

# Theory, methods and applications of data clustering

Rudolf Scitovski

Department of Mathematics, University of Osijek

July 7, 2010

# Problem statement

$S = \{x_i \in \mathbb{R}^n : i = 1, \dots, m\} \subseteq \mathbb{R}^n$  – a set of data vectors

$\Pi(S) = \{\pi_1, \dots, \pi_k\}$  – partition of the set  $S$

$\pi_1, \dots, \pi_k$  – clusters

# Problem statement

$S = \{x_i \in \mathbb{R}^n : i = 1, \dots, m\} \subseteq \mathbb{R}^n$  – a set of data vectors

$\Pi(S) = \{\pi_1, \dots, \pi_k\}$  – partition of the set  $S$

$\pi_1, \dots, \pi_k$  – clusters

$$\bigcup_{i=1}^k \pi_i = S \quad (1)$$

$$\pi_i \cap \pi_j = \emptyset, \quad i \neq j \quad (2)$$

$$|\pi_j| \geq 1, \quad j = 1, \dots, k \quad (3)$$

# Problem statement

$S = \{x_i \in \mathbb{R}^n : i = 1, \dots, m\} \subseteq \mathbb{R}^n$  – a set of data vectors

$\Pi(S) = \{\pi_1, \dots, \pi_k\}$  – partition of the set  $S$

$\pi_1, \dots, \pi_k$  – clusters

$$\bigcup_{i=1}^k \pi_i = S \quad (1)$$

$$\pi_i \cap \pi_j = \emptyset, \quad i \neq j \quad (2)$$

$$|\pi_j| \geq 1, \quad j = 1, \dots, k \quad (3)$$

$\Psi(S, k)$  – the set of all partitions of the length  $k$

# Problem statement

Objective function:

$$F : \Psi(S, k) \rightarrow [0, +\infty), \quad F(\Pi) = \sum_{j=1}^k f(\pi_j)$$

# Problem statement

Objective function:

$$F : \Psi(S, k) \rightarrow [0, +\infty), \quad F(\Pi) = \sum_{j=1}^k f(\pi_j)$$

$$f : \mathcal{P}(S) \rightarrow [0, +\infty)$$

convex or quasiconvex function defined  
on the partitive set  $\mathcal{P}(S)$

# Problem statement

Objective function:

$$F : \Psi(S, k) \rightarrow [0, +\infty), \quad F(\Pi) = \sum_{j=1}^k f(\pi_j)$$

$$f : \mathcal{P}(S) \rightarrow [0, +\infty)$$

convex or quasiconvex function defined  
on the partitive set  $\mathcal{P}(S)$

Optimal partition  $\Pi^{(o)}$ :

$$F(\Pi^{(o)}) = \min_{\Pi \in \Psi(S, k)} F(\Pi).$$

# Least Squares – optimal partition

$$f(\pi_j) = \sum_{x_i \in \pi_j} \|x_i - c_j\|_2^2,$$

$$c_j = \operatorname{argmin}_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_2^2 = \frac{1}{|\pi_j|} \sum_{x_i \in \pi_j} x_i \in \mathbb{R}^n \quad (\text{centroid of the cluster } \pi_j)$$

# Least Squares – optimal partition

$$f(\pi_j) = \sum_{x_i \in \pi_j} \|x_i - c_j\|_2^2,$$

$$c_j = \operatorname{argmin}_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_2^2 = \frac{1}{|\pi_j|} \sum_{x_i \in \pi_j} x_i \in \mathbb{R}^n \quad (\text{centroid of the cluster } \pi_j)$$

$$\min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_2^2 = \min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \left( \sum_{s=1}^n (x_s^{(i)} - a_s)^2 \right)$$

# Least Squares – optimal partition

$$f(\pi_j) = \sum_{x_i \in \pi_j} \|x_i - c_j\|_2^2,$$

$$c_j = \operatorname{argmin}_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_2^2 = \frac{1}{|\pi_j|} \sum_{x_i \in \pi_j} x_i \in \mathbb{R}^n \quad (\text{centroid of the cluster } \pi_j)$$

$$\begin{aligned} \min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_2^2 &= \min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \left( \sum_{s=1}^n (x_s^{(i)} - a_s)^2 \right) \\ &= \min_{a \in \mathbb{R}^n} \sum_{s=1}^n \left( \sum_{x_i \in \pi_j} (x_s^{(i)} - a_s)^2 \right) = \sum_{s=1}^n \min_{a_s \in \mathbb{R}} \sum_{x_i \in \pi_j} (x_s^{(i)} - a_s)^2 \end{aligned}$$

# Least Squares – optimal partition

$$f(\pi_j) = \sum_{x_i \in \pi_j} \|x_i - c_j\|_2^2,$$

$$c_j = \operatorname{argmin}_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_2^2 = \frac{1}{|\pi_j|} \sum_{x_i \in \pi_j} x_i \in \mathbb{R}^n \quad (\text{centroid of the cluster } \pi_j)$$

$$\begin{aligned} \min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_2^2 &= \min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \left( \sum_{s=1}^n (x_s^{(i)} - a_s)^2 \right) \\ &= \min_{a \in \mathbb{R}^n} \sum_{s=1}^n \left( \sum_{x_i \in \pi_j} (x_s^{(i)} - a_s)^2 \right) = \sum_{s=1}^n \min_{a_s \in \mathbb{R}} \sum_{x_i \in \pi_j} (x_s^{(i)} - a_s)^2 \\ &= \sum_{s=1}^n \sum_{x_i \in \pi_j} \left( x_s^{(i)} - A(x_s^{(i)}) \right)^2 = \sum_{x_i \in \pi_j} \sum_{s=1}^n \left( x_s^{(i)} - A(x_s^{(i)}) \right)^2 = \sum_{x_i \in \pi_j} \|x_i - a^*\|_2^2, \end{aligned}$$

# Least Squares – optimal partition

$$f(\pi_j) = \sum_{x_i \in \pi_j} \|x_i - c_j\|_2^2,$$

$$c_j = \operatorname{argmin}_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_2^2 = \frac{1}{|\pi_j|} \sum_{x_i \in \pi_j} x_i \in \mathbb{R}^n \quad (\text{centroid of the cluster } \pi_j)$$

$$\begin{aligned} \min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_2^2 &= \min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \left( \sum_{s=1}^n (x_s^{(i)} - a_s)^2 \right) \\ &= \min_{a \in \mathbb{R}^n} \sum_{s=1}^n \left( \sum_{x_i \in \pi_j} (x_s^{(i)} - a_s)^2 \right) = \sum_{s=1}^n \min_{a_s \in \mathbb{R}} \sum_{x_i \in \pi_j} (x_s^{(i)} - a_s)^2 \\ &= \sum_{s=1}^n \sum_{x_i \in \pi_j} \left( x_s^{(i)} - A \left( x_s^{(i)} \right) \right)^2 = \sum_{x_i \in \pi_j} \sum_{s=1}^n \left( x_s^{(i)} - A \left( x_s^{(i)} \right) \right)^2 = \sum_{x_i \in \pi_j} \|x_i - a^*\|_2^2, \\ a^* &= \left( \frac{1}{|\pi_j|} \sum_{x_i \in \pi_j} x_1^{(i)}, \dots, \frac{1}{|\pi_j|} \sum_{x_i \in \pi_j} x_n^{(i)} \right)^T = \frac{1}{|\pi_j|} \sum_{x_i \in \pi_j} x_i \end{aligned}$$

# Least Squares – optimal partition

- I. S. Dhillon, Y. Guan, B. Kulis, *Kernel  $k$ -means, spectral clustering and normalized cuts*, 2004.
- G. Gan, C Ma, J. Wu, *Data clustering: theory, algorithms, and applications*, SIAM, Philadelphia, 2007.
- L. Kaufman, P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*, John Wiley & Sons, Hoboken, 2005.
- J. Kogan, *Introduction to Clustering Large and High-Dimensional Data*, Cambridge University Press, 2007.
- H. Späth, *Cluster-Formation und -Analyse*, R. Oldenburg Verlag, München, 1983.

# Least Absolute Deviations – optimal partition

$$f(\pi_j) = \sum_{x_i \in \pi_j} \|x_i - c_j\|_1,$$

$$c_j = \operatorname{argmin}_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_1 = \left( \operatorname{med}_i x_1^{(i)}, \dots, \operatorname{med}_i x_n^{(i)} \right)^T \in \mathbb{R}^n,$$

# Least Absolute Deviations – optimal partition

$$f(\pi_j) = \sum_{x_i \in \pi_j} \|x_i - c_j\|_1,$$

$$c_j = \operatorname{argmin}_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_1 = \left( \operatorname{med}_i x_1^{(i)}, \dots, \operatorname{med}_i x_n^{(i)} \right)^T \in \mathbb{R}^n,$$

$$\min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_1 = \min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \left( \sum_{s=1}^n |x_s^{(i)} - a_s| \right)$$

# Least Absolute Deviations – optimal partition

$$f(\pi_j) = \sum_{x_i \in \pi_j} \|x_i - c_j\|_1,$$

$$c_j = \operatorname{argmin}_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_1 = \left( \operatorname{med}_i x_1^{(i)}, \dots, \operatorname{med}_i x_n^{(i)} \right)^T \in \mathbb{R}^n,$$

$$\begin{aligned} \min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_1 &= \min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \left( \sum_{s=1}^n |x_s^{(i)} - a_s| \right) \\ &= \min_{a \in \mathbb{R}^n} \sum_{s=1}^n \left( \sum_{x_i \in \pi_j} |x_s^{(i)} - a_s| \right) = \sum_{s=1}^n \min_{a_s \in \mathbb{R}} \sum_{x_i \in \pi_j} |x_s^{(i)} - a_s| \end{aligned}$$

