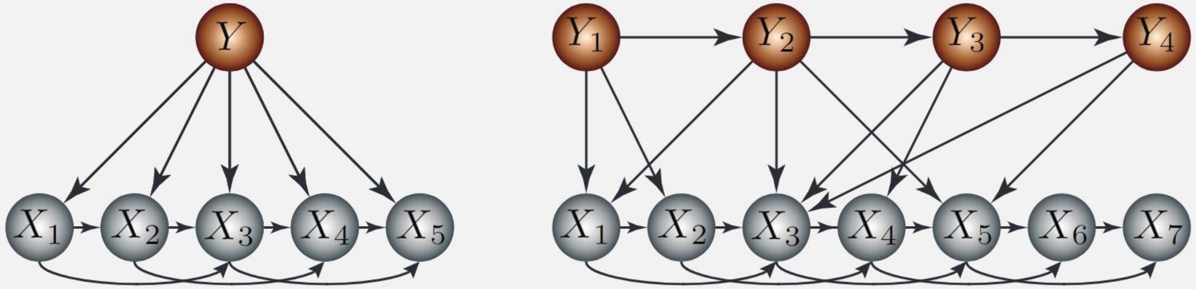
- When $X_A \perp\!\!\!\perp X_B | Y$ for all A, B (the Naïve Bayes assumption holds), then

$$f(A) = I(Y; X_A) = H(X_A) - H(X_A|Y) = H(X_A) - \sum_{a \in A} H(X_a|Y)$$

is submodular (submodular minus modular).

Feature Selection

- Naïve Bayes property fails:



- $f(A)$ naturally expressed as a difference of two submodular functions

$$f(A) = I(Y; X_A) = H(X_A) - H(X_A|Y),$$

which is a DS (difference of submodular) function.

- Alternatively, when Naïve Bayes assumption is false, we can make a submodular approximation (Peng-2005). E.g., functions of the form:

$$f(A) = \sum_{a \in A} I(X_a; Y) - \lambda \sum_{a,a' \in A} I(X_a; X_{a'}|Y)$$

where $\lambda \geq 0$ is a tradeoff constant.

Summary of Function Classes

Monotone Submodular:

1) Facility Location
2) Saturated Coverage
3) Feature Based Functions
4) Deep Submodular Functions
5) Set Cover
6) Probabilistic Set Cover
7) Complexity Function
8) Entropy

