Clustering What Matters in Constrained Settings
(Improved Outlier to Outlier-Free Reductions)

Ragesh Jaiswal

CSE, IIT Delhi

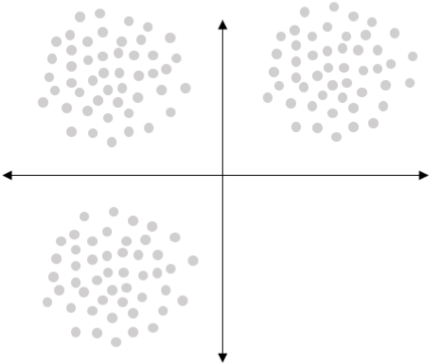
ISAAC Talk, December 06, 2023

[Joint work with Amit Kumar (IIT Delhi)]

The k-Median Problem

The k-median problem

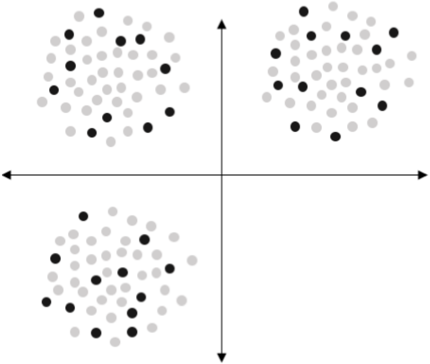
Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, and an integer k, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of every point in X to the nearest center in C is minimized.



The k-Median Problem

The k-median problem

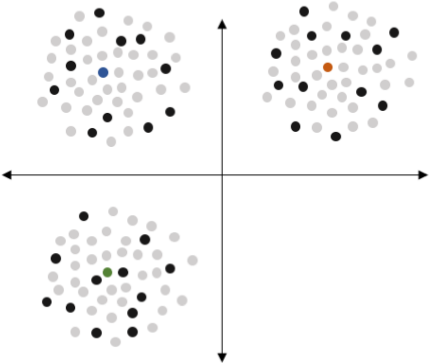
Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, and an integer k, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of every point in X to the nearest center in C is minimized.



The k-Median Problem

The k-median problem

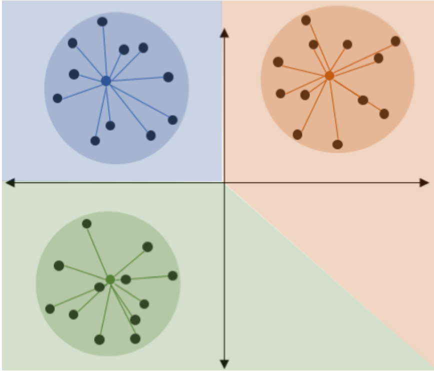
Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, and an integer k, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of every point in X to the nearest center in C is minimized.



The k-Median Problem

The k-median problem

Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, and an integer k, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of every point in X to the nearest center in C is minimized.



The k-Median Problem

The k-median problem

Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, and an integer k, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of every point in X to the nearest center in C is minimized.

- Known results:

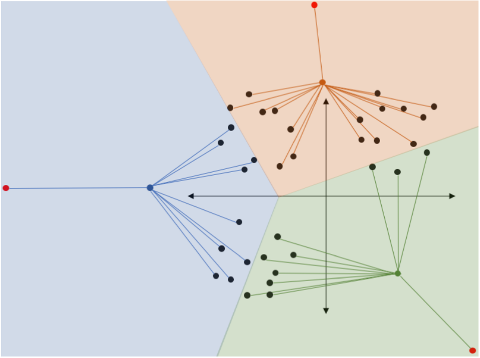
	Poly-time	FPT-time
Lower bound	$(1 + \frac{2}{e} - \varepsilon) \approx 1.735 - \varepsilon$ Guha and Khuller (1999)	$(1 + \frac{2}{e} - \varepsilon)$ Guha and Khuller (1999)
Upper bound	$2.675 + \varepsilon$ Byrka et al. (2017)	$(1 + \frac{2}{e} + \varepsilon)$ Cohen-Addad et al. (2019)

- FPT-time algorithms have running time of the form $f(k) \cdot n^{O(1)}$, where k is a parameter of interest ($f(k)$ *can be an exponential function*). FPT-time algorithms are poly-time for constant k.