# Least Absolute Deviations – optimal partition

$$f(\pi_j) = \sum_{x_i \in \pi_j} \|x_i - c_j\|_1,$$

$$c_j = \operatorname{argmin}_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_1 = \left( \operatorname{med}_i x_1^{(i)}, \dots, \operatorname{med}_i x_n^{(i)} \right)^T \in \mathbb{R}^n,$$

$$\begin{aligned} \min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_1 &= \min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \left( \sum_{s=1}^n |x_s^{(i)} - a_s| \right) \\ &= \min_{a \in \mathbb{R}^n} \sum_{s=1}^n \left( \sum_{x_i \in \pi_j} |x_s^{(i)} - a_s| \right) = \sum_{s=1}^n \min_{a_s \in \mathbb{R}} \sum_{x_i \in \pi_j} |x_s^{(i)} - a_s| \\ &= \sum_{s=1}^n \sum_{x_i \in \pi_j} |x_s^{(i)} - \operatorname{med}_{i=1, \dots, m} x_s^{(i)}| = \sum_{x_i \in \pi_j} \sum_{s=1}^n |x_s^{(i)} - \operatorname{med}_{i=1, \dots, m} x_s^{(i)}| \end{aligned}$$

# Least Absolute Deviations – optimal partition

$$f(\pi_j) = \sum_{x_i \in \pi_j} \|x_i - c_j\|_1,$$

$$c_j = \operatorname{argmin}_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_1 = \left( \operatorname{med}_i x_1^{(i)}, \dots, \operatorname{med}_i x_n^{(i)} \right)^T \in \mathbb{R}^n,$$

$$\begin{aligned} \min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \|x_i - a\|_1 &= \min_{a \in \mathbb{R}^n} \sum_{x_i \in \pi_j} \left( \sum_{s=1}^n |x_s^{(i)} - a_s| \right) \\ &= \min_{a \in \mathbb{R}^n} \sum_{s=1}^n \left( \sum_{x_i \in \pi_j} |x_s^{(i)} - a_s| \right) = \sum_{s=1}^n \min_{a_s \in \mathbb{R}} \sum_{x_i \in \pi_j} |x_s^{(i)} - a_s| \\ &= \sum_{s=1}^n \sum_{x_i \in \pi_j} |x_s^{(i)} - \operatorname{med}_{i=1, \dots, m} x_s^{(i)}| = \sum_{x_i \in \pi_j} \sum_{s=1}^n |x_s^{(i)} - \operatorname{med}_{i=1, \dots, m} x_s^{(i)}| \\ &= \sum_{x_i \in \pi_j} \|x_i - a^*\|_1, \quad a^* = \left( \operatorname{med}_i x_1^{(i)}, \dots, \operatorname{med}_i x_n^{(i)} \right)^T \in \mathbb{R}^n \end{aligned}$$

# Least Absolute Deviations – optimal partition

- H. Späth, *Clusterwise linear least absolute deviations regression*, Computing **37**(1986), 371–378
- R. Cupec, R. Grbić, K. Sabo, R. Scitovski, *Three points method for searching the best least absolute deviations plane*, Applied Mathematics and Computation **215**(2009), 983–994
- K. Sabo, R. Scitovski, *The best least absolute deviations line – properties and two efficient methods*, ANZIAM Journal **50**(2008), 185–198
- R. Cupec, R. Grbic, K. E. Nyarko, K. Sabo, and R. Scitovski, *Detection of planar surfaces based on RANSAC and LAD plane fitting*, European Conference on Mobile Robotics ECMR'09

# Number of all partitions

The problem of finding an optimal partition is an NP-hard problem.

An exact and efficient method for finding an optimal partition does not exist.

The number of all partitions of the set  $S$  containing  $k$  clusters:

$$|\Psi(S, k)| = \frac{1}{k!} \sum_{j=1}^k (-1)^{k-j} \binom{k}{j} j^m.$$

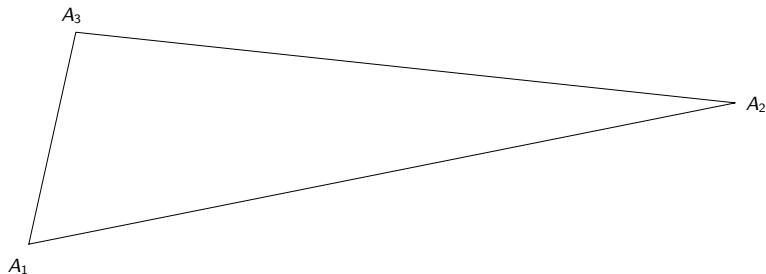
| $ \Psi(S, k) $ | $k = 2$         | $k = 3$         | $k = 5$         | $k = 8$         | $k = 10$    |
|----------------|-----------------|-----------------|-----------------|-----------------|-------------|
| $m = 10$       | 511             | 9330            | 42525           | 750             | 1           |
| $m = 50$       | $10^{15}$       | $10^{23}$       | $10^{33}$       | $10^{40}$       | $10^{43}$   |
| $m = 10^3$     | $10^{300}$      | $10^{476}$      | $10^{697}$      | $10^{898}$      | $10^{993}$  |
| $m = 10^6$     | $10^{301\,029}$ | $10^{477\,120}$ | $10^{698\,968}$ | $10^{903\,085}$ | $10^{10^6}$ |

The applications are important and very attractive (biology, medicine, robotics, marketing, education, ...).

Approaches: Algebraic and Combinatorial Optimization, Statistics, Graph Theory, Fuzzy Clustering

# Historical remark

## Fermat–Torricelli–Weber problem

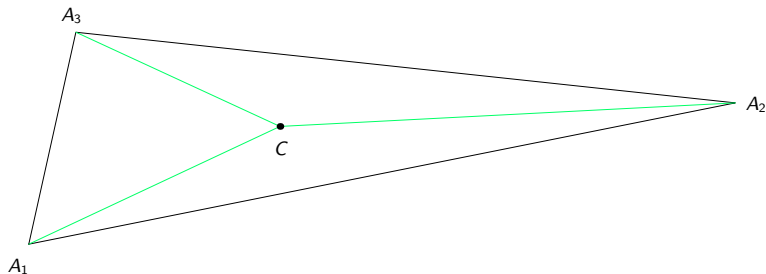


$$C : \sum_{i=1}^3 d_2^2(A_i, T(x, y)) \rightarrow \min_{x, y}$$

$$M : \sum_{i=1}^3 d_2(A_i, T(x, y)) \rightarrow \min_{x, y}$$

# Historical remark

## Fermat–Torricelli–Weber problem

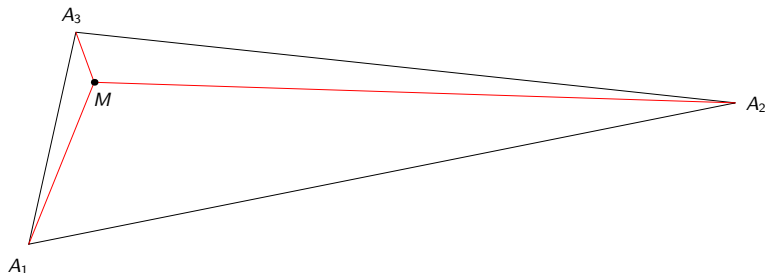


$$C : \sum_{i=1}^3 d_2^2(A_i, T(x, y)) \rightarrow \min_{x, y}$$

$$M : \sum_{i=1}^3 d_2(A_i, T(x, y)) \rightarrow \min_{x, y}$$

# Historical remark

## Fermat–Torricelli–Weber problem

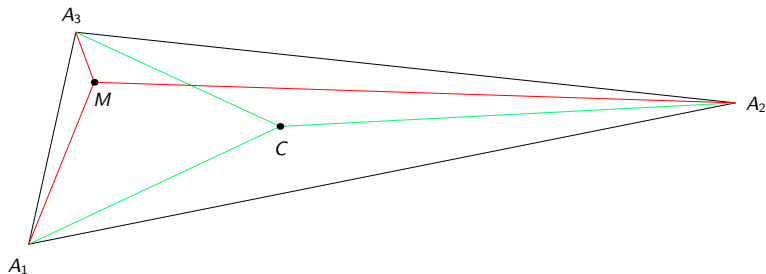


$$C : \sum_{i=1}^3 d_2^2(A_i, T(x, y)) \rightarrow \min_{x, y}$$

$$M : \sum_{i=1}^3 d_2(A_i, T(x, y)) \rightarrow \min_{x, y}$$

# Historical remark

## Fermat–Torricelli–Weber problem



$$C : \sum_{i=1}^3 d_2^2(A_i, T(x, y)) \rightarrow \min_{x, y}$$

$$M : \sum_{i=1}^3 d_2(A_i, T(x, y)) \rightarrow \min_{x, y}$$

# Clusters on $\mathbb{R}$

$$\|x\| = |x|, \quad x \in \mathbb{R}, \quad (\text{norm on } \mathbb{R})$$

# Clusters on $\mathbb{R}$

$$\|x\| = |x|, \quad x \in \mathbb{R}, \quad (\text{norm on } \mathbb{R})$$

$$d(x, y) = |x - y| \quad (\text{distance function on } \mathbb{R})$$

# Clusters on $\mathbb{R}$

$$\|x\| = |x|, \quad x \in \mathbb{R}, \quad (\text{norm on } \mathbb{R})$$

$$d(x, y) = |x - y| \quad (\text{distance function on } \mathbb{R})$$

$$\Pi = \{\pi_1, \dots, \pi_k\} \quad (\text{partition of the set } S = \{x_i \in \mathbb{R} : i = 1, \dots, m\})$$

# Clusters on $\mathbb{R}$

$$\|x\| = |x|, \quad x \in \mathbb{R}, \quad (\text{norm on } \mathbb{R})$$

$$d(x, y) = |x - y| \quad (\text{distance function on } \mathbb{R})$$

$$\Pi = \{\pi_1, \dots, \pi_k\} \quad (\text{partition of the set } S = \{x_i \in \mathbb{R} : i = 1, \dots, m\})$$

