Modified K-means Algorithm with Local Optimality Guarantees

Mingyi Li[1] Michael R. Metel[2] Akiko Takeda[1,3]

[1]The University of Tokyo [2]Huawei Noah's Ark Lab [3]RIKEN AIP

Summary & Contribution

- K-means is a classic, widely used clustering algorithm.
- We show by counterexample that **K-means does not necessarily converge to a locally optimal solution, let alone a global one**.

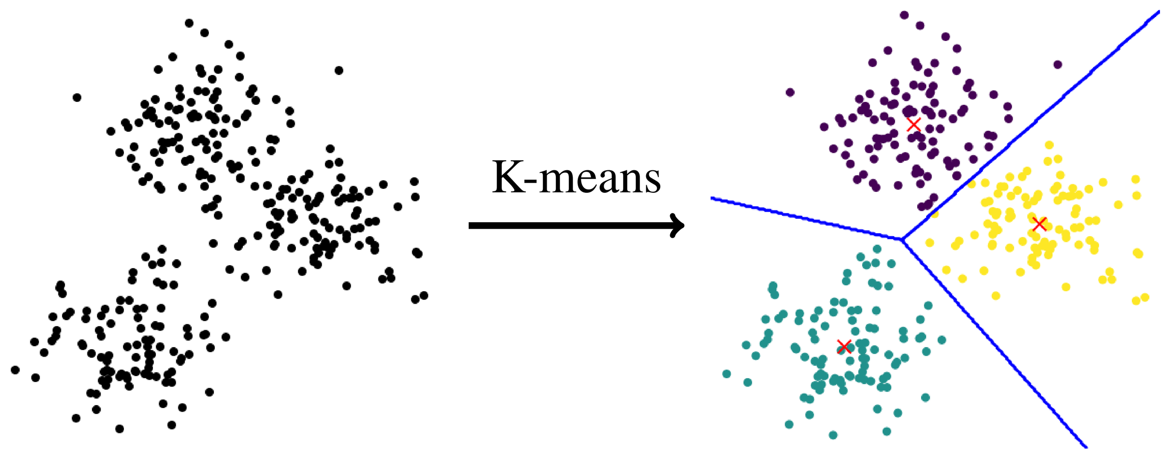
LO-K-means (Our Algorithm)

- **A simple modification** to the K-means algorithm that ensures local optimality with no additional complexity.
- Analysis of two local-optimality criteria—**continuous (C-local)** and **discrete (D-local)**—shows experimentally that LO-K-means consistently improves clustering quality.

Introduction

The K-means Clustering. Partition a set of N data points $X = \{x_i\}_{i=1}^N$ with weights $W = \{w_i\}_{i=1}^N$ into K distinct clusters by minimizing the total Bregman divergence to the cluster centers.

$$\min_{P,C} \; f(P,C) = \sum_{k=1}^{K} \sum_{n=1}^{N} p_{k,n} w_n D(x_n, c_k)$$



The K-means Algorithm (Lloyd, 1982).

1. Select K initial centers arbitrarily from X.
2. Assign each data point to the cluster with the nearest center.
3. Recalculate the center for each cluster as the mean of its assigned points.
4. Repeat 2 and 3 until cluster assignments no longer change.

Continuous Relaxation. Once an assignment matrix P is fixed, the optimal centers C are uniquely determined. Since $F(P) := \min_C f(P,C)$ is concave, relaxing P from $\{0,1\}^{K\times N}$ to $[0,1]^{K\times N}$ yields an equivalent continuous K-means formulation with the same optimal clustering loss.

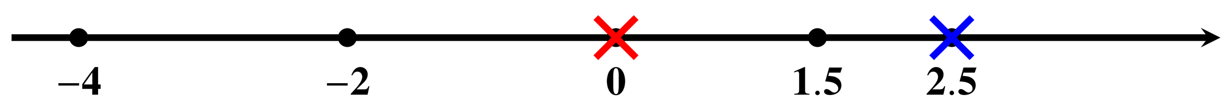
Common Misconception:

Although K-means is commonly assumed (e.g., in scikit-learn) to converge to a locally optimal solution, **it can fail.**