The k-median problem

Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, and an integer k, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of every point in X to the nearest center in C is minimized.

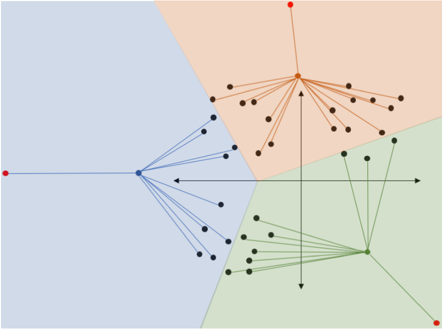
- Presence of outlier points may adversely impact the clustering.



The k-median problem

Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, and an integer k, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of every point in X to the nearest center in C is minimized.

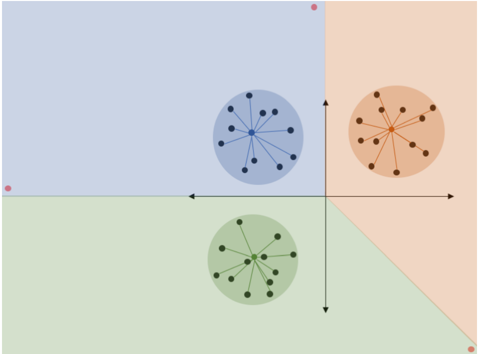
- Presence of outlier points may adversely impact the k-median clustering.
- So, we must allow ignoring a few points to cluster what matters.



The outlier k-median problem

Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, an integer k, and an integer m, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of ~~every point~~ all but m points in X to the nearest center in C is minimized.

- So, we must allow ignoring a few points to cluster what matters.



The outlier k-median problem

Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, an integer k, and an integer m, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of ~~every point~~ all but m points in X to the nearest center in C is minimized.

- Known results:

	Poly-time	FPT-time
Lower bound	$(1 + \frac{2}{e} - \varepsilon) \approx 1.735 - \varepsilon$ Guha and Khuller (1999)	$(1 + \frac{2}{e} - \varepsilon)$ Cohen-Addad et al. (2019)
Upper bound	$7 + \varepsilon$ Krishnaswamy et al. (2018)	$(1 + \frac{2}{e} + \varepsilon)$ Cohen-Addad et al. (2019)

The outlier k-median problem

Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, an integer k, and an integer m, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of ~~every point~~ all but m points in X to the nearest center in C is minimized.

- Known results for Outlier k-median:

	Poly-time	FPT-time
Lower bound	$(1 + \frac{2}{e} - \varepsilon) \approx 1.735 - \varepsilon$ Guha and Khuller (1999)	$(1 + \frac{2}{e} - \varepsilon)$ Cohen-Addad et al. (2019)
Upper bound	$7 + \varepsilon$ Krishnaswamy et al. (2018)	$(1 + \frac{2}{e} + \varepsilon)$ Cohen-Addad et al. (2019)

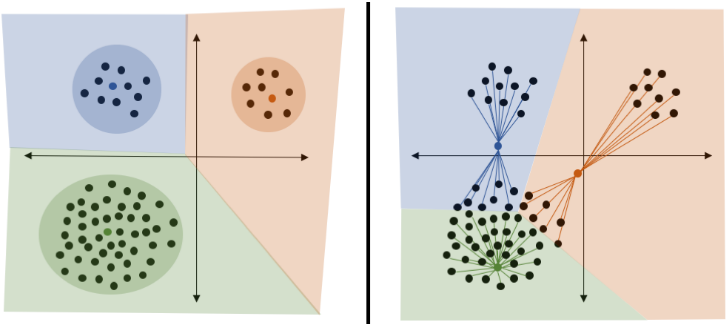
- Known results for Outlier-Free k-median:

	Poly-time	FPT-time
Lower bound	$(1 + \frac{2}{e} - \varepsilon) \approx 1.735 - \varepsilon$ Guha and Khuller (1999)	$(1 + \frac{2}{e} - \varepsilon)$ Guha and Khuller (1999)
Upper bound	$2.675 + \varepsilon$ Byrka et al. (2017)	$(1 + \frac{2}{e} + \varepsilon)$ Cohen-Addad et al. (2017)

The constrained k-median problem

Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, and an integer k, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of every point in X to the nearest feasible center in C is minimized, where constraints on centers/clusters determine feasibility.

- Example:



k-median | Constrained *k*-median

The constrained k-median problem

Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, and an integer k, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of every point in X to the nearest feasible center in C is minimized, where constraints on centers/clusters determine feasibility.

- The constraint may be on:
  - Centers: Restrictions on the number of points a center can service (e.g., capacitated clustering),
  - Clusters: Restrictions on the size of clusters (e.g., balanced clustering),
  - Label-based: Every point has an associated color (*indicating socio-economic groups*), and there are **fairness** restrictions such as proportional representation from each group in every cluster (e.g., fault-tolerant clustering),
  - or, a combination of the above (e.g., fair clustering).

The constrained k-median problem

Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, and an integer k, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of every point in X to the nearest feasible center in C is minimized, where constraints on centers/clusters determine feasibility.

- The constraint may be on:
  - Centers: Restrictions on the number of points a center can service (e.g., capacitated clustering),
  - Clusters: Restrictions on the size of clusters (e.g., balanced clustering),
  - Label-based: Every point has an associated color (*indicating socio-economic groups*), and there are **fairness** restrictions such as proportional representation from each group in every cluster (e.g., fault-tolerant clustering),
  - or, a combination of the above (e.g., fair clustering).

Clustering what matters in constrained settings

The constrained k-median problem

Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, and an integer k, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of every point in X to the nearest feasible center in C is minimized, where constraints on centers/clusters determine feasibility.

- The constraint may be on:
  - Centers: Restrictions on the number of points a center can service (e.g., capacitated clustering),
  - Clusters: Restrictions on the size of clusters (e.g., balanced clustering),
  - Label-based: Every point has an associated color (*indicating socio-economic groups*), and there are **fairness** restrictions such as proportional representation from each group in every cluster (e.g., fault-tolerant clustering),
  - or, a combination of the above (e.g., fair clustering).

The constrained k-median problem

Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, and an integer k, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of every point in X to the nearest feasible center in C is minimized, where constraints on centers/clusters determine feasibility.

- The constraint may be on:
  - Centers: Restrictions on the number of points a center can service (e.g., capacitated clustering),
  - Clusters: Restrictions on the size of clusters (e.g., balanced clustering),
  - Label-based: Every point has an associated color (*indicating socio-economic groups*), and there are **fairness** restrictions such as proportional representation from each group in every cluster (e.g., fault-tolerant clustering),
  - or, a combination of the above (e.g., fair clustering).