Centers  $c_j$  and objective function on the clusters  $\pi_j, j = 1, \dots, k$ :

$$c_j = \operatorname{argmin}_{\alpha} \sum_{x_i \in \pi_j} |x_i - \alpha| = \operatorname{med}_{x_i \in \pi_j} x_i, \quad f(\pi_j) = \sum_{x_i \in \pi_j} |x_i - \operatorname{med}_{x_i \in \pi_j} x_i|,$$

$$c_j^{LS} = \operatorname{argmin}_{\alpha} \sum_{x_i \in \pi_j} (x_i - \alpha)^2 = \frac{1}{|\pi_j|} \sum_{x_i \in \pi_j} x_i, \quad f_{LS}(\pi_j) = \sum_{x_i \in \pi_j} (x_i - c_j^{LS})^2.$$

## Example

*Each student is assigned the total number of points scored in the tests throughout the semester. Students should be divided into 4 clusters according to the number of points scored: grade (5), grade (4), grade (3) and grade (2).*

$x = \{418, 385, 357, 351, 347, 333, 330, 323, 321, 318, 308, 308, 302, 302, 300, 294, 293, 290, 287, 286, 286, 277, 271, 271, 269, 268, 266, 260, 253, 250, 248, 237, 237, 233, 232, 230, 228, 227, 226, 217, 214, 211, 208, 207, 187, 186, 181, 179, 175, 165, 161\}$

# Clusters on $\mathbb{R}$

Five partitions are selected (ways of dividing students into groups based upon the same grades):

$\Pi_1$  – defined by *Mathematica* function `FindClusters` with method “Optimize”,

$\Pi_2$  – defined by *Mathematica* function `FindClusters` with method  
“Agglomerate”,

$\Pi_3, \Pi_4, \Pi_5$  – by experience

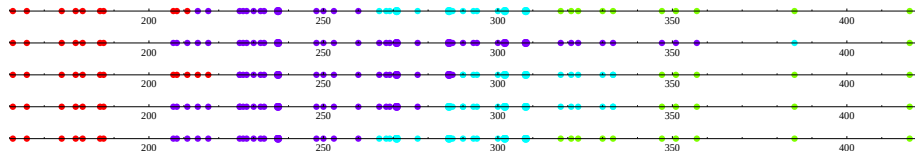
# Clusters on $\mathbb{R}$

Five partitions are selected (ways of dividing students into groups based upon the same grades):

$\Pi_1$  – defined by *Mathematica* function FindClusters with method “Optimize”,

$\Pi_2$  – defined by *Mathematica* function FindClusters with method  
“Agglomerate”,

$\Pi_3, \Pi_4, \Pi_5$  – by experience



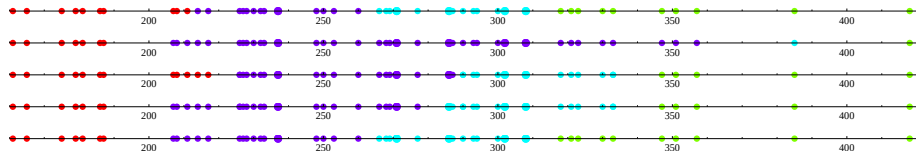
# Clusters on $\mathbb{R}$

Five partitions are selected (ways of dividing students into groups based upon the same grades):

$\Pi_1$  – defined by *Mathematica* function FindClusters with method “Optimize”,

$\Pi_2$  – defined by *Mathematica* function FindClusters with method  
“Agglomerate”,

$\Pi_3, \Pi_4, \Pi_5$  – by experience



| Partition | excellent      | very good      | good           | sufficient     | $FLAD$ | $FLS$ |
|-----------|----------------|----------------|----------------|----------------|--------|-------|
| $\Pi_1$   | [318-418] (10) | [266-308] (17) | [214-260] (14) | [161-211] (10) | 718    | 17657 |
| $\Pi_2$   | [418-418] (1)  | [385-385] (1)  | [207-357] (42) | [161-187] (7)  | 1545   | 74369 |
| $\Pi_3$   | [347-418] (5)  | [290-333] (15) | [226-290] (22) | [161-217] (12) | 936    | 22054 |
| $\Pi_4$   | [347-418] (5)  | [286-333] (18) | [207-277] (24) | [161-187] (7)  | 854    | 20348 |
| $\Pi_5$   | [318-418] (10) | [266-308] (17) | [207-260] (17) | [161-187] (7)  | 701    | 17187 |

# Clusters on $\mathbb{R}^2$

$$\|x\|_p = (|x_1|^p + |x_2|^p)^{\frac{1}{p}}, \quad x = (x_1, x_2)^T \in \mathbb{R}^2, \quad p \in [1, \infty) \cup \{+\infty\} \quad (\text{norm on } \mathbb{R}^2)$$

# Clusters on $\mathbb{R}^2$

$$\|x\|_p = (|x_1|^p + |x_2|^p)^{\frac{1}{p}}, \quad x = (x_1, x_2)^T \in \mathbb{R}^2, \quad p \in [1, \infty) \cup \{+\infty\} \quad (\text{norm on } \mathbb{R}^2)$$

$$d_p(x, y) = \|x - y\|_p \quad (\text{distance function on } \mathbb{R}^2)$$

# Clusters on $\mathbb{R}^2$

$$\|x\|_p = (|x_1|^p + |x_2|^p)^{\frac{1}{p}}, \quad x = (x_1, x_2)^T \in \mathbb{R}^2, \quad p \in [1, \infty) \cup \{+\infty\} \quad (\text{norm on } \mathbb{R}^2)$$

$$d_p(x, y) = \|x - y\|_p \quad (\text{distance function on } \mathbb{R}^2)$$

$$\Pi = \{\pi_1, \dots, \pi_k\} \quad (\text{partition})$$

# Clusters on $\mathbb{R}^2$

$$\|x\|_p = (|x_1|^p + |x_2|^p)^{\frac{1}{p}}, \quad x = (x_1, x_2)^T \in \mathbb{R}^2, \quad p \in [1, \infty) \cup \{+\infty\} \quad (\text{norm on } \mathbb{R}^2)$$

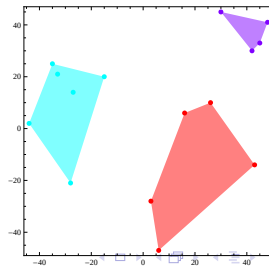
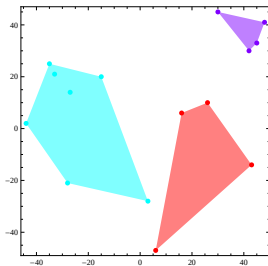
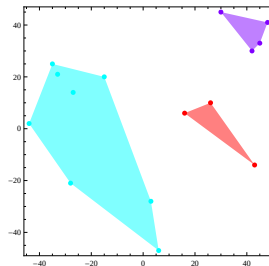
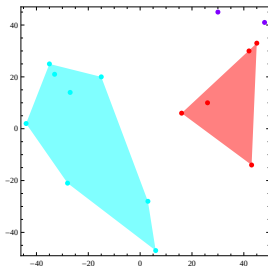
$$d_p(x, y) = \|x - y\|_p \quad (\text{distance function on } \mathbb{R}^2)$$

$$\Pi = \{\pi_1, \dots, \pi_k\} \quad (\text{partition})$$

$$c_j^p = \operatorname{argmin}_{\alpha \in \mathbb{R}^2} \sum_{x_i \in \pi_j} \|x_i - \alpha\|_p, \quad f_p(\pi_j) = \sum_{x_i \in \pi_j} \|x_i - c_j^p\|_p,$$

$$c_j^{LS} = \operatorname{argmin}_{\alpha \in \mathbb{R}^2} \sum_{x_i \in \pi_j} \|x_i - \alpha\|_2^2 = \frac{1}{|\pi_j|} \sum_{x_i \in \pi_j} x_i, \quad f_{LS}(\pi_j) = \sum_{x_i \in \pi_j} \|x_i - c_j^{LS}\|_2^2.$$

Example:  $m = 15$ ,  $k = 3$



Example:  $m = 15$ ,  $k = 3$

| Partition | Objective function |             |                  |         |
|-----------|--------------------|-------------|------------------|---------|
|           | $l_1$ -norm        | $l_2$ -norm | $l_\infty$ -norm | $LS$    |
| $\Pi_1$   | 420                | 326.89      | 278.0            | 9666.58 |
| $\Pi_2$   | 376                | 284.42      | 241.5            | 8402.21 |
| $\Pi_3$   | 354                | 269.42      | 234.0            | 7268.14 |
| $\Pi_4$   | 308                | 248.53      | 215.5            | 5603.67 |

$$S = \{x_i = (x_1^{(i)}, \dots, x_n^{(i)})^T \in \mathbb{R}^n : i = 1, \dots, m\}$$

$\pi_j$  – cluster with  $m_j = |\pi_j| \geq 1$  elements

$$c_j = \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i \quad - \text{center of the cluster } \pi_j$$

$$a \in \mathbb{R}^n$$

$$S = \{x_i = (x_1^{(i)}, \dots, x_n^{(i)})^T \in \mathbb{R}^n : i = 1, \dots, m\}$$

$\pi_j$  – cluster with  $m_j = |\pi_j| \geq 1$  elements

$$c_j = \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i \quad - \text{center of the cluster } \pi_j$$

$$a \in \mathbb{R}^n$$

$$\sum_{x_i \in \pi_j} (x_i - a)(x_i - a)^T = \sum_{x_i \in \pi_j} (x_i - c_j)(x_i - c_j)^T + m_j(c_j - a)(c_j - a)^T,$$

$$\sum_{x_i \in \pi_j} (x_i - a)(x_i - a)^T = \sum_{x_i \in \pi_j} (x_i - c_j)(x_i - c_j)^T + m_j(c_j - a)(c_j - a)^T$$

Proof:

$$\sum_{x_i \in \pi_j} (x_i - a)(x_i - a)^T = \sum_{x_i \in \pi_j} ((x_i - c_j) + (c_j - a))((x_i - c_j) + (c_j - a))^T$$

$$\sum_{x_i \in \pi_j} (x_i - a)(x_i - a)^T = \sum_{x_i \in \pi_j} (x_i - c_j)(x_i - c_j)^T + m_j(c_j - a)(c_j - a)^T$$

