אלגוריתמים ברשתות - 2011

גיל אלון

תכנות לינארי מנקודת המבט של כיסוי קבוצות

- Consider a matrix $A \in M_{m\times n}$ and vectors $\begin{cases} x \in \mathbb{R}^n \\ b \in \mathbb{R}^m \end{cases}$
- System of equations $Ax = b$

  solution: an (affine) subspace of $\mathbb{R}^n$

  $z + \{x : Ax = b\}$ where $Az = b$

  algorithm: Gauss elimination

What about inequalities?

$$\begin{bmatrix} Ax = b \\ x \geq 0 \end{bmatrix} \qquad \begin{bmatrix} Ax \leq b \end{bmatrix} \qquad \begin{bmatrix} x_1 + 2x_2 + 3x_3 \leq 4 \\ x_1 + x_2 + x_3 = 1 \\ x_1 \geq 0 \\ x_2 \geq 0 \\ x_3 \geq 0 \end{bmatrix}$$

Question #1 Find an efficient algorithm for the following problem:

Input: a set Π of constraints

((in)equalities over variables $x_1, \dots, x_n \in \mathbb{R}$)

[Each constraint is of the form:

$$a_{i1} \cdot x_1 + a_{i2} x_2 + \cdots + a_{in} x_n \ \substack{=\\ \leq \\ \geq}\ b_i$$]

Output: "not feasible" OR $\vec{x} \in \mathbb{R}^n$ that satisfies all the constraints

---

Question #2: Find an efficient algorithm for the following problem:

Input: $m \times n$ matrix $A \in M_{m\times n}$

vector $b \in \mathbb{R}^m$

vector $c \in \mathbb{R}^n$

Goal: find $x \in \mathbb{R}^n$ that minimizes $c^t x \triangleq \sum_j c_j x_j$

subject to $A \cdot x = b$, $x \geq 0$.

Output: "not feasible" or $x^* = \operatorname{argmin}\{c^t x \mid Ax = b,\ x \geq 0\}$.

Example: $\min\ x_1 + x_2$

s.t. $x_1 + 10x_2 \geq 21$, $x_1 \geq 0$, $x_2 \geq 0$.

Both questions are about linear programming !

- Q1 : Is a polyhedron empty ?
- Q2 : Linear programming in standard form.

Lots of : structure

& algorithms ...

Combinatorial Optimization Problems

find "best" solution for a combinatorial problem (as opposed to a "numerical problem").

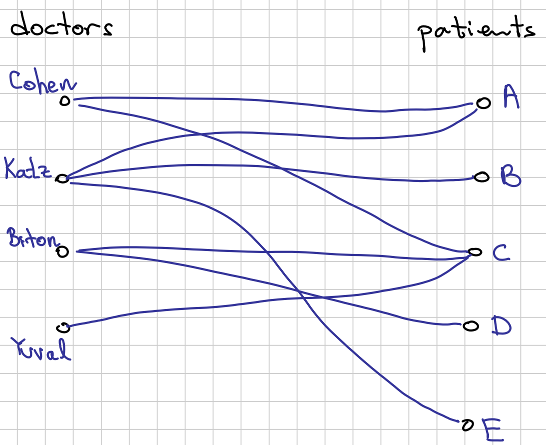
Examples : Set Cover

Vertex Cover

Minimum weight matching

Maximum flow

Set Cover - Example



Rules:

Every doctor can serve a subset of the patients.

Goal:

Choose fewest # doctors, so that each patient is served.

Definition : Set-Cover

Input : a set of elements $[1..n]$ patients

a family of subsets : $S_1, \dots, S_m \subseteq [1..n]$ doctors

$\{S_{i_1}, S_{i_2}, \dots, S_{i_k}\}$ is a Set Cover if

$S_{i_1} \cup S_{i_2} \cup \dots \cup S_{i_k} = [1..n]$ every patient served by a doctor

Goal : find a set cover of minimum size.

History: Set Cover

1) Lots of applications & special cases.

