Set Cover

Input: Set system $(U, \{S_1, \dots S_m\})$ (each $S_i \subseteq U$)
with weight $w: [m] \to \mathbb{R}^+$.

Goal: Find $C \subseteq [m]$ s.t. $\left(\bigcup_{i \in C} S_i\right) = U$ and $w(C)$ is minimized.

NP-Complete to solve exactly! So need approximation algorithms!

Parameters: $n = |U|$, $m = (\#\text{ sets})$

$k = \max_{i \in [m]} |S_i|$ $\qquad$ $d = \max_{j \in U} |\{i \in [m] : j \in S_i\}|$

max set size $\qquad$ max "degree"

Will see 4 algorithms, 2 guarantees.

LP Relaxation (vars: $(x_i)_{i \in [m]}$)

$$\min \sum_{i=1}^{m} w_i x_i$$
$$\text{s.t.} \sum_{S_i \ni j} x_i \geq 1 \quad \forall j \in U$$
$$x \geq 0$$

Dual (vars: $(y_j)_{j \in U}$)

$$\max \sum_{j \in U} y_j$$
$$\text{s.t.} \sum_{j \in S_i} y_j \leq w_i \quad \forall i \in [m]$$
$$y \geq 0$$

Rounding

① Solve the primal LP to get (fractional) optimal $x \in \mathbb{R}^m$

② "Round" x to an integral solution C.

Deterministic Rounding, $C = \{ i \in [m] : x_i \geq 1/d \}$.

$*$ C is a set cover since $\forall j \in U$, $\exists i \in [m]$ s.t. $x_i \geq 1/d$ and $j \in S_i$

(otherwise, $\sum_{S_i \ni j} x_i < d \cdot (1/d) = 1$, so

x is not feasible in the LP.)

$*$ $w(C) = \sum_{i \in C} w_i = d \cdot \sum_{i \in C} w_i / d \leq d \cdot \sum_{i \in C} w_i \cdot x_i \leq d \sum_{i \in [m]} w_i \cdot x_i = d \cdot OPT_{LP}$

by def. of C.

So, d-approximation!

Check when $d=2$, the problem is (Weighted) Vertex Cover.

Randomized Rounding, For each $i \in [m]$, put i to C w.p. $\min(\alpha \cdot \log n \cdot x_i, 1)$.
independently.

$*$ $\forall j \in U$, if $\exists S_i \ni j$ with $x_i \geq 1/\alpha \log n$, i is always in C.

otherwise, $\Pr[j \notin (\bigcup_{i \in C} S_i)] = \prod_{S_i \ni j} (1 - \alpha \cdot \log n \cdot x_i)$

$(1-y) \leq e^{-y}$ $\longrightarrow$ $\leq \prod_{S_i \ni j} e^{-\alpha \log n \cdot x_i}$

$e^{-\alpha \log n \cdot x_i} = n^{-\alpha x_i}$ $\longrightarrow$ $= n^{-(\alpha \sum_{S_i \ni j} x_i)}$

$\leq n^{-\alpha}$.

if $\alpha = 2$, $\Pr[\exists j \text{ s.t. } j \notin (\bigcup_{i \in C} S_i)] \leq n \cdot n^{-\alpha} \leq 1/n$ by union bound.

* $\mathbb{E}[w(C)] \le \alpha \cdot \log n \cdot \sum_{i \in [m]} w_i x_i$, so

w.p. ≥ 0.9, $w(C) \le 10\alpha \log n \cdot OPT_{LP}$.

$\therefore$ w.p. $\ge 0.9 - 1/n$, C will be a set cover with $w(C) \le 20 \log n \cdot OPT_{LP}$!

So, $\min(d, O(\log n))$-approximation. Both are optimal.

Thm. $\forall \varepsilon > 0$, $\nexists$ poly-time $(1-\varepsilon)\ln n$-approximation algorithm unless $P \ne NP$.

Thm. $\forall \varepsilon > 0$, $\nexists$ " $(d-\varepsilon)$ - " " assuming "Unique Games Conjecture"

Dual Fitting

"Combinatorial (often greedy) algo", analyzed by constructing a dual feasible solution.