Proof:

$$\begin{aligned} \sum_{x_i \in \pi_j} (x_i - a)(x_i - a)^T &= \sum_{x_i \in \pi_j} ((x_i - c_j) + (c_j - a))((x_i - c_j) + (c_j - a))^T \\ &= \sum_{x_i \in \pi_j} (x_i - c_j)(x_i - c_j)^T + m_j(c_j - a)(c_j - a)^T, \end{aligned}$$

$$\sum_{x_i \in \pi_j} (x_i - a)(x_i - a)^T = \sum_{x_i \in \pi_j} (x_i - c_j)(x_i - c_j)^T + m_j(c_j - a)(c_j - a)^T$$

Proof:

$$\begin{aligned} \sum_{x_i \in \pi_j} (x_i - a)(x_i - a)^T &= \sum_{x_i \in \pi_j} ((x_i - c_j) + (c_j - a))((x_i - c_j) + (c_j - a))^T \\ &= \sum_{x_i \in \pi_j} (x_i - c_j)(x_i - c_j)^T + m_j(c_j - a)(c_j - a)^T, \end{aligned}$$

because of

$$\sum_{x_i \in \pi_j} (x_i - c_j) = 0 \quad \text{and}$$

$$\sum_{x_i \in \pi_j} (x_i - c_j)(x_i - a)^T = \left( \sum_{x_i \in \pi_j} (x_i - c_j) \right) (x_i - a)^T = 0$$

$$\sum_{x_i \in \pi_j} (x_i - a)(x_i - a)^T = \sum_{x_i \in \pi_j} (x_i - c_j)(x_i - c_j)^T + m_j(c_j - a)(c_j - a)^T$$

$$\sum_{x_i \in \pi_j} (x_i - a)(x_i - a)^T = \sum_{x_i \in \pi_j} (x_i - c_j)(x_i - c_j)^T + m_j(c_j - a)(c_j - a)^T$$

$$a \leftrightarrow c = \frac{1}{m} \sum_{i=1}^m x_i$$

$$\sum_{x_i \in \pi_j} (x_i - a)(x_i - a)^T = \sum_{x_i \in \pi_j} (x_i - c_j)(x_i - c_j)^T + m_j(c_j - a)(c_j - a)^T$$

$$a \leftrightarrow c = \frac{1}{m} \sum_{i=1}^m x_i$$

$$\sum_{j=1}^k \sum_{x_i \in \pi_j} (x_i - c)(x_i - c)^T = \sum_{j=1}^k \sum_{x_i \in \pi_j} (x_i - c_j)(x_i - c_j)^T + \sum_{j=1}^k m_j(c_j - c)(c_j - c)^T$$

$$\sum_{x_i \in \pi_j} (x_i - a)(x_i - a)^T = \sum_{x_i \in \pi_j} (x_i - c_j)(x_i - c_j)^T + m_j(c_j - a)(c_j - a)^T$$

$$a \leftrightarrow c = \frac{1}{m} \sum_{i=1}^m x_i$$

$$\sum_{j=1}^k \sum_{x_i \in \pi_j} (x_i - c)(x_i - c)^T = \sum_{j=1}^k \sum_{x_i \in \pi_j} (x_i - c_j)(x_i - c_j)^T + \sum_{j=1}^k m_j(c_j - c)(c_j - c)^T$$

$$\sum_{i=1}^m (x_i - c)(x_i - c)^T = \sum_{j=1}^k \sum_{x_i \in \pi_j} (x_i - c_j)(x_i - c_j)^T + \sum_{j=1}^k m_j(c_j - c)(c_j - c)^T$$

# Theorem 1

## Theorem

*With previous notations the following holds:*

$$T = W(\Pi) + B(\Pi), \quad (4)$$

*where  $T, W, B \in \mathbb{R}^{n \times n}$ ,*

$$T = \sum_{i=1}^m (x_i - c)(x_i - c)^T,$$

$$W(\Pi) = \sum_{j=1}^k \sum_{x_i \in \pi_j} (x_i - c_j)(x_i - c_j)^T,$$

$$B(\Pi) = \sum_{j=1}^k m_j (c_j - c)(c_j - c)^T,$$

*are symmetric matrices, whereby  $W$  and  $B$  depend, and  $T$  does not depend on the partition  $\Pi$ .*

# LS-optimality criterion

By using:

$$\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B, \quad A, B \in \mathbb{R}^{n \times n}$$

$$\operatorname{tr}(xx^T) = x^T x = \|x\|^2, \quad x \in \mathbb{R}^n$$

# LS-optimality criterion

By using:

$$\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B, \quad A, B \in \mathbb{R}^{n \times n}$$

$$\operatorname{tr}(xx^T) = x^T x = \|x\|^2, \quad x \in \mathbb{R}^n$$

$$T = W(\Pi) + B(\Pi) \quad \Rightarrow \quad \sum_{i=1}^m \|x_i - c\|^2 = \sum_{j=1}^k \sum_{x_i \in \pi_j} \|x_i - c_j\|^2 + \sum_{j=1}^k m_j \|c_j - c\|^2$$

# LS-optimality criterion

By using:

$$\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B, \quad A, B \in \mathbb{R}^{n \times n}$$

$$\operatorname{tr}(xx^T) = x^T x = \|x\|^2, \quad x \in \mathbb{R}^n$$

$$T = W(\Pi) + B(\Pi) \Rightarrow \sum_{i=1}^m \|x_i - c\|^2 = \sum_{j=1}^k \sum_{x_i \in \pi_j} \|x_i - c_j\|^2 + \sum_{j=1}^k m_j \|c_j - c\|^2$$

## Remark

Since

$$\operatorname{tr} W(\Pi) = \operatorname{tr} \sum_{j=1}^k \sum_{x_i \in \pi_j} (x_i - c_j)(x_i - c_j)^T = \sum_{j=1}^k \sum_{x_i \in \pi_j} \|x_i - c_j\|^2 =: F(\Pi),$$

LS-optimality criterion can be written:

$$F(\Pi^{(o)}) = \min_{\Pi \in \Psi(S, k)} F(\Pi) = \min_{\Pi \in \Psi(S, k)} \operatorname{tr} W(\Pi).$$

$$\sum_{i=1}^m \|x_i - c\|^2 = \sum_{j=1}^k \sum_{x_i \in \pi_j} \|x_i - c_j\|^2 + \sum_{j=1}^k m_j \|c_j - c\|^2$$

## Remark

If  $\varphi, \psi \in C^2(\mathbb{R})$ , are such functions that  $\varphi(x) + \psi(x) = \text{const}$ , then it holds

$$x_0 \in \mathbb{R}, \quad \varphi'(x_0) = 0 \quad \& \quad \varphi''(x_0) > 0 \quad \Rightarrow \quad \psi'(x_0) = 0 \quad \& \quad \psi''(x_0) < 0$$

$$\sum_{i=1}^m \|x_i - c\|^2 = \sum_{j=1}^k \sum_{x_i \in \pi_j} \|x_i - c_j\|^2 + \sum_{j=1}^k m_j \|c_j - c\|^2$$

## Remark

If  $\varphi, \psi \in C^2(\mathbb{R})$ , are such functions that  $\varphi(x) + \psi(x) = \text{const}$ , then it holds

$$x_0 \in \mathbb{R}, \quad \varphi'(x_0) = 0 \quad \& \quad \varphi''(x_0) > 0 \quad \Rightarrow \quad \psi'(x_0) = 0 \quad \& \quad \psi''(x_0) < 0$$

$$\min_{x \in \mathbb{R}} \varphi(x) = \varphi(x_0) \quad \Rightarrow \quad \max_{x \in \mathbb{R}} \psi(x) = \psi(x_0),$$

$$\sum_{i=1}^m \|x_i - c\|^2 = \sum_{j=1}^k \sum_{x_i \in \pi_j} \|x_i - c_j\|^2 + \sum_{j=1}^k m_j \|c_j - c\|^2$$

## Remark

If  $\varphi, \psi \in C^2(\mathbb{R})$ , are such functions that  $\varphi(x) + \psi(x) = \text{const}$ , then it holds

$$x_0 \in \mathbb{R}, \quad \varphi'(x_0) = 0 \quad \& \quad \varphi''(x_0) > 0 \quad \Rightarrow \quad \psi'(x_0) = 0 \quad \& \quad \psi''(x_0) < 0$$

$$\min_{x \in \mathbb{R}} \varphi(x) = \varphi(x_0) \quad \Rightarrow \quad \max_{x \in \mathbb{R}} \psi(x) = \psi(x_0),$$

## Theorem

With previos notations the following holds (Späth, 1983, Zha, 2002):

$$\operatorname{argmin}_{\Pi \in \Psi(S,k)} \sum_{j=1}^k \sum_{x_i \in \pi_j} \|x_i - c_j\|^2 = \operatorname{argmax}_{\Pi \in \Psi(S,k)} \sum_{j=1}^k m_j \|c_j - c\|^2 =: \Pi^{(o)};$$

$$F(\Pi^{(o)}) = \min_{\Pi \in \Psi(S,k)} \sum_{j=1}^k \sum_{x_i \in \pi_j} \|x_i - c_j\|^2 = \sum_{i=1}^m \|x_i - c\|^2 - \max_{\Pi \in \Psi(S,k)} \sum_{j=1}^k m_j \|c_j - c\|^2$$

# Transformation of the data

$$S = \{x_i = (x_1^{(i)}, \dots, x_n^{(i)})^T \in \mathbb{R}^n : i = 1, \dots, m\}$$

$$\Pi = \{\pi_1, \dots, \pi_k\}$$

# Transformation of the data

$$S = \{x_i = (x_1^{(i)}, \dots, x_n^{(i)})^T \in \mathbb{R}^n : i = 1, \dots, m\}$$

$$\Pi = \{\pi_1, \dots, \pi_k\}$$

$$\pi_j \text{ cluster with } m_j = |\pi_j| \geq 1 \text{ and center } c_j = \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i$$

