CS7800: Advanced Algorithms

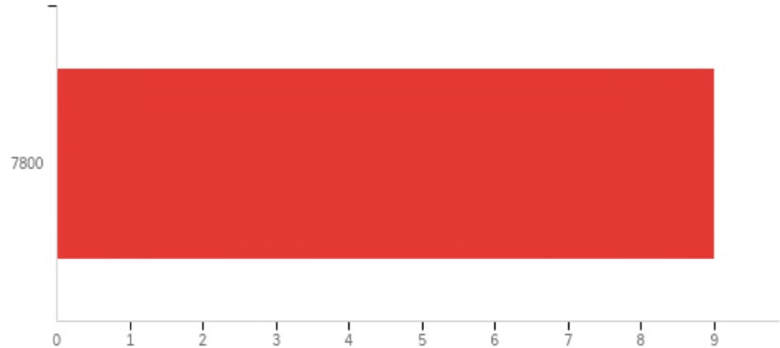
Lecture 15: Approximation Algorithms II

- Covering Problems

Jonathan Ullman

10-28-2022

CourseNumber



#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	CourseNumber	7800.00	7800.00	7800.00	0.00	0.00	9

#	Answer	%	Count
1	7800	100.00%	9
	Total	100%	9

Maximum Coverage (variant of Set Cover)

Inputs: Sets $S_1, \ldots, S_m \subseteq \{1, \ldots, n\} = U$

A budget k

Objective: Output sets $\{A_1 \ldots A_k\} \subseteq \{S_1, \ldots, S_m\}$

maximizing $\left|\bigcup_{i=1}^{k} A_i\right|$

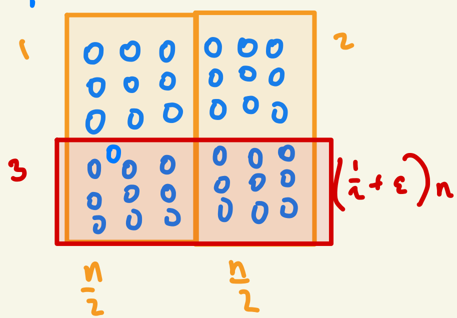
Recap: Problem is NP-hard to solve exactly

Greedy Maximum Coverage

For $i = 1, \dots, k$:

Pick the set $A_i \in \{S_1, \dots, S_m\}$ that covers the most <u>new</u> elements

bad example $(k=2)$

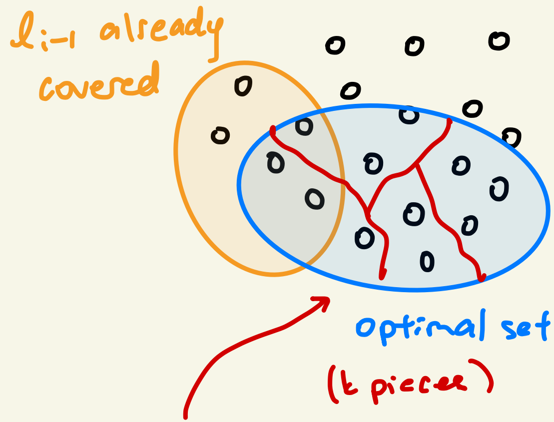


$OPT = n$

$GREEDY \approx \frac{3n}{4}$

Analyzing Greedy Set Cover

<u>Fact #1</u>: There exists a set that covers $\geq \frac{OPT}{k}$ elts



iteration i of greedy

ℓ_i = # of elements covered by greedy after i iterations

exists a set that covers $\frac{OPT - \ell_{i-1}}{k}$ <u>new</u> elts

for the remaining elements, we can cover $OPT - \ell_{i-1}$ with k sets

<u>Fact #2</u>: At iteration i, exists a set that covers $\frac{OPT - \ell_{i-1}}{k}$