Greedy

$C \leftarrow \phi$.

While $U \neq \phi$.

choose set with best "bang-for-buck" ratio.

$\quad i^* \leftarrow \underset{i \in [m]}{\operatorname{argmin}} \frac{w(i)}{|S_i \cap U|}$.

$\quad C \leftarrow C \cup \{i^*\}$, $U \leftarrow U \setminus S_{i^*}$

* C is a set cover at the end by design.
* In each iteration, $\forall j \in S_{i^*} \cap U$, (element covered for the first time)

  let $y_j = w(i^*) / |S_{i^*} \cap U|$.

  At the end, $w(C) = \sum_{j \in U} y_j$.

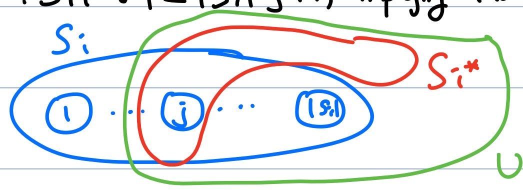
  Want to show (y / H_k) is feasible for dual, which implies that $w(C) \leq H_k \cdot LP_{OPT}$! ($1 + 1/2 + \dots + 1/k$.)

Dual (vars: $(y_j)_{j \in U}$)

$$\max \sum_{j \in U} y_j$$
$$\text{s.t.} \sum_{j \in S_i} y_j \leq w_i \quad \forall i \in [m]$$
$$y \geq 0$$

Claim. $\forall i \in [m]$, $\sum_{j \in S_i} y_j \leq w_i \cdot H_{|S_i|}$.

Pf. WLOG, let $S_i = \{1, 2, 3, \dots, |S_i|\}$, sorted in increasing order of times they are covered. Then $\forall j \in [S_i]$, $y_j \leq w_i / (|S_i| - j + 1)$, because in the iteration that covered j first time, $|S_i \cap U| \geq |S_i| - j + 1$, implying that $w(i^*) / |S_{i^*} \cap U| \leq w(i) / (|S_i| - j + 1)$ in that iteration. □



when j is covered first time, S_i^* is "better" than S_i

Primal-Dual

(When primal problem is "covering")

$\min \langle c, x\rangle$ s.t. $Ax \geq b$, $x \geq 0$
$(A, b, c \geq 0)$

Increase dual variables and "pick" primal variable corresponding to tight dual constraints

Primal-Dual.

$y \leftarrow 0$, $C \leftarrow \phi$.

While $\exists j \in U \setminus (\bigcup_{i \in C} S_i)$

  Choose an arbitrary such j.

  Increase y_j until $\exists i \in [m]$ s.t. $\sum_{j \in S_i} y_j = w_i$

  $C = \{i \in [m] : \sum_{j \in S_i} y_j = w_i\}$.

Dual (vars: $(y_j)_{j \in U}$)

$\max \sum_{j \in U} y_j$

s.t. $\sum_{j \in S_i} y_j \leq w_i \quad \forall i \in [m]$

$y \geq 0$

Correctness.

* y is feasible throughout the algo.
* When j is chosen, all $S_i \ni j$ has $\sum_{j \in S_i} y_j < w_i$.
  $\Rightarrow |C|$ will increase by at least 1 in each iteration.
* Algo will end and C will be a set cover.

Analysis. If x is a feasible primal soln, weak-duality goes like:

$$\sum_{i \in [m]} w_i \cdot x_i \geq \sum_{i \in [m]} \Big(\sum_{j \in S_i} y_j\Big) x_i = \sum_{j \in U} \Big(\Big(\sum_{i : S_i \ni j} x_i\Big) y_j\Big) \geq \sum_{j \in U} y_j.$$

Now, let x be the indicator vector of C.

"=" holds since $\forall i \in [m]$, either $x_i = 0$ or $w_i = \sum_{j \in S_i} y_j$.

"$\leq d$" holds since $\sum_{S_i \ni j} x_i \leq |S_i| \leq d$.

So, $d \cdot \sum_j y_j \geq w(C) \geq OPT_{SC} \geq LP_{SC} \geq \sum_j y_j$. d-approximation