Aalto University
School of Science

CS-E4070 — Computational learning theory

Slide set 09 : submodular functions

Cigdem Aslay and Aris Gionis

Aalto University

spring 2019

motivation

- combinatorial optimization is concerned with problems of the form

$$\max\{f(S) : S \in \mathcal{F}\}$$

$$\min\{f(S) : S \in \mathcal{F}\}$$

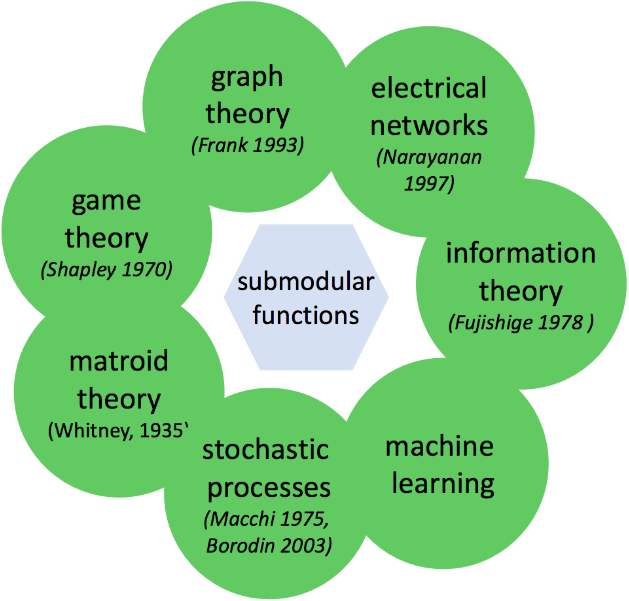
  where $f : 2^U \to \mathbb{R}$ is a function defined on a ground set U and $\mathcal{F}$ is the discrete set of feasible solutions.

- $\mathcal{F}$ could contain exponentially many solutions
- one way to make the problem tractable is to exploit the properties of f and $\mathcal{F}$

motivation

- in the continuous case, a function $g : \mathbb{R} \to \mathbb{R}$ can be
  - minimized efficiently if g is convex
  - maximized efficiently if g is concave
- submodularity plays the role of concavity / convexity in the discrete regime

motivation


graph theory (Frank 1993)
electrical networks (Narayanan 1997)
information theory (Fujishige 1978)
machine learning
stochastic processes (Macchi 1975, Borodin 2003)
matroid theory (Whitney, 1935)
game theory (Shapley 1970)
submodular functions

motivation

- submodular optimization can provide elegant solutions to many machine learning problems including
  - MAP inference in Markov Random Fields
  - variable selection / regularization
  - clustering
  - structure learning in graphical models
  - active learning
  - determinantal point processes

submodular (set) functions

- consider a ground set U
- a function $f : 2^U \to \mathbb{R}$ is submodular if

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$$

  for all $A, B \subseteq U$

- equivalently ("diminishing returns")

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

  for all $A \subseteq B \subseteq U$ and $x \in U \setminus B$

submodular (set) functions

- intuitively submodularity says that “you have more to gain from something new, if you have less to begin with"
- in some sense submodular functions are the discrete analogue of concave functions (but also has strong ties to convexity as we will see later)

submodular (set) functions

- a function $g : \mathbb{R} \to \mathbb{R}$ is concave if the derivative $g'(x)$ is non-increasing in x
- f is submodular if the “discrete derivative"

$$f(A \cup \{x\}) - f(A)$$

  is non-increasing in x
- suppose $h : \mathbb{N} \to \mathbb{R}$ and

$$f(A) = h(|A|)$$

  then f is submodular iff h is concave

submodular (set) functions

- closedness: submodularity is closed under nonnegative linear combinations

  $f_1, \cdots, f_m$ submodular functions on U and $\alpha_1, \cdots, \alpha_m > 0$. then

  $$f(A) = \sum_{i=1}^{m} \alpha_i f_i(A)$$

  is submodular

submodular (set) functions

- restriction: f is submodular on U, $W \subseteq U$. then

$$f'(A) = f(A \cap W)$$

  is submodular

- conditioning: f is submodular on U, $W \subseteq U$. then

$$f'(A) = f(A \cup W)$$

  is submodular

- reflection: f is submodular on U then

$$f'(A) = f(U \setminus A)$$

  is submodular

submodular (set) functions

- if f is submodular then $-f$ is supermodular
- f is modular (additive) if f is both supermodular and submodular. e.g.,