Analyzing Greedy Cont'd

<u>Fact #2</u>: At iteration i, greedy covers $\frac{OPT - \ell_{i-1}}{k}$ new elts

$$\ell_k = (\ell_k - \ell_{k-1}) + \ell_{k-1}$$

$$\geq \frac{OPT - \ell_{k-1}}{k} + \ell_{k-1} = \frac{OPT}{k} + \left(1 - \frac{1}{k}\right)\ell_{k-1}$$

$$= \frac{OPT}{k} + \left(1 - \frac{1}{k}\right)\left(\frac{OPT}{k} + \left(1 - \frac{1}{k}\right)\ell_{k-2}\right)$$

$$= \frac{OPT}{k}\left(1 + \left(1 - \frac{1}{k}\right) + \left(1 - \frac{1}{k}\right)^2 + \dots + \left(1 - \frac{1}{k}\right)^{k-1}\right)$$

$$= \frac{OPT}{k} \cdot \frac{1 - \left(1 - \frac{1}{k}\right)^k}{1 - \left(1 - \frac{1}{k}\right)} = OPT \cdot \left(1 - \left(1 - \frac{1}{k}\right)^k\right) \longrightarrow OPT \cdot \left(1 - \frac{1}{e}\right)$$

Vertex Cover (special case of Set Cover)

<u>Inputs:</u> An undirected graph $G = (V, E)$

<u>Objective:</u> Output nodes $C \subseteq V$ such that for every edge $(u,v) \in E$ either $u \in C$ or $v \in C$ (or both).

Recap: Problem is NP-hard to solve exactly

LP Relaxations and Rounding

IP for Vertex Cover

$$\min_x \sum_u x_u$$

$$x_u + x_v \geq 1 \quad \text{for all } (u,v) \in E$$

$$x_u \geq 0$$

$$x_u \text{ integer}$$

- Exactly solves Vertex Cover
- Optimal value is OPT_{IP}

LP Relaxation

$$\min_x \sum_u x_u$$

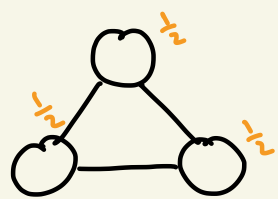
$$x_u + x_v \geq 1 \quad \text{for } (u,v) \in E$$

$$x_u \geq 0$$

$$x_u \in \mathbb{R}$$

- Optimal value $OPT_{LP} \leq OPT_{IP}$

$OPT_{IP} = 2$



$OPT_{LP} = \frac{3}{2}$

LP Relaxations and Rounding

IP for Vertex Cover

$$\min_x \sum_u x_u$$

$$x_u + x_v \geq 1 \quad \text{for all } (u,v) \in E$$

$$x_u \geq 0$$

$$x_u \text{ integer}$$

- Exactly solves Vertex Cover
- Optimal value is OPT_{IP}

LP Relaxation

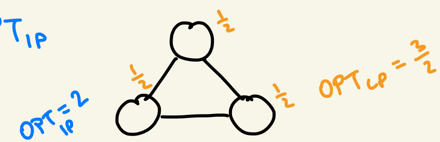
$$\min_x \sum_u x_u$$

$$x_u + x_v \geq 1 \quad \text{for } (u,v) \in E$$

$$x_u \geq 0$$

$$x_u \in \mathbb{R}$$

- Optimal value $OPT_{LP} \leq OPT_{IP}$



A rounding algorithm takes a fractional solution x_{LP} and outputs an integer solution x_{IP} s.t. $\text{value}(x_{IP}) \leq c \cdot \text{value}(x_{LP})$

$$ROUND_{LP} \leq c \cdot OPT_{LP} \leq c \cdot OPT_{IP}$$

LP Relaxations and Rounding

IP for Vertex Cover

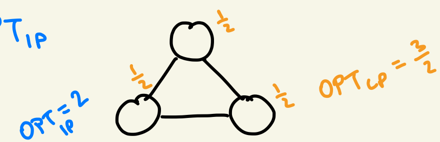
$$\min_x \sum_u x_u$$
$$x_u + x_v \geq 1 \quad \text{for all } (u,v) \in E$$
$$x_u \geq 0$$
$$x_u \text{ integer}$$

- Exactly solves Vertex Cover
- Optimal value is OPT_{IP}

LP Relaxation

$$\min_x \sum_u x_u$$
$$x_u + x_v \geq 1 \quad \text{for } (u,v) \in E$$
$$x_u \geq 0$$
$$x_u \in \mathbb{R}$$

- Optimal value $OPT_{LP} \leq OPT_{IP}$



Given x^{LP} for the linear program, define $x_u^{IP} = \begin{cases} 1 & \text{if } x_u^{LP} \geq \frac{1}{2} \\ 0 & \text{if } x_u^{LP} < \frac{1}{2} \end{cases}$

