k-Median Problem

Given: Positive integer k; set of facilities F; set of clients D; metric connection costs $c_{ij} \geq 0$, $i \in F$, $j \in D$.

Task: Find $S \subseteq F$ with $|S| \leq k$ minimizing $\sum_{j \in D} \min_{i \in S} c_{ij}$.

IP formulation:

$$
\begin{aligned}
\mathrm{OPT}_k \;:=\; \min \quad & \sum_{i \in F, j \in D} c_{ij} \cdot x_{ij} \\
\text{s.t.} \quad & \sum_{i \in F} x_{ij} = 1 && \text{for all } j \in D, \\
& y_i - x_{ij} \geq 0 && \text{for all } i \in F, j \in D, \\
& \sum_{i \in F} y_i \leq k \\
& x_{ij}, y_i \in \{0,1\} && \text{for all } i \in F, j \in D.
\end{aligned}
$$

Lagrangean Relaxation

Take LP relaxation and move constraint $\sum y_i \leq k$ to objective function.

Let $\lambda \geq 0$ (Langrange multiplier):

$$
\begin{aligned}
\min \quad & \sum_{i \in F, j \in D} c_{ij} \cdot x_{ij} + \sum_{i \in F} \lambda \cdot y_i - \lambda \cdot k \\
\text{s.t.} \quad & \sum_{i \in F} x_{ij} = 1 && \text{for all } j \in D, \\
& y_i - x_{ij} \geq 0 && \text{for all } i \in F, j \in D, \\
& x_{ij}, y_i \geq 0 && \text{for all } i \in F, j \in D.
\end{aligned}
$$

Dual LP:

$$
\begin{aligned}
\max \quad & \sum_{j \in D} v_j - \lambda \cdot k \\
\text{s.t.} \quad & \sum_{j \in D} w_{ij} \leq \lambda && \text{for all } i \in F, \\
& v_j - w_{ij} \leq c_{ij} && \text{for all } i \in F, j \in D, \\
& w_{ij} \geq 0 && \text{for all } i \in F, j \in D.
\end{aligned}
$$

Algorithmic Idea

Use primal-dual algorithm for uncap. facility location with $f_i := \lambda$, $i \in F$.

This yields $S \subseteq F$ and feasible dual (v, w) with

$$\tfrac{1}{3}\sum_{j\in D}\min_{i\in S} c_{ij} + \sum_{i\in S} f_i \;\le\; \sum_{j\in D} v_j \;.$$

That is,

$$c(S) := \sum_{j\in D}\min_{i\in S} c_{ij} \;\le\; 3\left(\sum_{j\in D} v_j - \lambda\cdot|S|\right) \;.$$

Observation 7.12.

If the algorithm happens to output a set $S \subseteq F$ with $|S| = k$, then

$$c(S) \;\le\; 3\left(\sum_{j\in D} v_j - \lambda\cdot k\right) \;\le\; 3\cdot \mathrm{OPT}_k \;.$$

□

Question: Can we find $\lambda \ge 0$ such that $|S| = k$?

Choosing a Suitable $\lambda \ge 0$

Observation.

- For $\lambda = 0$ the algorithm outputs $S \subseteq F$ with $|S| \ge k$ (w.l.o.g.).
- For $\lambda = \sum_{j\in D}\sum_{i\in F} c_{ij}$, the algorithm opens a single facility only.

Idea: Use binary search until one of the following two things happen:

- find λ such that the algorithm opens exactly k facilities;
- find $\lambda_1 < \lambda_2$ with $\lambda_2 - \lambda_1 \le \frac{\varepsilon}{3} c_{\min}/|F|$; ($c_{\min} := \min\{c_{ij} \mid c_{ij} > 0\}$) for $\lambda = \lambda_\ell$ the algorithm outputs $S_\ell \subseteq F$ with $|S_1| > k > |S_2|$ and

$$c(S_\ell) \le 3\left(\sum_{j\in D} v_j^\ell - \lambda_\ell\cdot|S_\ell|\right) \;, \qquad \ell = 1, 2.$$

Notice that binary search terminates after $O(\log \frac{|F|\sum c_{ij}}{\varepsilon\cdot c_{\min}})$ iterations.

- In the first case, we have found a solution of cost at most $3\cdot\mathrm{OPT}_k$.
- In the second case we still need to work a little...

Finding a Suitable $S \subseteq F$ with $|S| \leq k$

Goal: Find $S \subseteq F$ with $|S| \leq k$ and $c(S) \leq 2(3+\varepsilon) \cdot \mathrm{OPT}_k$.

- Let $\alpha_1 := \dfrac{k - |S_2|}{|S_1| - |S_2|}$ and $\alpha_2 := \dfrac{|S_1| - k}{|S_1| - |S_2|}$.

  Notice that $\alpha_1 + \alpha_2 = 1$, $\alpha_1, \alpha_2 \geq 0$, and $\alpha_1 |S_1| + \alpha_2 |S_2| = k$.
- Let $\tilde{v} := \alpha_1 v^1 + \alpha_2 v^2$ and $\tilde{w} := \alpha_1 w^1 + \alpha_2 w^2$.

  Notice that $(\tilde{v}, \tilde{w})$ is a feasible dual solution for facility costs λ_2.

Lemma 7.13.

It holds that $\alpha_1 \cdot c(S_1) + \alpha_2 \cdot c(S_2) \leq (3+\varepsilon) \cdot \mathrm{OPT}_k$.

Proof:... □

Thus, if $\alpha_2 \geq \frac{1}{2}$, we reach our goal by setting $S := S_2$.

In the following we can thus assume that $\alpha_2 < \frac{1}{2}$...

Finding a Suitable $S \subseteq F$ with $|S| \leq k$ (cont.)

Algorithm.

1. $S := \emptyset$; for each facility $i \in S_2$, add the closest facility $h \in S_1$ to S;
2. while $|S| < k$, choose random facility in $S_1 \setminus S$ and add it to S;

Lemma 7.14.

If $\alpha_2 < \frac{1}{2}$, opening facilities as above has cost $\mathrm{E}[c(S)] \leq 2(3+\varepsilon) \cdot \mathrm{OPT}_k$.

Proof:... □

The algorithm can be derandomized by method of conditional probabilities.

Theorem 7.15.

There is a $(6+\varepsilon)$-approximation algorithm for the k-median problem. □

Theorem 7.16.

There is no 1.735-approximation algorithm unless each problem in NP has an $n^{O(\log \log n)}$ time algorithm. □