$$f(A) = \sum_{i \in A} w(i)$$

where $w(i)$ is some constant, $\forall i \in U$

submodular (set) functions

may or may not satisfy the following properties

- non-negative : $f(A) \geq 0$ for all $A \subseteq U$
- monotone : $f(A) \leq f(B)$ for all $A \subseteq B \subseteq U$
- symmetric : $f(A) = f(U \setminus A)$ for all $A \subseteq U$

examples

- coverage in set systems

  $\Rightarrow$ monotone and non-negative

- cut functions in undirected graphs and hypergraphs

  $\Rightarrow$ symmetric and non-negative

- cut functions in directed graphs

  $\Rightarrow$ non-negative

example: coverage in set systems

- $V = \{1, \dots, n\}$
- $U = \{S_1, \dots, S_m\}$ a collection of subsets of V
- function $f : 2^U \to \mathbb{R}_+$
- coverage :

$$f(A) = |\cup_{i \in A} S_i|$$

- weighted coverage :

$$w : U \to \mathbb{R}_+ \text{ and } f(A) = \sum_{x \in \cup_{i \in A} S_i} w(x)$$

example: set cover and maximum coverage

- set cover: choose the set A of minimum cardinality such that all elements of U is covered

$$f(A) = |\cup_{i \in A} S_i| = |U|$$

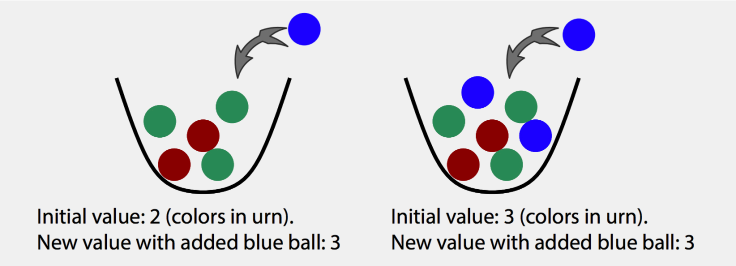
- maximum coverage: given $k \in \mathbb{N}^+$, select a set A of cardinality k such that

$$f(A) = |\cup_{i \in A} S_i|$$

is maximized

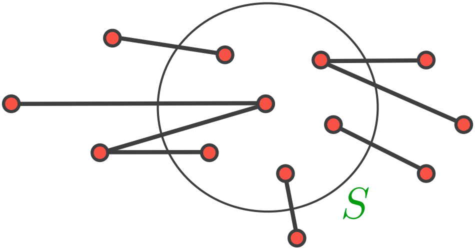
example: nr of colors of balls in an urn

- consider an urn containing colored balls.
- given a set S of balls, $f(S)$ counts the number of *distinct* colors in S



example: cut in graphs

- consider undirected graph $G = (V, E)$
- cut function $f : 2^V \to R_+$ defined as $f(S) = |E(S, V \setminus S)|$



example: cut in graphs

- define $f(S)$ from

$$f(S) = \sum_{(i,j) \in E} f_{i,j}(S \cap \{i, j\})$$

  where $f_{i,j}$ is the cut function in subgraph $\{i, j\}$

- $f_{i,j}$ is submodular so is f
- follows from closedness and restriction properties

example: entropy

- entropy of a discrete random variable X

$$H(X) = -\sum_{x} \mathbf{Pr}(x) \log \mathbf{Pr}(x)$$

- entropy of X conditioned on Y

$$H(X \mid Y) = -\sum_{x,y} \mathbf{Pr}(x,y) \log \frac{\mathbf{Pr}(x,y)}{\mathbf{Pr}(y)}$$

example: entropy

- given n random variables $U = \{X_i\}_{i \in [1,n]}$, define

$$f(A) = H(X_A)$$

  to be the joint entropy of the variables indexed by A.
- f is submodular

example: entropy

- suppose that $A \subseteq B$, $X_e \in U$, then

$$
\begin{aligned}
f(A \cup \{X_e\}) - f(A) &= H(X_A, X_e) - H(X_A) \\
&= H(X_e \mid X_A) \text{ “\textit{information never hurts}"} \\
&\geq H(X_e \mid X_B)
\end{aligned}
$$

- information never hurts: conditioning on data never increases uncertainty
- mutual information is also submodular

$$I(A) = f(A) + f(U \setminus A) - f(U)$$

the maximization problem

- given submodular function $f : 2^U \to \mathbb{R}$
  find $S \subseteq X$ to maximize $f(S)$
  subject to constraints