Clustering what matters in constrained settings

Problem	Description
Unconstrained k-median (*Constraint type: unconstrained*)	*Input*: (F, X, k) *Output*: $(X_1, ..., X_k, f_1, ..., f_k)$ *Constraints*: None, i.e., $\mathsf{check}(X_1, ..., X_k, f_1, ..., f_k)$ always equals 1. *Objective*: Minimise $\sum_i \sum_{x \in X_i} D(x, f_i)$. (This includes various versions corresponding to specific metrics such as Ulam metric on permutations, metric spaces with constant doubling dimension etc.)
Fault-tolerant k-median (*Constraint type: unconstrained but labelled*) [21, 23]	*Input*: (F, X, k) and a number $h(x) \leq k$ for every facility $x \in X$ *Output*: $(f_1, ..., f_k)$ *Constraints*: None. *Objective*: Minimise $\sum_{x \in X} \sum_{j=1}^{h(x)} D(x, f_{\pi_x(j)})$, where $\pi_x(j)$ is the index of j^{th} nearest center to x in $(f_1, ..., f_k)$ (*Label*: $h(x)$ may be regarded as the label of the client x. So, the number of distinct labels $\ell \leq k$.)
Balanced k-median (*Constraint type: size*) [1, 18]	*Input*: (F, X, k) and integers $(r_1, ..., r_k)$, $(l_1, ..., l_k)$, *Output*: $(X_1, ..., X_k, f_1, ..., f_k)$ *Constraints*: X_i should have at least r_i and at most l_i clients, i.e., $\mathsf{check}(X_1, ..., X_k, f_1, ..., f_k) = 1$ iff $\forall i, r_i \leq \lvert X_i \rvert \leq l_i$. *Objective*: Minimise $\sum_i \sum_{x \in X_i} D(x, f_i)$. (Versions corresponding to specific values of r_i's and l_i's are known by different names. The version corresponding to $l_1 = ... = l_k = \lvert X \rvert$ is called the r-gather problem and the version where $r_1 = ... = r_k = 0$ is called the l-capacity problem.)
Capacitated k-median (*Constraint type: center + size*) [15]	*Input*: (F, X, k) and with capacity $s(f)$ for every facility $f \in F$ *Output*: $(X_1, ..., X_k, f_1, ..., f_k)$ *Constraints*: The number of clients, X_i, assigned to f_i is at most $s(f_i)$, i.e., $\mathsf{check}(X_1, ..., X_k, f_1, ..., f_k) = 1$ iff $\forall i, \lvert X_i \rvert \leq s(f_i)$. *Objective*: Minimise $\sum_i \sum_{x \in X_i} D(x, f_i)$.
Matroid k-median (*Constraint type: center*) [24, 14]	*Input*: (F, X, k) and a Matroid on F *Output*: $(X_1, ..., X_k, f_1, ..., f_k)$ *Constraints*: The number of clients, X_i, assigned to f_i is at most $s(f_i)$, i.e., $\mathsf{check}(X_1, ..., X_k, f_1, ..., f_k) = 1$ iff $\{f_1, ..., f_k\}$ is an independent set of the Matroid . *Objective*: Minimise $\sum_i \sum_{x \in X_i} D(x, f_i)$.
Strongly private k-median (*Constraint type: label + size*) [27]	*Input*: (F, X, k) and numbers $(l_1, ..., l_w)$. Each client has a label $\in \{1, ..., w\}$. *Output*: $(X_1, ..., X_k, f_1, ..., f_k)$ *Constraints*: Every X_i has at least l_j clients with label j, i.e., $\mathsf{check}(X_1, ..., X_k, f_1, ..., f_k) = 1$ iff $\forall i, j, \lvert X_i \cap S_j \rvert \geq l_j$, where S_j is the set of clients with label j . *Objective*: Minimise $\sum_i \sum_{x \in X_i} D(x, f_i)$. (*Labels*: The number of distinct labels $\ell = w$).
l-diversity k-median (*Constraint type: label + size*) [7]	*Input*: (F, X, k) and a number $l > 1$. Each client has one colour from $\in \{1, ..., w\}$ *Output*: $(X_1, ..., X_k, f_1, ..., f_k)$ *Constraints*: The fraction of clients with colour j in every X_i is at least $1/l$, i.e., $\mathsf{check}(X_1, ..., X_k, f_1, ..., f_k) = 1$ iff $\forall i, j, \lvert X_i \cap S_j \rvert \leq \lvert X_i \rvert / l$, where S_j is the set of clients with colour j . *Objective*: Minimise $\sum_i \sum_{x \in X_i} D(x, f_i)$. (*Labels*: Each colour can be regarded as a label and hence the number of distinct labels $\ell = w$).
Fair k-median (*Constraint type: label + size*) [7, 6]	*Input*: (F, X, k) and fairness values $(\alpha_1, ..., \alpha_w), (\beta_1, ..., \beta_w)$. Each client has colours from $\in \{1, ..., w\}$ *Output*: $(X_1, ..., X_k, f_1, ..., f_k)$ *Constraints*: The fraction of clients with colour j in every X_i is between α_j and β_j, i.e., $\mathsf{check}(X_1, ..., X_k, f_1, ..., f_k) = 1$ iff $\forall i, j, \alpha_j \lvert X_i \rvert \leq \lvert X_i \cap S_j \rvert \leq \beta \lvert X_i \rvert$, where S_j is the set of clients with colour j . *Objective*: Minimise $\sum_i \sum_{x \in X_i} D(x, f_i)$. (There are two versions: (i) each client has a unique label, and (ii) a client can have multiple labels.) (*Labels*: For the first version $\ell = w$ and for the second version $\ell = 2^w$.)