2) Among the first problems proven to be NP-Complete [Karp]

3) Greedy algorithm gives $H_n \approx \ln n$ approximation.

4) NP-Complete to find "better" approximations [Feige, RS]

What is the greedy algorithm for Set Cover?

What does linear programming have to say?

Set Cover: the greedy algorithm

Input: $S_1, \dots, S_m \subseteq [1..n]$

goal: find smallest set $\{S_{i_1}, \dots, S_{i_k}\}$ s.t. $S_{i_1} \cup \dots \cup S_{i_k} = [1..n]$

method: (Greedy)

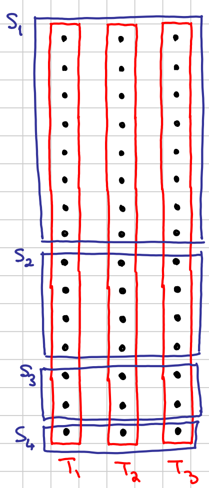
$C \leftarrow \emptyset$

while $C \neq [1..n]$ do

- $S \leftarrow \operatorname{argmax}\{\, |S_i \setminus C| : 1 \le i \le m \,\}$
- $C \leftarrow C \cup S$ (add S to the cover)

Question: How many sets in the cover (compared to an optimal solution)?

Greedy - Example



Greedy picks:

$\{S_1, S_2, S_3, S_4\}$

But OPT picks

$opt = \{T_1, T_2, T_3\}$

So $|Greedy| = 4$

& $|OPT| = 3$.

Question 1: Give a sequence of examples where

Greedy/OPT $\ge \ln n$ (easier: $\ge \Omega(\ln n)$)

Set Cover: the greedy algorithm (CONT)

Theorem [Johnson]

greedy returns an (H_n)-approx.

$H_n = 1 + \frac{1}{2} + \dots + \frac{1}{n}$

$\le 1 + \ln n$

$(|Greedy| \le (1 + \ln n) \cdot OPT)$

proof – not hard, we will prove using LP & duality...

An Integer Programming formulation for Set Cover

For each set S, define a variable $x_S \in \{0,1\}$

$$\min \sum_S x_S$$
$$\text{s.t.} \sum_{\{S:\, e\in S\}} x_S \geq 1 \quad \forall \text{element } e \in [1..n]$$

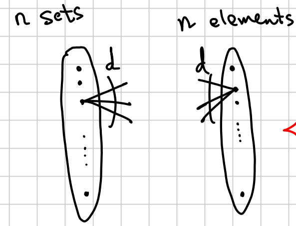
Now relax integrality constraint $x_S \in \{0,1\}$ and use fractional constraint $0 \leq x_S \leq 1$:

$$\min \sum_S x_S$$
$$\text{s.t.} \sum_{\{S:\, e\in S\}} x_S \geq 1 \quad \forall \text{element } e$$
$$0 \leq x_S \leq 1 \quad \forall \text{set } S$$

fractional relaxation for Set-Cover

Example of fractional Set Cover

$$\min \sum_S x_S$$
$$\text{s.t.} \sum_{\{S:\, e\in S\}} x_S \geq 1 \quad \forall \text{element } e$$
$$0 \leq x_S \leq 1 \quad \forall \text{set } S$$



d-regular bipartite graph (superposition of d matchings)

Suppose $\forall$ set S : $|S| = d$

$\forall$ element e : $|\{S : e \in S\}| = d$

$\Longrightarrow$ $x_S = \frac{1}{d}$ is a feasible solution.

Question #2: prove that $\sum x_S = \frac{n}{d}$ is optimal

What can we do with (optimal) solution $\{x^*_S\}_S$ of LP ?

$$\min \sum_S x_S$$
$$\text{s.t.} \sum_{\{S:\, e\in S\}} x_S \geq 1 \quad \forall \text{element } e$$
$$0 \leq x_S \leq 1 \quad \forall \text{set } S$$

1) Lower bound : $\sum_S x^*_S \triangleq OPT_f \leq OPT$

2) If (almost) integral then $\{x^*_S\}$ is (almost) a set cover.