- value-oracle model
- generalizes many interesting **NP**-hard problems
- minimization problem is polynomial in the unconstrained case (e.g., min-cut)

monotone submodular functions

- $f(U)$ trivial maximizer in the unconstrained case
- more interesting to maximize under cardinality constraints

- find $S \subseteq U$ subject to $|S| \leq k$ that maximizes $f(S)$

- MAX k-COVER is a special case
- greedy gives $(1 - 1/e)$ approximation
  [Nemhauser et al., 1978]
- no better approximation unless **P**=**NP**

the greedy algorithm

1. $S \leftarrow \emptyset$
2. while $|S| < k$
3. $\quad i \leftarrow \arg\max_j f(S \cup \{j\}) - f(S)$
4. $\quad S \leftarrow S \cup \{i\}$
5. return S

analysis of the greedy

- S^* : the optimal solution
- $S_j = \{x_1, \cdots, x_j\}$: the first j elements picked by the greedy
- let $f(x_j \mid S_{j-1})$ denote the marginal gain of adding the j-th element to S_{j-1}

$$f(x_j \mid S_{j-1}) = f(S_j) - f(S_{j-1})$$

- hence

$$f(S) = \sum_{j=1}^{k} f(x_j \mid S_{j-1})$$

analysis of the greedy

- claim:

$$f(x_j \mid S_{j-1}) \geq \frac{f(S^*) - f(S_{j-1})}{k}$$

- proof. first we need to state a property of submodular functions:
  - if f is submodular, then the following holds $\forall A, B \subseteq U$:

$$f(A) \leq f(B) + \sum_{x \in A \setminus B} f(x \mid B) - \sum_{x \in B \setminus A} f(x \mid A \cup B \setminus \{x\})$$

  (see Proposition 2.1 in [Nemhauser et al., 1978] for all similar properties)

analysis of the greedy

- proof (cont'd). using this property, we have

$$f(S^*) \leq f(S_{j-1}) + \sum_{x \in S^* \setminus S_{j-1}} f(x \mid S_{j-1}) - \sum_{x \in S_{j-1} \setminus S^*} f(x \mid S^* \cup S_{j-1} \setminus \{x\})$$

  which further implies (due to monotonicity of f):

$$f(S^*) - f(S_{j-1}) \leq \sum_{x \in S^* \setminus S_{j-1}} f(x \mid S_{j-1})$$

analysis of the greedy

- proof (cont'd). using also the fact that $\forall x \in V \setminus S_{j-1}$:

  $$f(x_j \mid S_{j-1}) \geq f(x \mid S_{j-1})$$

  since otherwise x_j wouldn't be selected by greedy, we have:

  $$\begin{aligned} f(S^*) - f(S_{j-1}) &\leq \sum_{x \in S^* \setminus S_{j-1}} f(x \mid S_{j-1}) \\ &\leq k \cdot f(x_j \mid S_{j-1}) \end{aligned}$$

- we have just proved our claim

analysis of the greedy

- continuing the analysis of greedy, we have

$f(S^*) - f(S_j) \leq (1 - 1/k)^j f(S^*)$ (by induction)

$f(S^*) - f(S_k) \leq (1 - 1/k)^k f(S^*)$

$$\begin{aligned} f(S_k) &\geq (1 - (1 - 1/k)^k) f(S^*) \\ &\geq \left(1 - \frac{1}{e}\right) f(S^*) \end{aligned}$$

widely occurs in algorithmic data mining

- example : maximize the spread of influence in social networks [Kempe et al., 2003]

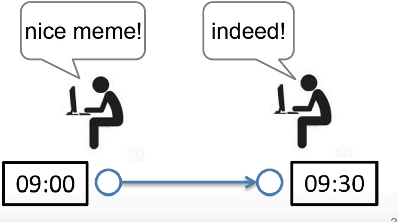
influence maximization

- Social Influence Induced Viral Phenomena

Grumpy Cat

- 25K+ votes in Reddit (< 1 day)
- 1M+ views in Imgur
- 300+ variants in Reddit
- 100+ Quickmeme macros

(< 2 days)



Influence in Online Social Networks

Viral Marketing*

exploit the "*word of mouth*" effect in a social network to achieve marketing goals through self-replicating viral processes

Hotmail

- Attached a promotional message with a clickable URL for free sign up
- Merely spent $50K
- 12M users signed up within the first 18 months