Clustering what matters in constrained settings

The outlier constrained k-median problem

Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, an integer k, and an integer m, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of ~~every point~~ all but m points in X to the nearest feasible center in C is minimized, where constraints on centers/clusters determine feasibility.

- General observation: Gap between developments in outlier versus outlier-free versions of constrained clustering.

Clustering what matters in constrained settings

Problem	Outlier-free	Outlier version [20]	Outlier version [2]	Outlier version This work
Euclidean k-means (i.e., $F = \mathbb{R}^d, X \subset \mathbb{R}^d$)	$(1+\varepsilon)$ [9]	$\times$	$(1+\varepsilon)$	$(1+\varepsilon)$
k-median	$(1+\frac{2}{e}+\varepsilon)$ [14]	$(3+\varepsilon)$	$(1+\frac{2}{e}+\varepsilon)$	$(1+\frac{2}{e}+\varepsilon)$
k-means	$(1+\frac{8}{e}+\varepsilon)$ [14]	$(9+\varepsilon)$	$(1+\frac{8}{e}+\varepsilon)$	$(1+\frac{8}{e}+\varepsilon)$
k-median/means in metrics: (i) constant doubling dimension (ii) metrics induced by graphs of bounded treewidth (iii) metrics induced by graphs that exclude a fixed graph as a minor	$(1+\varepsilon)$ [16]	$(3+\varepsilon)$ k-median $(9+\varepsilon)$ k-means	$(1+\varepsilon)$	$(1+\varepsilon)$
Matroid k-median	$(2+\varepsilon)$ [14]	$(3+\varepsilon)$	$(2+\varepsilon)$	$(2+\varepsilon)$
Colourful k-median	$(1+\frac{2}{e}+\varepsilon)$ [14]	$(3+\varepsilon)$	$(1+\frac{2}{e}+\varepsilon)$	$(1+\frac{2}{e}+\varepsilon)$
Ulam k-median (here $F = X$)	$(2-\delta)$ [11]	$(2+\varepsilon)$	$\times$	$(2-\delta)$
Euclidean Capacitated k-median/means	$(1+\varepsilon)$ [15]	$\times$	$\times$	$(1+\varepsilon)$
Capacitated k-median Capacitated k-means	$(3+\varepsilon)$ $(9+\varepsilon)$ [15]	$\times$ $\times$	$\times$ $\times$	$(3+\varepsilon)$ $(9+\varepsilon)$
Uniform/non-uniform r-gather k-median/means (uniform implies $r_1 = r_2 = \ldots = r_k$) Uniform/non-uniform l-capacity k-median/means (uniform implies $l_1 = l_2 = \ldots = l_k$) Uniform/non-uniform balanced k-median/means (uniform implies $r_1 = r_2 = \ldots = r_k$ and $l_1 = l_2 = \ldots = l_k$) Uniform/non-uniform fault tolerant k-median/means (uniform implies same $h(x)$ for every x) Strongly private k-median/means l-diversity k-median/means Fair k-median/means	$(3+\varepsilon)$ (k-median) $(9+\varepsilon)$ (k-means) [20]	$(3+\varepsilon)$ (k-median) $(9+\varepsilon)$ (k-means)	$\times$ $\times$	$(3+\varepsilon)$ (k-median) $(9+\varepsilon)$ (k-means)