3) Round $\{x^*_S\}$ ...

First, let's use it to analyze Greedy.

Greedy-with charging

$C \leftarrow \emptyset$

while $C \neq [1..n]$ do

- $S \leftarrow \arg\max\{ |S_i \setminus C| : 1 \leq i \leq m \}$
- $\forall e \in S \setminus C$: $p(e) \leftarrow \frac{1}{|S \setminus C|}$ (charge covered elements)
- $C \leftarrow C \cup S$ (add S to the cover)

In each round $\sum_{e \in S \setminus C} p(e) = 1$ pays for S.

$\Longrightarrow$ In the end : $\sum_e p(e) = |\text{cover}|$

Why does this help ?

Primal LP (covering)

min $\sum_S x_S$

s.t. $\sum_{\{S : e \in S\}} x_S \geq 1$ $\forall$ element e

$0 \leq x_S \leq 1$ $\forall$ set S

Dual LP (packing)

max $\sum_e y_e$

s.t. $\sum_{e \in S} y_e \leq 1$ $\forall S$

$0 \leq y_e$

Claim (weak duality) : $\sum_S x_S \geq \sum_e y_e$

proof : $\sum_S 1 \cdot x_S \geq \sum_S \left(\sum_{e \in S} y_e\right) x_S = \sum_e \left(\sum_{S: e \in S} x_S\right) y_e \geq \sum_e y_e$.

Hope : If $\{p_e\}$ is a dual solution, then

$$|\text{Greedy}| = \sum_e p_e \leq OPT_f \leq OPT \ldots$$

Question #3: show that $\{p_e\}_e$ is NOT a dual solution.

Lovasz : $y_e = \frac{p_e}{1+\ln n}$ is a dual solution !

and then $|\text{Greedy}| = \sum_e p_e = (1+\ln n) \sum_e y_e$

$$\leq (1+\ln n) \cdot OPT_f \leq (1+\ln n) \cdot OPT$$

$\Rightarrow$ Greedy is an $O(\lg n)$ – approximation.

So, now left to prove feasibility of $\{y_e\}$.

Consider any set S'

$S' = \{e_1, e_2, \ldots, e_\ell\}$.

let $T(e) \triangleq$ iteration in which e is covered

while $C \neq [1..n]$ do

- $S \leftarrow \operatorname{argmax}\{ |S_i \setminus C| : 1 \leq i \leq m\}$
- $\forall e \in S \setminus C$: $p(e) \leftarrow \frac{1}{|S \setminus C|}$ (charge covered elements)
- $C \leftarrow C \cup S$ (add S to the cover)

assume: $T(e_1) \leq T(e_2) \leq \cdots \leq T(e_\ell)$

$$\Rightarrow p(e_1) \leq p(e_2) \leq \cdots \leq p(e_\ell)$$

Consider iteration in which e_i is covered.

$\{e_i, \ldots, e_\ell\} \subseteq S' \setminus C$ & $|S \setminus C| \geq |S' \setminus C|$

$$\Rightarrow p(e_i) = \frac{1}{|S \setminus C|} \leq \frac{1}{|S' \setminus C|} = \frac{1}{\ell - i + 1}.$$

$$\Rightarrow p(e_1) \leq \frac{1}{\ell},\ p(e_2) \leq \frac{1}{\ell - 1},\ \ldots,\ p(e_\ell) \leq 1.$$

$\Rightarrow \forall S : \sum_{e \in S} p(e) \leq \sum_{i=1}^{n} \frac{1}{i} \triangleq H_n \leq 1 + \ln n$.

$\Rightarrow y_e = \frac{p(e)}{1+\ln n}$ is dual feasible. $\square$

Question #4: prove that $\ln(n+1) \leq H_n \leq 1 + \ln n$

Recap

1) Set Cover Problem

2) Greedy Algorithm

3) Integer Programming Formulation.

4) Linear Programming Formulation.

5) Primal & Dual LP's.

6) Weak Duality

7) Analysis of approximation ratio of Greedy.