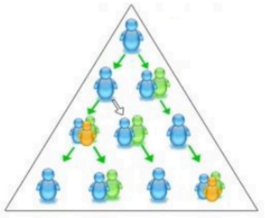
Gmail by Google

- Sign-up to the service only through invitation from a friend
- No money spent on marketing
- Resulted in bidding on Ebay for invites

influence maximization

Influence Maximization

Discrete Optimization Problem*



- Given
  - a directed social network G = (V,E)
  - a propagation model m
  - a cardinality budget k
- Define
  - S: initial set of k (seed) nodes to start the propagation
  - $\sigma_m(S)$: expected size of the influence propagation from S
- Find

$$S^* = \underset{S \subseteq V, |S| = k}{\operatorname{argmax}} \; \sigma_m(S)$$

non-monotone functions

- unconstrained version becomes interesting
- find $S \subseteq X$ to maximize $f(S)$

- generalizes MAX-CUT
- what do we know about approximation?
- random set gives $1/2$ ($1/4$ for MAX-DICUT)
- SDP gives 0.878 (0.796 for MAX-DICUT)
  major breakthrough [Goemans and Williamson, 1995]
- 0.53 by spectral approach [Trevisan, 2012]

unconstrained problem

[Feige et al., 2011]

- first constant-factor approximations for non-negative submodular functions
- simple algorithms: randomized / deterministic, non-adaptive / adaptive
- $1/2$ approx for symmetric functions
- $2/5 = 0.4$ approx for the non-negative functions
- lower bound: better than $1/2$ approx requires exponential number of value queries

unconstrained problem

[Feige et al., 2011]

- pick a random set

  $1/4$ for non-negative function (on expectation)

  $1/2$ for symmetric function (on expectation)

- local search
  - initialize S to best singleton
  - S = local optimum (add or delete elements)
  - return the best of S and $U \setminus S$

  $1/3$ approx for non-negative function

  $1/2$ for non-negative symmetric function

random set analysis

- for $A \subseteq U$, $A(p)$ is a random set where each element of A is selected with prob p
- algorithm returns $R = U(1/2)$
- lemma I

$$E[f(A(p))] \geq (1-p)\,f(\emptyset) + p\,f(A)$$

  can prove by induction on the size of A
  and using the submodularity property

- lemma II

$$\begin{aligned} E[f(A(p) \cup B(q))] \quad \geq \quad & (1-p)(1-q)\,f(\emptyset) + \\ & p(1-q)\,f(A) + \\ & (1-p)q\,f(B) + \\ & pq\,f(A \cup B) \end{aligned}$$

  to prove use lemma I

random set analysis

- algorithm returns

$$R = U(1/2) = S^*(1/2) \cup \overline{S^*}(1/2)$$

- by applying lemma II

$$\begin{aligned} E[f(R)] &= E[f(S^*(1/2) \cup \overline{S^*}(1/2))] \\ &= \frac{1}{4}f(\emptyset) + \frac{1}{4}f(S^*) + \frac{1}{4}f(\overline{S^*}) + \frac{1}{4}f(U) \end{aligned}$$

- gives $1/4$ for non-negative and $1/2$ for symmetric function

unconstrained problem

[Feige et al., 2011]

- local search
  - initialize S to best singleton
  - S = local optimum (add or delete elements)
  - return the best of S and $U \setminus S$

$1/3$ approx for non-negative function

$1/2$ for non-negative symmetric function

analysis of local search

- lemma if S is a local optimum then
  $f(S) \geq f(T)$ for all $S \subseteq T$ and $T \subseteq S$

- proof
  take $S \subseteq T$ and consider $S = X_0 \subseteq \dots X_\ell = T$
  by submodularity and local optimality

$$0 \geq f(S \cup \{x_i\}) - f(S) \geq f(X_i) - f(X_{i-1})$$

  summing up gives $0 \geq f(X_\ell) - (X_0)$ or $f(S) \geq (T)$

- corollary
  for optimum S^* and local optimum S it is
  $f(S) \geq f(S \cup S^*)$ and $f(S) \geq f(S \cap S^*)$

analysis of local search (cont)

- it is
$$f(S) \geq f(S \cup S^*) \quad \text{and} \quad f(S) \geq f(S \cap S^*)$$

- by submodularity and non-negativity
$$f(S \cup S^*) + f(U \setminus S) \geq f(S^* \setminus S) + f(U) \geq f(S^* \setminus S)$$
$$f(S \cap S^*) + f(S^* \setminus S) \geq f(S^*) + f(\emptyset) \geq f(S^*)$$

- combining we get
$$2f(S) + f(U \setminus S) \geq f(S^*)$$

- and so
$$\max\{f(S), f(U \setminus S)\} \geq \frac{1}{3} f(S^*)$$

unconstrained problem

[Buchbinder et al., 2015]