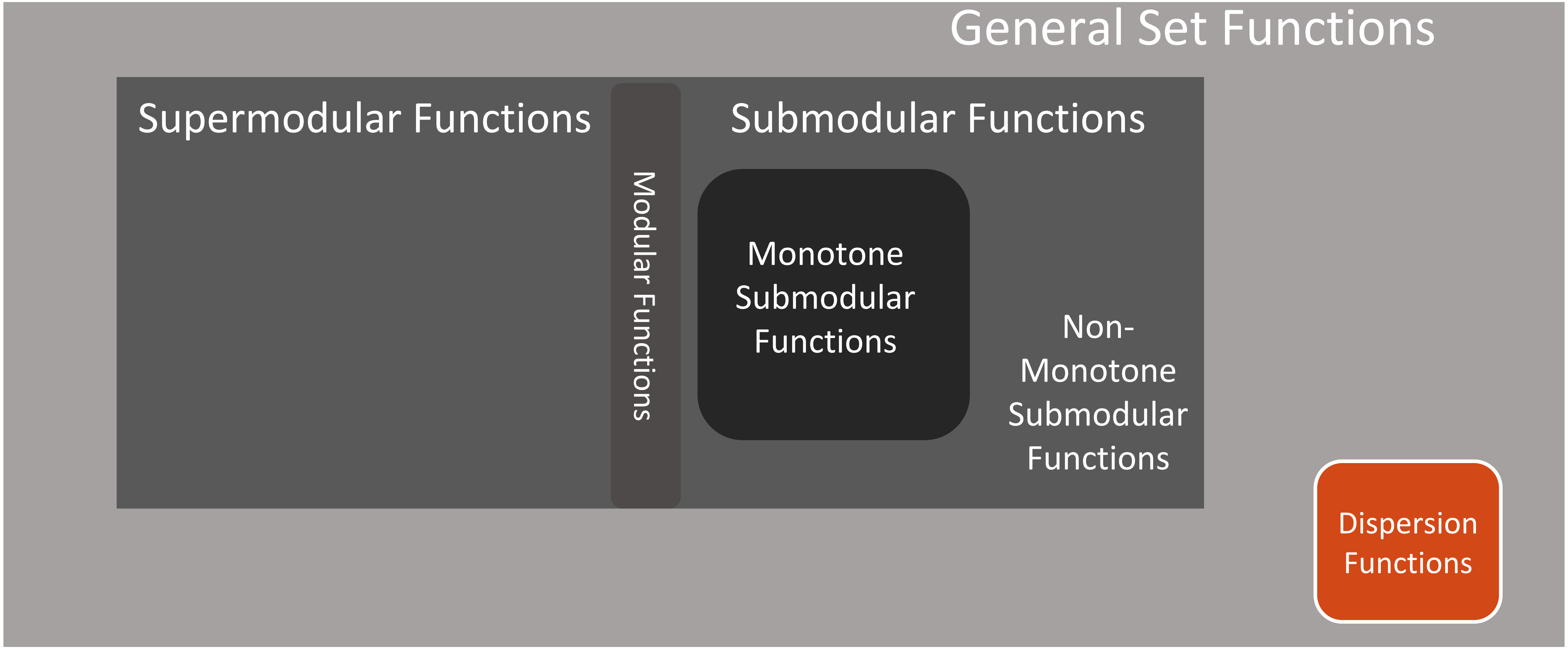
Non Monotone Submodular

1) Log Determinant (DPP)
2) Graph Cut ($\lambda < 2$)
3) Mutual Information (NB)

Non Submodular Functions

1) Disparity Min
2) Disparity Sum

Summary of Function Classes



Properties of Submodular Functions

- Convex Combinations of Submodular Functions are Submodular
- Intersections with Fixed Sets is Submodular (Restrictions)
- Unions with Fixed Sets is Submodular (Conditioning)
- Complement Functions are Submodular (Reflection)
- Minimum and Maximum of Submodular Functions
- Submodularity and Convexity
- Submodularity and Concavity

Convex Combinations of Submodular Functions

$F_1,\ldots,F_m$ submodular functions on V and $\lambda_1,\ldots,\lambda_m > 0$

Then: $F(A) = \sum_i \lambda_i F_i(A)$ is submodular

Submodularity closed under nonnegative linear combinations!

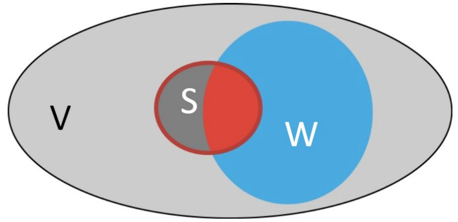
Extremely useful fact:

- $F_\theta(A)$ submodular ➔ $\sum_\theta P(\theta) F_\theta(A)$ submodular!
- Multicriterion optimization
- A basic proof technique! ☺

More Properties

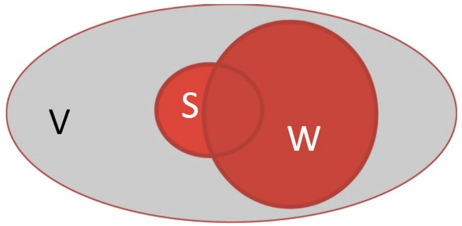
- Restriction: F(S) submodular on V, W subset of V

  Then $F'(S) = F(S \cap W)$ is submodular



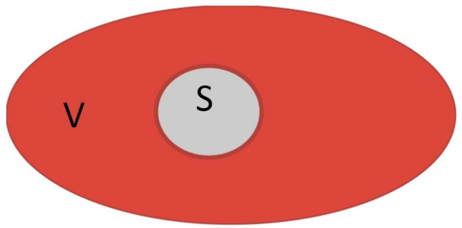
- Conditioning: F(S) submodular on V, W subset of V