Clustering what matters in constrained settings

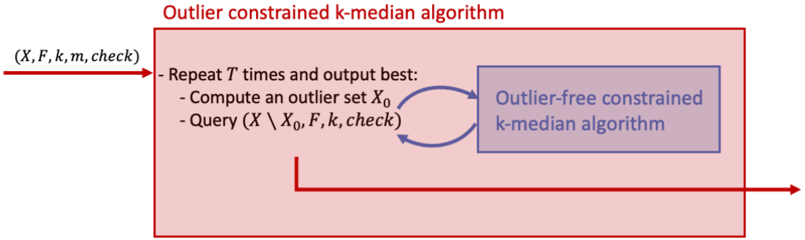
The outlier constrained k-median problem

Let $(\mathcal{X}, D)$ be any metric space. Given a facility set $F \subseteq \mathcal{X}$, a client set $X \subseteq \mathcal{X}$, an integer k, and an integer m, find k points $C \subseteq F$ (*called centers*) such that the sum of distances of ~~every point~~ all but m points in X to the nearest feasible center in C is minimized, where constraints on centers/clusters determine feasibility.

- General observation: Gap between developments in outlier versus outlier-free versions of constrained clustering.
- General goal: Bridge the gap using an approximation-preserving reduction from outlier to outlier-free version.

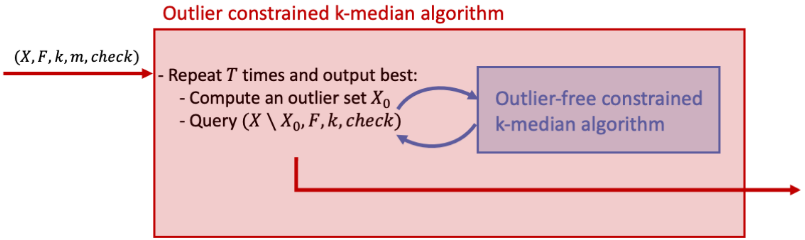
Clustering what matters in constrained settings

- General goal: Bridge the gap using an approximation-preserving reduction from outlier to outlier-free version.

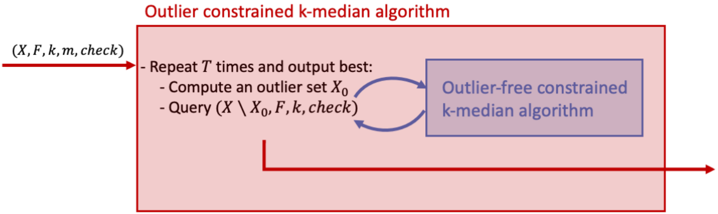


Clustering what matters in constrained settings

- General goal: Bridge the gap using an approximation-preserving reduction from outlier to outlier-free version.
  - Approximation-preserving: α-approximation gives $(1+\varepsilon)\cdot\alpha$-approximation

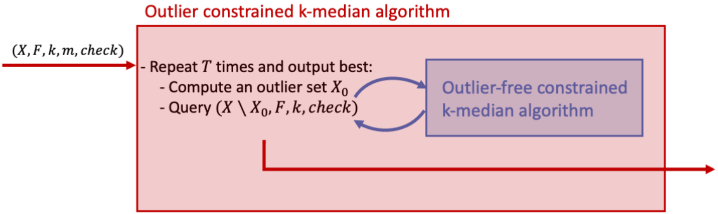


Clustering what matters in constrained settings



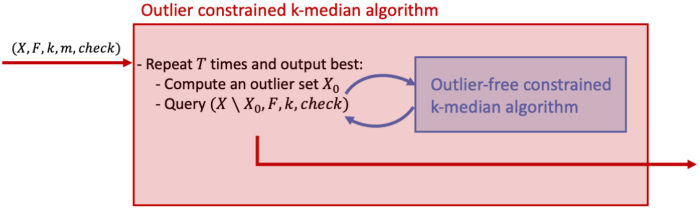
- A trivial reduction: For outlier set X_0 , try all combinations of m points from X.
  - Issue: $T = O(n^m)$, where m is the number of outliers.
  - Ideally, we would want T to be independent of the problem size and dependent only on the parameters k, m, ε.