- $\sum_u x_u^{IP} \leq 2 \cdot \sum_u x_u^{LP}$
- if $x_u^{LP} + x_v^{LP} \geq 1$ then at least one of $x_u^{IP} = 1$ and $x_v^{IP} = 1$ is true

Weighted Set Cover

Inputs: Sets $S_1, \ldots, S_m \subseteq \mathcal{U}$ $(|\mathcal{U}| = n)$

Costs $c_1, \ldots, c_m$

Objective: Find a minimum cost collection of sets whose union is $\mathcal{U}$

Recap: Problem is NP-hard to solve exactly

LP Relaxation for Set Cover

Primal

$$\min \sum_{i=1}^{m} c_i x_i$$

$$\text{s.t.} \sum_{i: e \in S_i} x_i \geq 1 \quad \text{for } e \in \mathcal{U} \quad (p_e)$$

$$x_i \geq 0$$

Dual

$$\max \sum_{e} p_e$$

$$\text{s.t.} \sum_{e \in S_i} p_e \leq c_i \quad \text{for } i = 1 \dots m$$

$$p_e \geq 0$$

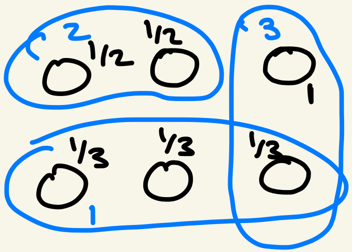
For any feasible dual $\{p_e\}$ $\quad \sum_e p_e \leq OPT_{LP} \leq OPT_{IP} \leq GREEDY$

Greedy Alg for Weighted Set Cover

Until all of $\mathcal{U}$ is covered
- choose S_i minimizes $\boxed{\dfrac{c_i}{\#\text{of new elts covered}}}$

ratio

costs = 1



- q_e = ratio of the first set that covers e
- $\sum_e q_e$ = GREEDY

Fix some iteration. Suppose ℓ elts of S_i are covered.

Then the ratio of S_i is $\dfrac{c_i}{|S_i| - \ell}$

Greedy Alg for Weighted Set Cover

Until all of $\mathcal{U}$ is covered
- choose S_i minimizes $\boxed{\dfrac{c_i}{\#\text{ of new elts covered}}}$

ratio

- q_e = ratio of the first set that covers e

Fix some iteration. Suppose ℓ elts of S_i are covered.
Then the ratio of S_i is $\dfrac{c_i}{|S_i| - \ell}$

<u>Clm:</u> For every S_i, the j^{th} elt covered satisfies $q_e \le \dfrac{c_i}{|S_i| - (j-1)}$

Greedy Alg for Weighted Set Cover

- q_e = ratio of the first set that covers e

Fix some iteration. Suppose ℓ elts of S_i are covered.
Then the ratio of S_i is $\frac{c_i}{|S_i| - \ell}$

<u>Clm:</u> For every S_i, the j^{th} elt covered satisfies $q_e \leq \frac{c_i}{|S_i| - (j-1)}$

$$\sum_{e \in S_i} q_e \leq \frac{c_i}{|S_i|} + \frac{c_i}{|S_i|-1} + \frac{c_i}{|S_i|-2} + \ldots + \frac{c_i}{1}$$
$$= c_i \left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{|S_i|} \right) \approx c_i \cdot \ln|S_i| \leq c_i \cdot \ln(n)$$

$$\sum_{e \in S_i} q_e \leq c_i \cdot \ln(n) \iff \sum_{e \in S_i} \frac{q_e}{\ln(n)} \leq c_i$$

$= p_e$

LP Relaxation for Set Cover

Primal

$$\min \sum_{i=1}^{m} c_i x_i$$

$$\text{s.t.} \quad \sum_{i: e \in S_i} x_i \geq 1 \quad \text{for } e \in \mathcal{U} \quad (p_e)$$

$$x_i \geq 0$$

Dual

$$\max \sum_e p_e$$

$$\text{s.t.} \quad \sum_{e \in S_i} p_e \leq c_i \quad \text{for } i = 1 \ldots m$$

$$p_e \geq 0$$

For any feasible dual $\{p_e\}$

$$\sum_e p_e \leq OPT_{LP} \leq OPT_{IP} \leq GREEDY$$

$$\sum_e p_e = \sum_e \frac{q_e}{\ln(n)} = \frac{GREEDY}{\ln(n)}$$