- tight $1/2$ approximation for general non-negative submodular function
- randomized algorithm, approximation $1/2$
- deterministic algorithm, approximation $1/3$

deterministic algorithm

[Buchbinder et al., 2015]

Algorithm 1: DeterministicUSM$(f, \mathcal{N})$

```
X_0 ← ∅, Y_0 ← N.
for i = 1 to n do
    a_i ← f(X_{i-1} ∪ {u_i}) − f(X_{i-1}).
    b_i ← f(Y_{i-1} \ {u_i}) − f(Y_{i-1}).
    if a_i ≥ b_i then X_i ← X_{i-1} ∪ {u_i}, Y_i ← Y_{i-1}.
    else X_i ← X_{i-1}, Y_i ← Y_{i-1} \ {u_i}.
return X_n (or equivalently Y_n).
```

randomized algorithm

[Buchbinder et al., 2015]

Algorithm 2: RandomizedUSM$(f, \mathcal{N})$

$X_0 \leftarrow \emptyset$, $Y_0 \leftarrow \mathcal{N}$.
for i = 1 *to* n **do**
    $a_i \leftarrow f(X_{i-1} \cup \{u_i\}) - f(X_{i-1})$.
    $b_i \leftarrow f(Y_{i-1} \setminus \{u_i\}) - f(Y_{i-1})$.
    $a'_i \leftarrow \max\{a_i, 0\}$, $b'_i \leftarrow \max\{b_i, 0\}$.
    with probability $a'_i/(a'_i + b'_i)^*$ **do**:
    $X_i \leftarrow X_{i-1} \cup \{u_i\}$, $Y_i \leftarrow Y_{i-1}$.
    else (with the compliment probability $b'_i/(a'_i + b'_i)$)
    do: $X_i \leftarrow X_{i-1}$, $Y_i \leftarrow Y_{i-1} \setminus \{u_i\}$.
return X_n (or equivalently Y_n).

* If $a'_i = b'_i = 0$, we assume $a'_i/(a'_i + b'_i) = 1$.

max-sum diversification

[Borodin et al., 2012]

- U is a ground set
- $d : U \times U \to \mathbb{R}$ is a metric distance function on U
- $f : 2^U \to \mathbb{R}$ is a submodular function

- we want to find $S \subseteq U$ such that
  $\phi(S) = f(S) + \lambda \sum_{u,v \in S} d(u, v)$ is maximized and
  $|S| \leq k$

max-sum diversification

[Borodin et al., 2012]

- consider $S \subseteq U$ and $x \in U \setminus S$
- define the following types of marginal gain

$$d_x(S) = \sum_{v \in S} d(x, v)$$

$$f_x(S) = f(S \cup \{x\}) - f(S)$$

$$\phi_x(S) = \frac{1}{2} f_x(S) + \lambda d_x(S)$$

- greedy algorithm on marginal gain $\phi_x(S)$ gives factor 2 approximation

max-sum diversification – the greedy

[Borodin et al., 2012]

1. $S \leftarrow \emptyset$
2. while $|S| < k$
3. $\quad i \leftarrow \arg\max_{\{j \in U \setminus S\}} \phi_j(S)$
4. $\quad S \leftarrow S \cup \{i\}$
5. return S

references

Borodin, A., Lee, H. C., and Ye, Y. (2012).
Max-sum diversification, monotone submodular functions and dynamic updates.
In *Proceedings of the 31st symposium on Principles of Database Systems*, pages 155–166. ACM.

Buchbinder, N., Feldman, M., Seffi, J., and Schwartz, R. (2015).
A tight linear time (1/2)-approximation for unconstrained submodular maximization.
SIAM Journal on Computing, 44(5):1384–1402.

Feige, U., Mirrokni, V. S., and Vondrak, J. (2011).
Maximizing non-monotone submodular functions.
SIAM Journal on Computing, 40(4):1133–1153.

references (cont.)

- Goemans, M. X. and Williamson, D. P. (1995).
  Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming.
  Journal of the ACM (JACM), 42(6):1115–1145.

- Kempe, D., Kleinberg, J., and Tardos, E. (2003).
  Maximizing the spread of influence through a social network.
  In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM Press.

- Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L. (1978).
  An analysis of approximations for maximizing submodular set functions I.
  Mathematical Programming, 14(1):265–294.

references (cont.)

Trevisan, L. (2012).
Max cut and the smallest eigenvalue.
SIAM Journal on Computing, 41(6):1769–1786.