Clustering what matters in constrained settings



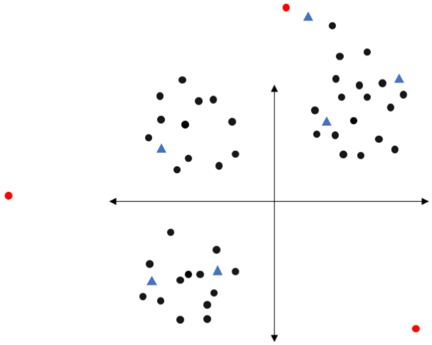
- Better reductions:
  1. Bhattacharya et al. (2020): D^2-sampling based reduction for k-means in the Euclidean setting.
  2. Agrawal et al. (2023): Coreset based reduction for metric spaces.
     - Coreset: Compressed dataset that mimics the k-median cost.
     - $T = \left(\frac{(k+m)\log n}{\varepsilon}\right)^{O(m)}$
     - Issue: Constrained setting.

Clustering what matters in constrained settings



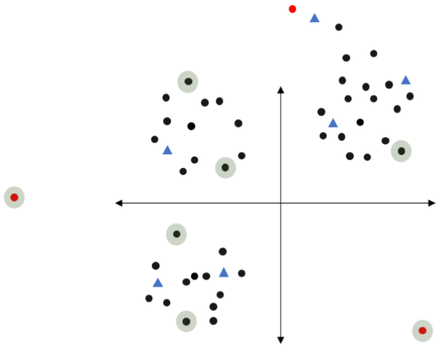
- Better reductions:
  1. Bhattacharya et al. (2020): D^2-sampling based reduction for k-means in the Euclidean setting.
  2. Agrawal et al. (2023): Coreset based reduction for metric spaces.
     - Coreset: Compressed dataset that mimics the k-median cost.
     - $T = \left(\frac{(k+m) \cdot \log n}{\varepsilon}\right)^{O(m)}$
     - Issue: Constrained setting.
  3. This work: D^z-sampling based reduction for metric space in constrained settings.
     - $T = \left(\frac{(k+m)}{\varepsilon}\right)^{O(m)}$

The Reduction: Key Ideas



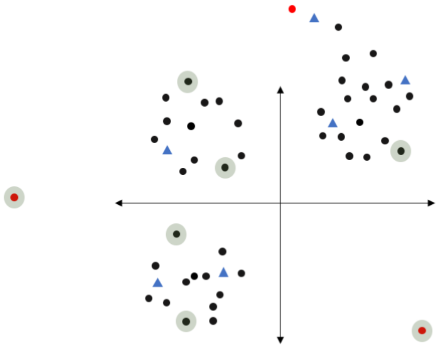
1. Start with a $(k+m)$ centers C that give constant approximation to the unconstrained $(k+m)$-median problem. (*see* ▲ *in Figure*)
   - Interesting observation: C gives constant approximation for the outlier version.

The Reduction: Key Ideas



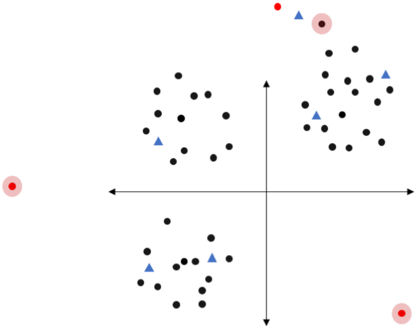
1. Start with a $(k+m)$ centers C that give constant approximation to the unconstrained $(k+m)$-median problem. (*see* ▲ *in Figure*)
2. D-sample $O(m \log m)$ points $S \subseteq X$ with respect to C. (*see* ● *in Figure*)
   - D-sampling: The probability of a point being sampled is proportional to its distance from the nearest center in C.