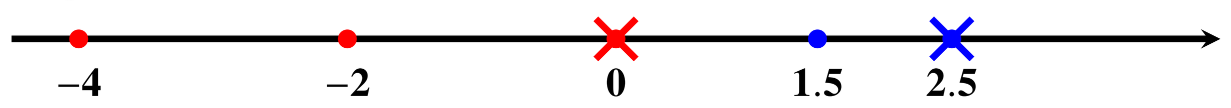
- **The most-cited proof (Selim & Ismail, 1984) for local optimality has some flaws.**

Counterexample

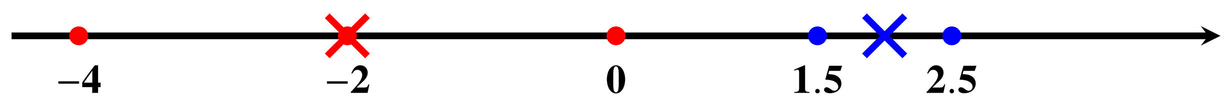
Initial setup: $N = 5, K = 2, x = \{-4, -2, 0, 1.5, 2.5\}$,
Initial centers: $c_1 = 0, c_2 = 2.5$.



Update assignment



Update center & Converged to $c_1^* = -2,\ c_2^* = 2$



Not a Locally Optimal Solution!

- Shifting a small part of point 0 to the other cluster can further reduce the clustering loss.

Theoretical Guarantees

Two Definitions of Local Optimality:

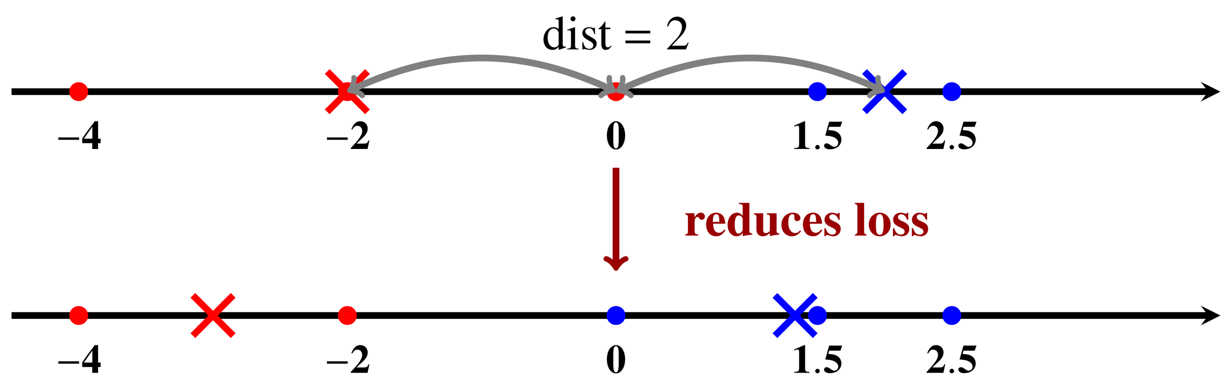
- **C-local:** (P,C) is a local optimum in the continuous relaxation.
  i.e. no P' with $\|P' - P\| \le \varepsilon$ such that $F(P') < F(P)$.
- **D-local:** (P,C) is a local optimum in hard clustering.
  i.e. no P' adjacent to P such that $F(P') < F(P)$.

Key Condition for Local Optimality:

K-means solution (P,C) is C-local.
$\Leftrightarrow$ The optimal assignment for the solution centers C is **unique**.

- If the optimal assignment for centers C is not unique, then **switching to any other assignment strictly decreases the clustering loss**.

- Simply check if any point is at the same distance from two or more centers.
- Compute the exact change in loss when moving a single point to another cluster by **a simple explicit formula**.



Numerical Experiments

- Guarantees convergence to local optimality (both continuous and discrete).
- Same per-iteration complexity as the original K-means $O(NKd)$.