# Transformation of the data

$$S = \{x_i = (x_1^{(i)}, \dots, x_n^{(i)})^T \in \mathbb{R}^n : i = 1, \dots, m\}$$

$$\Pi = \{\pi_1, \dots, \pi_k\}$$

$$\pi_j \text{ cluster with } m_j = |\pi_j| \geq 1 \text{ and center } c_j = \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i$$

$$\mathcal{A} : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

$$S' = \{y_i \in \mathbb{R}^n : y_i = \mathcal{A}x_i, x_i \in S\}$$

$$\Pi' = \{\pi'_1, \dots, \pi'_k\},$$

# Transformation of the data

$$S = \{x_i = (x_1^{(i)}, \dots, x_n^{(i)})^T \in \mathbb{R}^n : i = 1, \dots, m\}$$

$$\Pi = \{\pi_1, \dots, \pi_k\}$$

$$\pi_j \text{ cluster with } m_j = |\pi_j| \geq 1 \text{ and center } c_j = \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i$$

$$\mathcal{A} : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

$$S' = \{y_i \in \mathbb{R}^n : y_i = \mathcal{A}x_i, x_i \in S\}$$

$$\Pi' = \{\pi'_1, \dots, \pi'_k\},$$

$$\pi'_j = \{y_i \in S' : y_i = \mathcal{A}x_i, x_i \in \pi_j\} \text{ with center } c'_j = \frac{1}{m_j} \sum_{y_i \in \pi'_j} y_i$$

# Transformation of the data

$$S = \{x_i = (x_1^{(i)}, \dots, x_n^{(i)})^T \in \mathbb{R}^n : i = 1, \dots, m\}$$

$$\Pi = \{\pi_1, \dots, \pi_k\}$$

$$\pi_j \text{ cluster with } m_j = |\pi_j| \geq 1 \text{ and center } c_j = \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i$$

$$\mathcal{A} : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

$$S' = \{y_i \in \mathbb{R}^n : y_i = \mathcal{A}x_i, x_i \in S\}$$

$$\Pi' = \{\pi'_1, \dots, \pi'_k\},$$

$$\pi'_j = \{y_i \in S' : y_i = \mathcal{A}x_i, x_i \in \pi_j\} \text{ with center } c'_j = \frac{1}{m_j} \sum_{y_i \in \pi'_j} y_i$$

$$\min_{\Pi' \in \Psi(S', k)} \sum_{j=1}^k \sum_{y_i \in \pi'_j} \|y_i - c'_j\|^2 = \min_{\Pi \in \Psi(S, k)} \sum_{j=1}^k \sum_{x_i \in \pi_j} \|x_i - c_j\|^2 \quad ???$$

## Theorem

*LS-optimality criterion is rotation and translation invariant. More precisely, let  $Q \in \mathbb{R}^{n \times n}$  be an orthogonal matrix and  $b \in \mathbb{R}^n$  some vector and*

$$S' = \{y_i \in \mathbb{R}^n : y_i = Qx_i + b, i = 1, \dots, m\},$$

$$\Pi' = \{\pi'_1, \dots, \pi'_k\}, \quad \pi'_j = \{y_i \in S' : x_i \in \pi_j\}, \quad c'_j = \frac{1}{m_j} \sum_{y_i \in \pi'_j} y_i.$$

*Then  $c'_j = Qc_j + b$  and*

$$\min_{\Pi' \in \Psi(S', k)} \sum_{j=1}^k \sum_{y_i \in \pi'_j} \|y_i - c'_j\|^2 = \min_{\Pi \in \Psi(S, k)} \sum_{j=1}^k \sum_{x_i \in \pi_j} \|x_i - c_j\|^2.$$

## Theorem

*LS-optimality criterion is rotation and translation invariant. More precisely, let  $Q \in \mathbb{R}^{n \times n}$  be an orthogonal matrix and  $b \in \mathbb{R}^n$  some vector and*

$$S' = \{y_i \in \mathbb{R}^n : y_i = Qx_i + b, i = 1, \dots, m\},$$

$$\Pi' = \{\pi'_1, \dots, \pi'_k\}, \quad \pi'_j = \{y_i \in S' : x_i \in \pi_j\}, \quad c'_j = \frac{1}{m_j} \sum_{y_i \in \pi'_j} y_i.$$

*Then  $c'_j = Qc_j + b$  and*

$$\min_{\Pi' \in \Psi(S', k)} \sum_{j=1}^k \sum_{y_i \in \pi'_j} \|y_i - c'_j\|^2 = \min_{\Pi \in \Psi(S, k)} \sum_{j=1}^k \sum_{x_i \in \pi_j} \|x_i - c_j\|^2.$$

Proof: 
$$c'_j = \frac{1}{m_j} \sum_{y_i \in \pi'_j} y_i = \frac{1}{m_j} \sum_{x_i \in \pi_j} (Qx_i + b) = Q \left( \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i \right) + b = Qc_j + b.$$

# Homothety – scaling of the data

## Theorem

*LS-optimality criterion is scale invariant  $y_i = Hx_i$ ,  $H = \text{diag}(h_1, \dots, h_n)$  if and only if  $h_i = \pm 1$ , for every  $i = 1, \dots, n$ , i.e. there holds*

$$\min_{\Pi \in \Psi(S, k)} \sum_{j=1}^k \sum_{y_i \in \pi'_j} \|y_i - c'_j\|^2 = \min_{\Pi \in \Psi(S, k)} \sum_{j=1}^k \sum_{x_i \in \pi_j} \|x_i - c_j\|^2,$$

where  $c_j = \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i$ ,  $c'_j = \frac{1}{m_j} \sum_{y_i \in \pi'_j} y_i$ .

# Homothety – scaling of the data

## Theorem

*LS-optimality criterion is scale invariant  $y_i = Hx_i$ ,  $H = \text{diag}(h_1, \dots, h_n)$  if and only if  $h_i = \pm 1$ , for every  $i = 1, \dots, n$ , i.e. there holds*

$$\min_{\Pi \in \Psi(S, k)} \sum_{j=1}^k \sum_{y_i \in \pi'_j} \|y_i - c'_j\|^2 = \min_{\Pi \in \Psi(S, k)} \sum_{j=1}^k \sum_{x_i \in \pi_j} \|x_i - c_j\|^2,$$

where  $c_j = \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i$ ,  $c'_j = \frac{1}{m_j} \sum_{y_i \in \pi'_j} y_i$ .

Proof:  $c'_j = \frac{1}{m_j} \sum_{y_i \in \pi'_j} y_i = \frac{1}{m_j} \sum_{x_i \in \pi_j} Hx_i = H \left( \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i \right) = Hc_j$  and

# Homothety – scaling of the data

## Theorem

*LS-optimality criterion is scale invariant  $y_i = Hx_i$ ,  $H = \text{diag}(h_1, \dots, h_n)$  if and only if  $h_i = \pm 1$ , for every  $i = 1, \dots, n$ , i.e. there holds*

$$\min_{\Pi \in \Psi(S, k)} \sum_{j=1}^k \sum_{y_i \in \pi'_j} \|y_i - c'_j\|^2 = \min_{\Pi \in \Psi(S, k)} \sum_{j=1}^k \sum_{x_i \in \pi_j} \|x_i - c_j\|^2,$$

where  $c_j = \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i$ ,  $c'_j = \frac{1}{m_j} \sum_{y_i \in \pi'_j} y_i$ .

Proof:  $c'_j = \frac{1}{m_j} \sum_{y_i \in \pi'_j} y_i = \frac{1}{m_j} \sum_{x_i \in \pi_j} Hx_i = H \left( \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i \right) = Hc_j$  and

$$\sum_{y_i \in \pi'_j} \|y_i - c'_j\|^2 = \sum_{x_i \in \pi_j} \|H(x_i - c_j)\|^2 = \sum_{x_i \in \pi_j} (x_i - c_j)^T H^T H (x_i - c_j)$$

# Homothety – scaling of the data

## Theorem

*LS-optimality criterion is scale invariant  $y_i = Hx_i$ ,  $H = \text{diag}(h_1, \dots, h_n)$  if and only if  $h_i = \pm 1$ , for every  $i = 1, \dots, n$ , i.e. there holds*

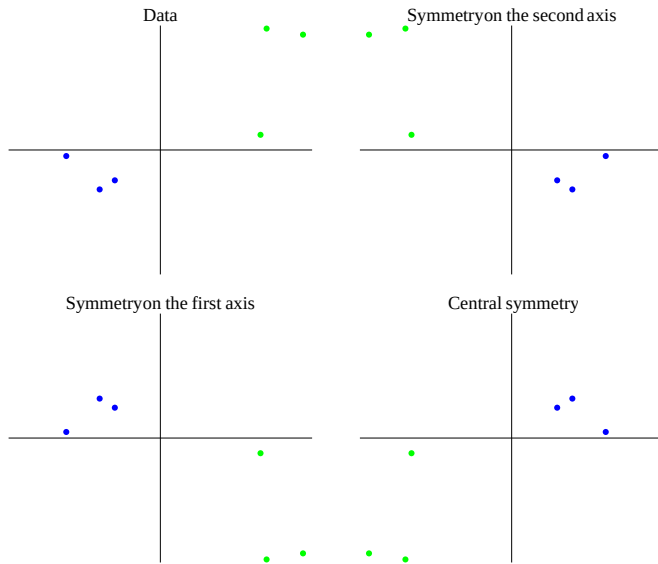
$$\min_{\Pi \in \Psi(S, k)} \sum_{j=1}^k \sum_{y_i \in \pi'_j} \|y_i - c'_j\|^2 = \min_{\Pi \in \Psi(S, k)} \sum_{j=1}^k \sum_{x_i \in \pi_j} \|x_i - c_j\|^2,$$

where  $c_j = \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i$ ,  $c'_j = \frac{1}{m_j} \sum_{y_i \in \pi'_j} y_i$ .