  Then $F'(S) = F(S \cup W)$ is submodular



- Reflection: F(S) submodular on V

  Then $F'(S) = F(V \setminus S)$ is submodular



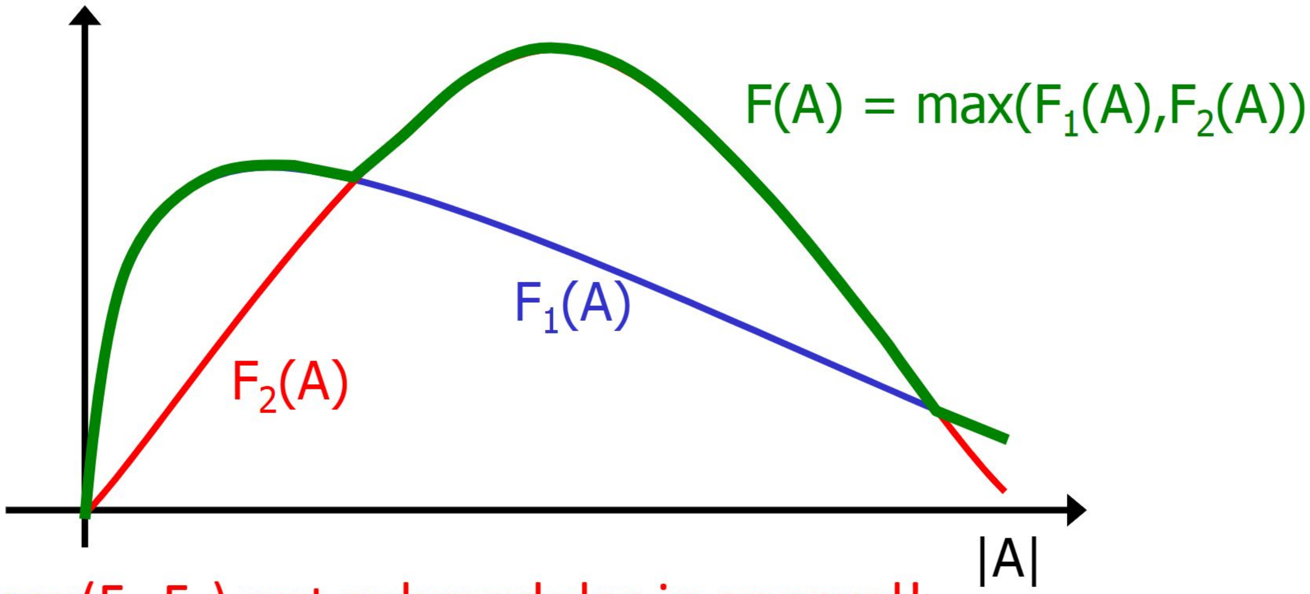
Concave over Submodular

- Concave over Monotone Submodular functions are Submodular:
  - Given a concave function g, and monotone submodular function f, g(f(S)) is submodular!
  - Hence given a positive modular function m, g(m(S)) is submodular and hence Feature based functions are submodular
  - This does not hold if m is positive and negative (i.e. f is non-monotone)

Maximum of Submodular Functions

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

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



max(F_1,F_2) not submodular in general!

Minimum of Submodular Functions

Well, maybe $F(A) = \min(F_1(A), F_2(A))$ instead?

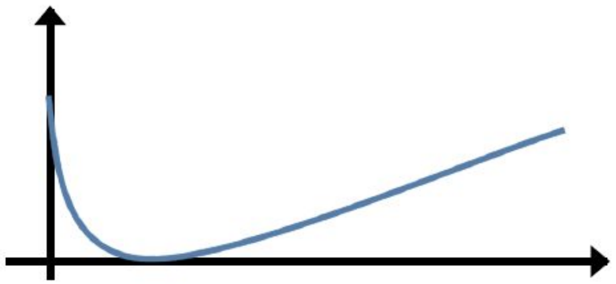
	$F_1(A)$	$F_2(A)$
{}	0	0
{a}	1	0
{b}	0	1
{a,b}	1	1

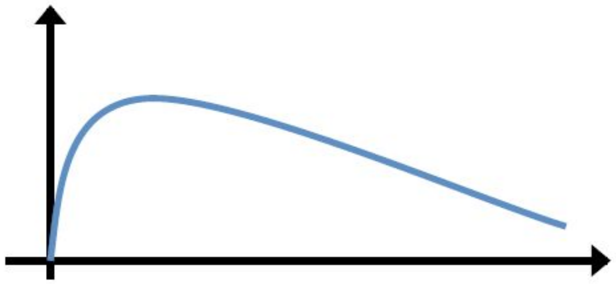
$$F(\{b\}) - F(\{\}) = 0$$
$$<$$
$$F(\{a,b\}) - F(\{a\}) = 1$$

$\min(F_1,F_2)$ not submodular in general!

Is Submodularity like Convexity or Concavity?

discrete convexity ....