Hint for Q1: Find an example in which $OPT \geq \Omega(\lg n) \cdot OPT_f$.

Hint: Elements $\mathcal{U} = \{0,1\}^k - 0^k$ (all nonzero strings)

Sets $\{S_i\}_{i \in \mathcal{U}}$

where $S_i = \{ u \in \mathcal{U} \mid \sum_{j=1}^{k} u_j \cdot i_j \text{ is odd} \}$

Show: $\forall i: \ |S_i| = 2^k/2$

$\forall u: \ |\{S_i : u \in S_i\}| = 2^k/2$

use prob. (2) to prove $OPT_f \leq 2$.

prove $OPT \geq k$ (system of $<k$ equations in $GF(2)$ has nonzero solution)

MORE QUESTIONS

5) Suppose that $\forall$ element: $|\{S : e \in S\}| \leq d$.

Find an approximation algorithm with

$$|ALG| \leq d \cdot OPT.$$

6) Improve the analysis of the approximation ratio of the Greedy algorithm if

$\forall$ Set: $|Set| \leq k$. (say, $k=3$).

* what if $k = 2$???

7) Weak duality

Primal	Dual
min $c^t x$	max $y^t b$
s.t. $Ax \geq b$	s.t $y^t A \leq c^t$
$x \geq 0$	$y \geq 0$

prove that: if x is a primal solution & y is a dual solution, then $c^t x \geq y^t b$.

Encore : Rounding a fractional set-cover

Suppose $\{x_S\}$ is an optimal fractional set-cover,

I.E., $\forall e: \sum_{e \in S} x_S \geq 1$ and $\sum_S x_S = OPT_f$.

How can we obtain an integral set-cover $\hat{x}_S \in \{0,1\}$ from x_S ?

2 Goals :

feasibility : $\forall e \ \exists S : \ e \in S \ \& \ \hat{x}_S = 1$

cost : $\frac{\sum_S \hat{x}_S}{\sum_S x_S}$ is small $\left(\sum_S \hat{x}_S \leq O(\lg n) \cdot OPT_f\right)$

Randomized Rounding

For each set S, randomly flip a "coin" $\hat{x}_S$ such that $\Pr[\hat{x}_S = 1] = x_S$.

Now : $\forall$ element e :

$$\Pr[e \text{ is covered}] = \Pr[\exists S: e \in S \ \& \ \hat{x}_S = 1]$$

$$= 1 - \Pr[\forall S: e \in S \Rightarrow \hat{x}_S = 0]$$

$e^{-x} \geq 1 - x$

$$= 1 - \prod_{S: e \in S} (1 - x_S)$$

$$\geq 1 - e^{-\sum_{e \in S} x_S} \geq 1 - e^{-1} \approx 0.632$$

$\Rightarrow$ Each element is covered with constant prob.

Repeat this experiment $\ln(4n)$ times

(S is in the cover if $\hat{x}_S = 1$ in one of experiments)

$$\Pr(e \text{ is not covered}) \leq \left(\frac{1}{e}\right)^{\ln(4n)} = \frac{1}{4n}$$

$$\Rightarrow \Pr(\exists e \text{ not covered}) \leq \sum_e \Pr(e \text{ not covered}) = \frac{1}{4}$$

$\Rightarrow$ with prob $\geq 3/4$ every element is covered if we repeat $\ln(4n)$ times.

Amplify prob. of success by repeating $k \cdot \ln(4n)$ times $\Longrightarrow \Pr(\text{Success}) \geq 1 - \left(\frac{1}{4}\right)^k$.

What about cost ?

$$E\left[\sum_S \hat{x}_S\right] = \sum_S E[\hat{x}_S] = \sum_S x_S$$

We repeat $\ln(4n)$ times $\Longrightarrow E(\text{cover}) = \ln(4n) \cdot OPT_f$.

Markov Inequality:

$$\Pr(\text{cover} \geq 4 \cdot \ln(4n)\, OPT_f) \leq \frac{1}{4}.$$