Proof:  $c'_j = \frac{1}{m_j} \sum_{y_i \in \pi'_j} y_i = \frac{1}{m_j} \sum_{x_i \in \pi_j} Hx_i = H \left( \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i \right) = Hc_j$  and

$$\begin{aligned} \sum_{y_i \in \pi'_j} \|y_i - c'_j\|^2 &= \sum_{x_i \in \pi_j} \|H(x_i - c_j)\|^2 = \sum_{x_i \in \pi_j} (x_i - c_j)^T H^T H (x_i - c_j) \\ &= \sum_{x_i \in \pi_j} (x_i - c_j)^T (x_i - c_j) \Leftrightarrow H^T H = I \Leftrightarrow h_i^2 = 1. \end{aligned}$$

# Homothety – scaling of the data



# The methods

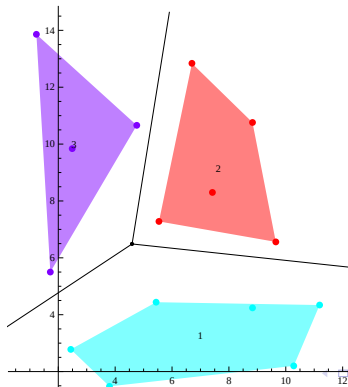
- $k$ -means clustering
- Substitution method
- Spectral Relaxation
- Agglomerative nesting

# $k$ -means optimal partition (Minimaldistanzpartition)

## Definition

A partition  $\hat{\Pi} = \{\hat{\pi}_1, \dots, \hat{\pi}_k\}$  is called a  $k$ -means optimal partition if

$$x_i \in \hat{\pi}_j \Leftrightarrow \|x_i - c_j\| = \min_{s=1,k} \|x_i - c_s\|, \quad c_j = \frac{1}{|\hat{\pi}_j|} \sum_{x_l \in \hat{\pi}_j} x_l.$$



# Separating Theorem

## Theorem

Let  $\hat{\Pi} = \{\hat{\pi}_1, \dots, \hat{\pi}_k\}$  be an  $k$ -means optimal partition. Then

- $\text{conv}(\hat{\pi}_i) \cap \text{conv}(\hat{\pi}_j) = \emptyset, \quad \forall 1 \leq i \neq j \leq k,$
- the hyperplane separating two different clusters  $\hat{\pi}_p, \hat{\pi}_q$  with centers  $c_p, c_q$  is given by

$$(c_q - c_p)^T \left( x - \frac{1}{2}(c_q + c_p) \right) = 0.$$

# Separating Theorem

## Theorem

Let  $\hat{\Pi} = \{\hat{\pi}_1, \dots, \hat{\pi}_k\}$  be an  $k$ -means optimal partition. Then

- $\text{conv}(\hat{\pi}_i) \cap \text{conv}(\hat{\pi}_j) = \emptyset, \quad \forall 1 \leq i \neq j \leq k,$
- the hyperplane separating two different clusters  $\hat{\pi}_p, \hat{\pi}_q$  with centers  $c_p, c_q$  is given by

$$(c_q - c_p)^T \left( x - \frac{1}{2}(c_q + c_p) \right) = 0.$$

The set of all points  $x \in \mathbb{R}^n$  equally distant from the point  $c_p \in \mathbb{R}^n$  and the point  $c_q \in \mathbb{R}^n$  is given by

$$H = \{x \in \mathbb{R}^n : \|x - c_q\| = \|x - c_p\|\}.$$

# Separating Theorem

## Theorem

Let  $\hat{\Pi} = \{\hat{\pi}_1, \dots, \hat{\pi}_k\}$  be an  $k$ -means optimal partition. Then

- $\text{conv}(\hat{\pi}_i) \cap \text{conv}(\hat{\pi}_j) = \emptyset, \quad \forall 1 \leq i \neq j \leq k,$
- the hyperplane separating two different clusters  $\hat{\pi}_p, \hat{\pi}_q$  with centers  $c_p, c_q$  is given by

$$(c_q - c_p)^T \left( x - \frac{1}{2}(c_q + c_p) \right) = 0.$$

The set of all points  $x \in \mathbb{R}^n$  equally distant from the point  $c_p \in \mathbb{R}^n$  and the point  $c_q \in \mathbb{R}^n$  is given by

$$H = \{x \in \mathbb{R}^n : \|x - c_q\| = \|x - c_p\|\}.$$

$$\|x\|^2 + \|c_q\|^2 - 2c_q^T x = \|x\|^2 + \|c_p\|^2 - 2c_p^T x,$$

# Separating Theorem

## Theorem

Let  $\hat{\Pi} = \{\hat{\pi}_1, \dots, \hat{\pi}_k\}$  be an  $k$ -means optimal partition. Then

- $\text{conv}(\hat{\pi}_i) \cap \text{conv}(\hat{\pi}_j) = \emptyset, \quad \forall 1 \leq i \neq j \leq k,$
- the hyperplane separating two different clusters  $\hat{\pi}_p, \hat{\pi}_q$  with centers  $c_p, c_q$  is given by

$$(c_q - c_p)^T \left( x - \frac{1}{2}(c_q + c_p) \right) = 0.$$

The set of all points  $x \in \mathbb{R}^n$  equally distant from the point  $c_p \in \mathbb{R}^n$  and the point  $c_q \in \mathbb{R}^n$  is given by

$$H = \{x \in \mathbb{R}^n : \|x - c_q\| = \|x - c_p\|\}.$$

$$\|x\|^2 + \|c_q\|^2 - 2c_q^T x = \|x\|^2 + \|c_p\|^2 - 2c_p^T x,$$

$$\|c_q\|^2 - \|c_p\|^2 = 2(c_q - c_p)^T x,$$

# Separating Theorem

## Theorem

Let  $\hat{\Pi} = \{\hat{\pi}_1, \dots, \hat{\pi}_k\}$  be an  $k$ -means optimal partition. Then

- $\text{conv}(\hat{\pi}_i) \cap \text{conv}(\hat{\pi}_j) = \emptyset, \quad \forall 1 \leq i \neq j \leq k,$
- the hyperplane separating two different clusters  $\hat{\pi}_p, \hat{\pi}_q$  with centers  $c_p, c_q$  is given by

$$(c_q - c_p)^T \left( x - \frac{1}{2}(c_q + c_p) \right) = 0.$$

The set of all points  $x \in \mathbb{R}^n$  equally distant from the point  $c_p \in \mathbb{R}^n$  and the point  $c_q \in \mathbb{R}^n$  is given by

$$H = \{x \in \mathbb{R}^n : \|x - c_q\| = \|x - c_p\|\}.$$

$$\|x\|^2 + \|c_q\|^2 - 2c_q^T x = \|x\|^2 + \|c_p\|^2 - 2c_p^T x,$$

$$\|c_q\|^2 - \|c_p\|^2 = 2(c_q - c_p)^T x,$$

$$(c_q - c_p)^T (c_q + c_p) = 2(c_q - c_p)^T x.$$

# $k$ -means algorithm (Minimaldistanzverfahren)

# $k$ -means algorithm (Minimaldistanzverfahren)

**Step 1:** Input  $1 \leq k \leq m$ ,  $it_{max}$ ,  $(x_i, i = 1, \dots, m)$ , set  $\mu = 0$ , choose the initial partition  $\Pi^{(\mu)} = \{\pi_1^{(\mu)}, \dots, \pi_k^{(\mu)}\}$  and calculate  $m_j^{(\mu)} = |\pi_j^{(\mu)}|$  for  $j = 1, \dots, k$  and

$$c_j^{(\mu)} = \frac{1}{m_j^{(\mu)}} \sum_{x_i \in \pi_j^{(\mu)}} x_i, \quad F_\mu = F(\Pi^{(\mu)}) = \sum_{j=1}^k \sum_{x_i \in \pi_j^{(\mu)}} \|x_i - c_j^{(\mu)}\|^2;$$

# k-means algorithm (Minimaldistanzverfahren)

**Step 1:** Input  $1 \leq k \leq m$ ,  $it_{max}$ ,  $(x_i, i = 1, \dots, m)$ , set  $\mu = 0$ , choose the initial partition  $\Pi^{(\mu)} = \{\pi_1^{(\mu)}, \dots, \pi_k^{(\mu)}\}$  and calculate  $m_j^{(\mu)} = |\pi_j^{(\mu)}|$  for  $j = 1, \dots, k$  and

$$c_j^{(\mu)} = \frac{1}{m_j^{(\mu)}} \sum_{x_i \in \pi_j^{(\mu)}} x_i, \quad F_\mu = F(\Pi^{(\mu)}) = \sum_{j=1}^k \sum_{x_i \in \pi_j^{(\mu)}} \|x_i - c_j^{(\mu)}\|^2;$$

**Step 2:** Define a new partition  $\Pi^{(\mu+1)} = \{\pi_1^{(\mu+1)}, \dots, \pi_k^{(\mu+1)}\}$

$$\pi_1^{(\mu+1)} = \{x_i \in S : \|x_i - c_1^{(\mu)}\| \leq \|x_i - c_l^{(\mu)}\|, l = 1, \dots, k\}$$

$$\pi_j^{(\mu+1)} = \{x_i \in S \setminus \bigcup_{s=1}^{j-1} \pi_s^{(\mu+1)} : \|x_i - c_j^{(\mu)}\| \leq \|x_i - c_l^{(\mu)}\|, \forall l = 1, \dots, k\}, \quad j = 2, \dots, k;$$

# k-means algorithm (Minimaldistanzverfahren)

**Step 1:** Input  $1 \leq k \leq m$ ,  $it_{max}$ ,  $(x_i, i = 1, \dots, m)$ , set  $\mu = 0$ , choose the initial partition  $\Pi^{(\mu)} = \{\pi_1^{(\mu)}, \dots, \pi_k^{(\mu)}\}$  and calculate  $m_j^{(\mu)} = |\pi_j^{(\mu)}|$  for  $j = 1, \dots, k$  and

$$c_j^{(\mu)} = \frac{1}{m_j^{(\mu)}} \sum_{x_i \in \pi_j^{(\mu)}} x_i, \quad F_\mu = F(\Pi^{(\mu)}) = \sum_{j=1}^k \sum_{x_i \in \pi_j^{(\mu)}} \|x_i - c_j^{(\mu)}\|^2;$$