Algorithms:

- K-means
- **C-LO** (LO-K-means; guarantees C-local optimality)
- **D-LO** (LO-K-means; guarantees D-local optimality)
- **Min-D-LO** (LO-K-means; guarantees D-local and enhances D-LO)

■ Synthetic Datasets

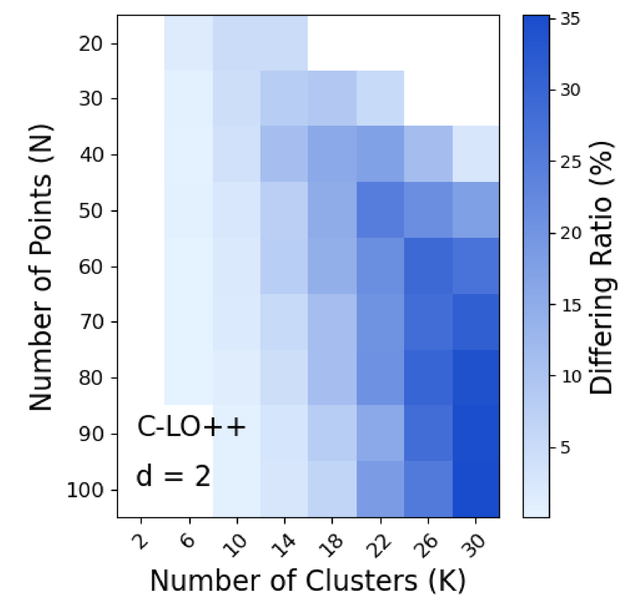


Figure 1. Proportion of runs where C-LO outperforms K-means (squared Euclidean; 1,000 trials).

- K-means sometimes fails to converge to a C-local solution.

■ Real-World Datasets

Table 1. Clustering loss (mean, min), runtime, and iterations for K-means, D-LO, and Min-D-LO (squared Euclidean; 20 trials) on real datasets.

	Dataset	Iris (N = 150, d = 4)				News20 (N = 2,000, d = 1,089)			
K	Algorithm	Mean	Minimum	Time(s)	Num Iter	Mean	Minimum	Time(s)	Num Iter
10	K-means++	29.57	26.01	< 0.001	7	697,527	643,583	0.48	23
	D-LO++	**28.92**	**25.94**	< 0.001	17	**634,216**	**625,467**	6.18	288
	Min-D-LO++	28.93	**25.94**	< 0.001	17	634,293	625,468	2.55	125
25	K-means++	13.73	12.70	< 0.001	6	529,028	487,823	1.25	26
	D-LO++	**12.58**	**11.83**	< 0.001	31	475,299	468,201	35.96	705
	Min-D-LO++	12.61	12.07	< 0.001	27	**474,431**	**467,745**	15.77	316
50	K-means++	6.40	5.52	< 0.001	5	439,029	418,754	3.02	31
	D-LO++	**5.36**	**5.04**	0.002	37	**392,016**	**388,746**	157.97	1,228
	Min-D-LO++	5.40	**5.04**	0.002	30	392,146	388,990	60.41	533

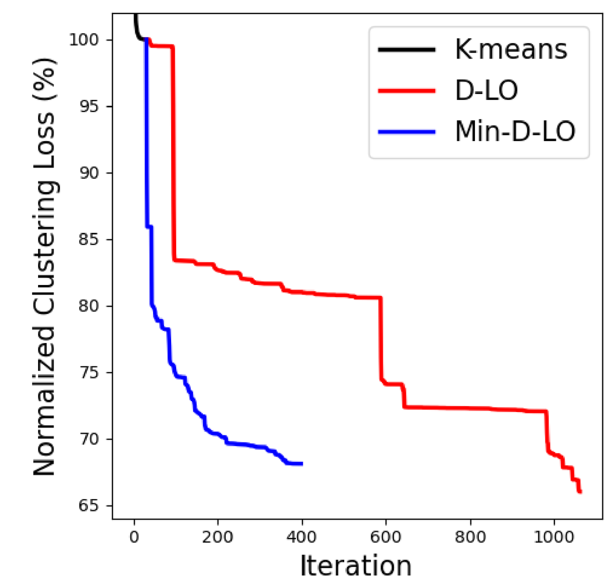


Figure 2. Clustering loss progression per iteration for K-means, D-LO, and Min-D-LO on News20 (N = 2000, d = 1089, K = 10).

- **Our methods (D-LO, Min-D-LO)** consistently find solutions with lower clustering losses.
- Even with a practical iteration limit (e.g., 300 iterations), our methods still provide significant accuracy improvements of over 15-25%.