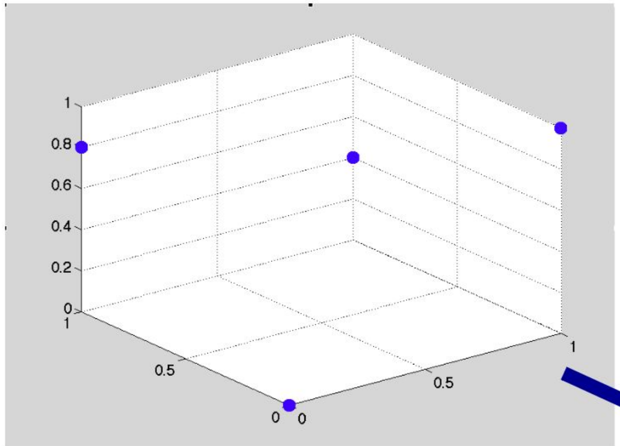


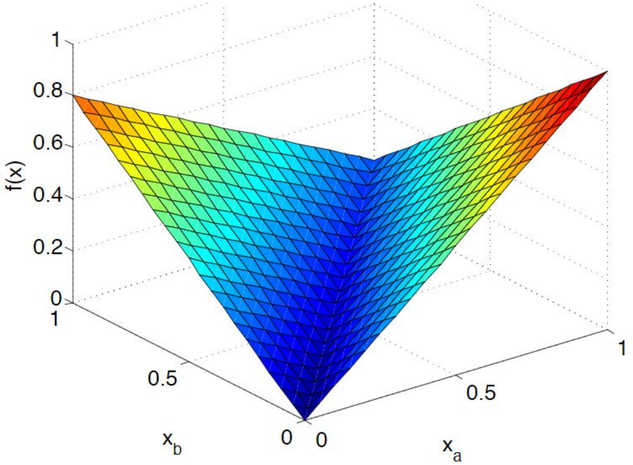


... or concavity?

Convex Aspects of Submodular Functions

- convex extension
  - duality
  - efficient minimization





But this is only half of the story…

Concave Aspects of Submodular Functions

- submodularity:

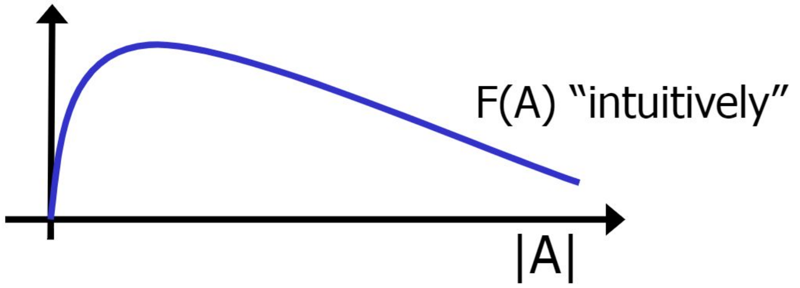
$$A \subseteq B, \quad s \notin B:$$

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

- concavity:

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

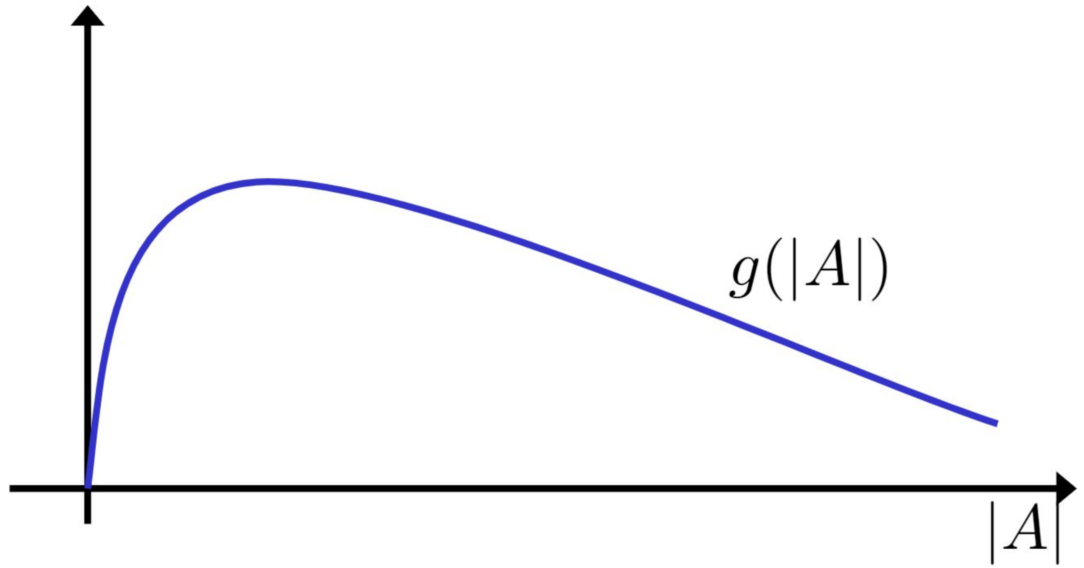
$$f(a+s) - f(a) \quad \geq \quad f(b+s) - f(b)$$



Concave Aspects of Submodular Functions

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

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



Two Faces of Submodularity