**Step 2:** Define a new partition  $\Pi^{(\mu+1)} = \{\pi_1^{(\mu+1)}, \dots, \pi_k^{(\mu+1)}\}$

$$\pi_1^{(\mu+1)} = \{x_i \in S : \|x_i - c_1^{(\mu)}\| \leq \|x_i - c_l^{(\mu)}\|, l = 1, \dots, k\}$$

$$\pi_j^{(\mu+1)} = \{x_i \in S \setminus \bigcup_{s=1}^{j-1} \pi_s^{(\mu+1)} : \|x_i - c_j^{(\mu)}\| \leq \|x_i - c_l^{(\mu)}\|, \forall l = 1, \dots, k\}, \quad j = 2, \dots, k;$$

**Step 3:** Calculate new centroids and a new value of the objective function;

$$c_j^{(\mu+1)} = \frac{1}{m_j^{(\mu+1)}} \sum_{x_i \in \pi_j^{(\mu+1)}} x_i, \quad F_{\mu+1} = F(\Pi^{(\mu+1)}) = \sum_{j=1}^k \sum_{x_i \in \pi_j^{(\mu+1)}} \|x_i - c_j^{(\mu+1)}\|^2;$$

# k-means algorithm (Minimaldistanzverfahren)

**Step 1:** Input  $1 \leq k \leq m$ ,  $it_{max}$ ,  $(x_i, i = 1, \dots, m)$ , set  $\mu = 0$ , choose the initial partition  $\Pi^{(\mu)} = \{\pi_1^{(\mu)}, \dots, \pi_k^{(\mu)}\}$  and calculate  $m_j^{(\mu)} = |\pi_j^{(\mu)}|$  for  $j = 1, \dots, k$  and

$$c_j^{(\mu)} = \frac{1}{m_j^{(\mu)}} \sum_{x_i \in \pi_j^{(\mu)}} x_i, \quad F_\mu = F(\Pi^{(\mu)}) = \sum_{j=1}^k \sum_{x_i \in \pi_j^{(\mu)}} \|x_i - c_j^{(\mu)}\|^2;$$

**Step 2:** Define a new partition  $\Pi^{(\mu+1)} = \{\pi_1^{(\mu+1)}, \dots, \pi_k^{(\mu+1)}\}$

$$\pi_1^{(\mu+1)} = \{x_i \in S : \|x_i - c_1^{(\mu)}\| \leq \|x_i - c_l^{(\mu)}\|, l = 1, \dots, k\}$$

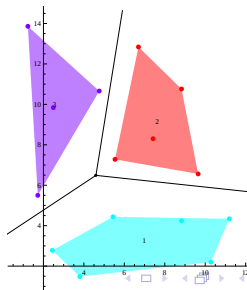
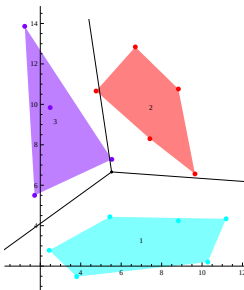
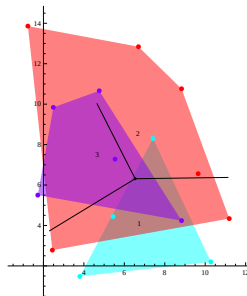
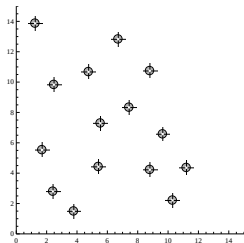
$$\pi_j^{(\mu+1)} = \{x_i \in S \setminus \bigcup_{s=1}^{j-1} \pi_s^{(\mu+1)} : \|x_i - c_j^{(\mu)}\| \leq \|x_i - c_l^{(\mu)}\|, \forall l = 1, \dots, k\}, \quad j = 2, \dots, k;$$

**Step 3:** Calculate new centroids and a new value of the objective function;

$$c_j^{(\mu+1)} = \frac{1}{m_j^{(\mu+1)}} \sum_{x_i \in \pi_j^{(\mu+1)}} x_i, \quad F_{\mu+1} = F(\Pi^{(\mu+1)}) = \sum_{j=1}^k \sum_{x_i \in \pi_j^{(\mu+1)}} \|x_i - c_j^{(\mu+1)}\|^2;$$

**Step 4:** If  $F_{\mu+1} < F_\mu$  and if  $\mu < it_{max}$ , set  $\mu := \mu + 1$  and GoTo Step 2; Else STOP

# k-means algorithm (Minimaldistanzverfahren)



# $k$ -means algorithm – properties

## Theorem

*The  $k$ -means algorithm does not increase the value of the objective function.*

## Theorem

*The  $k$ -means algorithm does not increase the value of the objective function.*

Proof:

$$F(\Pi^{(\mu)}) = \sum_{j=1}^k \sum_{x_i \in \pi_j^{(\mu)}} \|x_i - c_j^{(\mu)}\|^2$$

## Theorem

*The  $k$ -means algorithm does not increase the value of the objective function.*

Proof:

$$\begin{aligned} F(\Pi^{(\mu)}) &= \sum_{j=1}^k \sum_{x_i \in \pi_j^{(\mu)}} \|x_i - c_j^{(\mu)}\|^2 \\ &\geq \sum_{j=1}^k \sum_{x_i \in \pi_j^{(\mu)}} \min_{s=1,k} \|x_i - c_s^{(\mu)}\|^2 = \sum_{j=1}^k \sum_{x_i \in \pi_j^{(\mu+1)}} \|x_i - c_j^{(\mu)}\|^2 \end{aligned}$$

## Theorem

*The k-means algorithm does not increase the value of the objective function.*

Proof:

$$\begin{aligned} F(\Pi^{(\mu)}) &= \sum_{j=1}^k \sum_{x_i \in \pi_j^{(\mu)}} \|x_i - c_j^{(\mu)}\|^2 \\ &\geq \sum_{j=1}^k \sum_{x_i \in \pi_j^{(\mu)}} \min_{s=1,k} \|x_i - c_s^{(\mu)}\|^2 = \sum_{j=1}^k \sum_{x_i \in \pi_j^{(\mu+1)}} \|x_i - c_j^{(\mu)}\|^2 \\ &\geq \sum_{j=1}^k \sum_{x_i \in \pi_j^{(\mu+1)}} \|x_i - c_j^{(\mu+1)}\|^2 = F(\Pi^{(\mu+1)}) \end{aligned}$$

# $k$ -means algorithm – properties

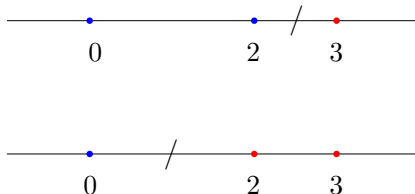
- In some cases, the  $k$ -means algorithm does not give an optimal partition:

$S = \{0, 2, 3\}$  – the set

$\Pi^0 = \{\{0, 2\}, \{3\}\}$  – initial partition

$\hat{\Pi}^k = \{\{0, 2\}, \{3\}\}$  –  $k$ -means optimal partition ( $F(\hat{\Pi}^k) = 2$ )

$\hat{\Pi} = \{\{0\}, \{2, 3\}\}$  – optimal partition ( $F(\hat{\Pi}) = 0.5$ )



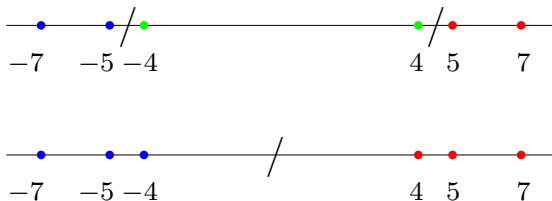
# $k$ -means algorithm – properties

- The  $k$ -means algorithm can reduce the number of clusters:

$S = \{-7, -5, -4, 4, 5, 7\}$  – the set

$\Pi^0 = \{\{-7, -5\}, \{-4, 4\}, \{5, 7\}\}$  – initial partition ( $F(\Pi^0) = 36$ )

$\hat{\Pi} = \{\{-7, -5, -4\}, \emptyset, \{4, 5, 7\}\}$  –  $k$ -means optimal partition ( $F(\hat{\Pi}) = 9.33$ )



# Replacement of cluster elements

Let  $c = \frac{1}{m} \sum_{i=1}^m x_i$  and  $a \in \mathbb{R}^n$ . Then

$$\sum_{i=1}^m (x_i - a)(x_i - a)^T = \sum_{i=1}^m ((x_i - c) + (c - a))((x_i - c) + (c - a))^T$$

# Replacement of cluster elements

Let  $c = \frac{1}{m} \sum_{i=1}^m x_i$  and  $a \in \mathbb{R}^n$ . Then

$$\sum_{i=1}^m (x_i - a)(x_i - a)^T = \sum_{i=1}^m ((x_i - c) + (c - a))((x_i - c) + (c - a))^T$$

# Replacement of cluster elements

Let  $c = \frac{1}{m} \sum_{i=1}^m x_i$  and  $a \in \mathbb{R}^n$ . Then

$$\begin{aligned} \sum_{i=1}^m (x_i - a)(x_i - a)^T &= \sum_{i=1}^m ((x_i - c) + (c - a))((x_i - c) + (c - a))^T \\ &= \sum_{i=1}^m (x_i - c)(x_i - c)^T + m(c - a)(c - a)^T, \end{aligned}$$

# Replacement of cluster elements

Let  $c = \frac{1}{m} \sum_{i=1}^m x_i$  and  $a \in \mathbb{R}^n$ . Then

$$\begin{aligned} \sum_{i=1}^m (x_i - a)(x_i - a)^T &= \sum_{i=1}^m ((x_i - c) + (c - a))((x_i - c) + (c - a))^T \\ &= \sum_{i=1}^m (x_i - c)(x_i - c)^T + m(c - a)(c - a)^T, \end{aligned}$$

because of  $\sum_{i=1}^m (x_i - c) = 0$  and  $\sum_{i=1}^m (x_i - c)(c - a)^T = \left( \sum_{i=1}^m (x_i - c) \right) (c - a)^T = 0$