The Reduction: Key Ideas



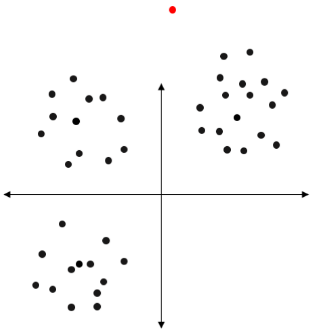
1. Start with a $(k + m)$ centers C that give constant approximation to the unconstrained $(k + m)$-median problem. (*see* ▲ *in Figure*)
2. D-sample $O(m \log m)$ points $S \subseteq X$ with respect to C. (*see* ● *in Figure*)
   - Observation: Outliers that are far from C get sampled in S, which can be located by trying out all subsets of S.

The Reduction: Key Ideas



1. Start with a $(k+m)$ centers C that give constant approximation to the unconstrained $(k+m)$-median problem.
2. D-sample $O(m \log m)$ points $S \subseteq X$ with respect to C.
   - Observation: Outliers that are far from C get sampled in S, which can be located by trying out all subsets of S.
3. For outliers close to C, locate appropriate replacement by matching.

The Reduction: Key Ideas



1. Start with a $(k+m)$ centers C that give constant approximation to the unconstrained $(k+m)$-median problem.
2. D-sample $O(m \log m)$ points $S \subseteq X$ with respect to C.
   - Observation: Outliers that are far from C get sampled in S, which can be located by trying out all subsets of S.
3. For outliers close to C, locate appropriate outlier replacement by matching.

The Reduction: Key Ideas

Algorithm sketch

1. Start with a $(k+m)$ centers C that give constant approximation to the unconstrained $(k+m)$-median problem.
2. D-sample $O(m \log m)$ points $S \subseteq X$ with respect to C.
   - Observation: Outliers that are far from C get sampled in S, which can be located by trying out all subsets of S.
3. For outliers close to C, locate appropriate outlier replacement by matching.

- Our reduction generalizes to the k-means problem and a wide range of center/size/label-based constrained settings.
- Our reduction matches the best-known approximation bounds for several constrained problems and gives the best results for others (e.g., capacitated k-median.)

See paper for details...

References

- Agrawal, A., Inamdar, T., Saurabh, S., and Xue, J. (2023). Clustering what matters: Optimal approximation for clustering with outliers. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(6):6666–6674.
- Bhattacharya, A., Goyal, D., Jaiswal, R., and Kumar, A. (2020). On Sampling Based Algorithms for k-Means. In Saxena, N. and Simon, S., editors, *40th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2020)*, volume 182 of Leibniz International Proceedings in Informatics (LIPIcs), pages 13:1–13:17, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz–Zentrum fuer Informatik.
- Byrka, J., Pensyl, T., Rybicki, B., Srinivasan, A., and Trinh, K. (2017). An improved approximation for k-median and positive correlation in budgeted optimization. *ACM Trans. Algorithms*, 13(2).
- Cohen-Addad, V., Gupta, A., Kumar, A., Lee, E., and Li, J. (2019). Tight FPT Approximations for k-Median and k-Means. In Baier, C., Chatzigiannakis, I., Flocchini, P., and Leonardi, S., editors, *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, volume 132 of Leibniz International Proceedings in Informatics (LIPIcs), pages 42:1–42:14, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz–Zentrum fuer Informatik.
- Guha, S. and Khuller, S. (1999). Greedy strikes back: Improved facility location algorithms. *Journal of Algorithms*, 31(1):228–248.
- Krishnaswamy, R., Li, S., and Sandeep, S. (2018). Constant approximation for k-median and k-means with outliers via iterative rounding. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 646–659, New York, NY, USA. Association for Computing Machinery.

Thank you