# Replacement of cluster elements

Let  $c = \frac{1}{m} \sum_{i=1}^m x_i$  and  $a \in \mathbb{R}^n$ . Then

$$\begin{aligned} \sum_{i=1}^m (x_i - a)(x_i - a)^T &= \sum_{i=1}^m ((x_i - c) + (c - a))((x_i - c) + (c - a))^T \\ &= \sum_{i=1}^m (x_i - c)(x_i - c)^T + m(c - a)(c - a)^T, \end{aligned}$$

because of  $\sum_{i=1}^m (x_i - c) = 0$  and  $\sum_{i=1}^m (x_i - c)(c - a)^T = \left( \sum_{i=1}^m (x_i - c) \right) (c - a)^T = 0$

Therefore,

$$\sum_{i=1}^m \|x_i - a\|^2 = \sum_{i=1}^m \|x_i - c\|^2 + m\|a - c\|^2, \quad \forall a \in \mathbb{R}^n. \quad (5)$$

$$\sum_{i=1}^m \|x_i - a\|^2 = \sum_{i=1}^m \|x_i - c\|^2 + m\|a - c\|^2, \quad \forall a \in \mathbb{R}^n.$$

## Lemma

Let  $A = \{x_1, \dots, x_p\}$  and  $B = \{y_1, \dots, y_q\}$  be two different sets in  $\mathbb{R}^n$ . Then

$$f(A \cup B) = f(A) + f(B) + p\|c(A) - c\|^2 + q\|c(B) - c\|^2,$$

$$c(A) = \frac{1}{p} \sum_{i=1}^p x_i, \quad c(B) = \frac{1}{q} \sum_{i=1}^q y_i, \quad c = c(A \cup B) = \frac{p}{p+q} c(A) + \frac{q}{p+q} c(B).$$

Proof:

$$\sum_{i=1}^m \|x_i - a\|^2 = \sum_{i=1}^m \|x_i - c\|^2 + m\|a - c\|^2, \quad \forall a \in \mathbb{R}^n.$$

## Lemma

Let  $A = \{x_1, \dots, x_p\}$  and  $B = \{y_1, \dots, y_q\}$  be two different sets in  $\mathbb{R}^n$ . Then

$$f(A \cup B) = f(A) + f(B) + p\|c(A) - c\|^2 + q\|c(B) - c\|^2,$$

$$c(A) = \frac{1}{p} \sum_{i=1}^p x_i, \quad c(B) = \frac{1}{q} \sum_{i=1}^q y_i, \quad c = c(A \cup B) = \frac{p}{p+q} c(A) + \frac{q}{p+q} c(B).$$

Proof:

$$c = c(A \cup B) = \frac{1}{p+q} \left( \sum_{i=1}^p x_i + \sum_{i=1}^q y_i \right) = \frac{p}{p+q} \frac{1}{p} \sum_{i=1}^p x_i + \frac{q}{p+q} \frac{1}{q} \sum_{i=1}^q y_i.$$

$$\sum_{i=1}^m \|x_i - a\|^2 = \sum_{i=1}^m \|x_i - c\|^2 + m\|a - c\|^2, \quad \forall a \in \mathbb{R}^n.$$

## Lemma

Let  $A = \{x_1, \dots, x_p\}$  and  $B = \{y_1, \dots, y_q\}$  be two different sets in  $\mathbb{R}^n$ . Then

$$f(A \cup B) = f(A) + f(B) + p\|c(A) - c\|^2 + q\|c(B) - c\|^2,$$

$$c(A) = \frac{1}{p} \sum_{i=1}^p x_i, \quad c(B) = \frac{1}{q} \sum_{i=1}^q y_i, \quad c = c(A \cup B) = \frac{p}{p+q} c(A) + \frac{q}{p+q} c(B).$$

Proof:

$$c = c(A \cup B) = \frac{1}{p+q} \left( \sum_{i=1}^p x_i + \sum_{i=1}^q y_i \right) = \frac{p}{p+q} \frac{1}{p} \sum_{i=1}^p x_i + \frac{q}{p+q} \frac{1}{q} \sum_{i=1}^q y_i.$$

$$f(A \cup B) = \sum_{i=1}^p \|x_i - c\|^2 + \sum_{i=1}^q \|y_i - c\|^2$$

$$\sum_{i=1}^m \|x_i - a\|^2 = \sum_{i=1}^m \|x_i - c\|^2 + m\|a - c\|^2, \quad \forall a \in \mathbb{R}^n.$$

## Lemma

Let  $A = \{x_1, \dots, x_p\}$  and  $B = \{y_1, \dots, y_q\}$  be two different sets in  $\mathbb{R}^n$ . Then

$$f(A \cup B) = f(A) + f(B) + p\|c(A) - c\|^2 + q\|c(B) - c\|^2,$$

$$c(A) = \frac{1}{p} \sum_{i=1}^p x_i, \quad c(B) = \frac{1}{q} \sum_{i=1}^q y_i, \quad c = c(A \cup B) = \frac{p}{p+q} c(A) + \frac{q}{p+q} c(B).$$

Proof:

$$c = c(A \cup B) = \frac{1}{p+q} \left( \sum_{i=1}^p x_i + \sum_{i=1}^q y_i \right) = \frac{p}{p+q} \frac{1}{p} \sum_{i=1}^p x_i + \frac{q}{p+q} \frac{1}{q} \sum_{i=1}^q y_i.$$

$$\begin{aligned} f(A \cup B) &= \sum_{i=1}^p \|x_i - c\|^2 + \sum_{i=1}^q \|y_i - c\|^2 \\ &= \sum_{i=1}^p \|x_i - c(A)\|^2 + p\|c - c(A)\|^2 + \sum_{i=1}^q \|y_i - c(B)\|^2 + q\|c - c(B)\|^2. \end{aligned}$$

## Lemma

Let  $A = \{x_1, \dots, x_p\}$  and  $B = \{y_1, \dots, y_q\}$  be two different sets in  $\mathbb{R}^n$ ,  $y_{i_0} \in B$ ,  $A^+ = A \cup \{y_{i_0}\}$ ,  $B^- = B \setminus \{y_{i_0}\}$ . Then

$$f(A^+) = f(A) + \frac{p}{p+1} \|c(A) - y_{i_0}\|^2, \quad (6)$$

$$f(B^-) = f(B) - \frac{q}{q-1} \|c(B) - y_{i_0}\|^2. \quad (7)$$

$$f(A \cup B) = f(A) + f(B) + p\|c(A) - c\|^2 + q\|c(B) - c\|^2$$

Proof: Because of  $f(\{y_{i_0}\}) = 0$  and  $c(A^+) = \frac{p}{p+1}c(A) + \frac{1}{p+1}y_{i_0}$ ,

$$f(A \cup B) = f(A) + f(B) + p\|c(A) - c\|^2 + q\|c(B) - c\|^2$$

Proof: Because of  $f(\{y_{i_0}\}) = 0$  and  $c(A^+) = \frac{p}{p+1}c(A) + \frac{1}{p+1}y_{i_0}$ ,

$$f(A^+) = f(A \cup \{y_{i_0}\}) = f(A) + p\|c(A) - c(A^+)\|^2 + 1 \cdot \|y_{i_0} - c(A^+)\|^2$$

$$f(A \cup B) = f(A) + f(B) + p\|c(A) - c\|^2 + q\|c(B) - c\|^2$$

Proof: Because of  $f(\{y_{i_0}\}) = 0$  and  $c(A^+) = \frac{p}{p+1}c(A) + \frac{1}{p+1}y_{i_0}$ ,

$$\begin{aligned} f(A^+) &= f(A \cup \{y_{i_0}\}) = f(A) + p\|c(A) - c(A^+)\|^2 + 1 \cdot \|y_{i_0} - c(A^+)\|^2 \\ &= f(A) + \frac{p}{(p+1)^2}\|c(A) - y_{i_0}\|^2 + \frac{p^2}{(p+1)^2}\|c(A) - y_{i_0}\|^2 \end{aligned}$$

$$f(A \cup B) = f(A) + f(B) + p\|c(A) - c\|^2 + q\|c(B) - c\|^2$$

Proof: Because of  $f(\{y_{i_0}\}) = 0$  and  $c(A^+) = \frac{p}{p+1}c(A) + \frac{1}{p+1}y_{i_0}$ ,

$$\begin{aligned} f(A^+) &= f(A \cup \{y_{i_0}\}) = f(A) + p\|c(A) - c(A^+)\|^2 + 1 \cdot \|y_{i_0} - c(A^+)\|^2 \\ &= f(A) + \frac{p}{(p+1)^2}\|c(A) - y_{i_0}\|^2 + \frac{p^2}{(p+1)^2}\|c(A) - y_{i_0}\|^2 \\ &= f(A) + \frac{p}{p+1}\|c(A) - y_{i_0}\|^2. \end{aligned}$$

$$f(A \cup B) = f(A) + f(B) + p\|c(A) - c\|^2 + q\|c(B) - c\|^2$$

Proof: Because of  $f(\{y_{i_0}\}) = 0$  and  $c(A^+) = \frac{p}{p+1}c(A) + \frac{1}{p+1}y_{i_0}$ ,

$$\begin{aligned} f(A^+) &= f(A \cup \{y_{i_0}\}) = f(A) + p\|c(A) - c(A^+)\|^2 + 1 \cdot \|y_{i_0} - c(A^+)\|^2 \\ &= f(A) + \frac{p}{(p+1)^2}\|c(A) - y_{i_0}\|^2 + \frac{p^2}{(p+1)^2}\|c(A) - y_{i_0}\|^2 \\ &= f(A) + \frac{p}{p+1}\|c(A) - y_{i_0}\|^2. \end{aligned}$$

Because of  $c(B) = \frac{q-1}{q}c(B^-) + \frac{1}{q}y_{i_0}$  and  $c(B^-) = \frac{q}{q-1}c(B) - \frac{1}{q-1}y_{i_0}$ ,

$$f(A \cup B) = f(A) + f(B) + p\|c(A) - c\|^2 + q\|c(B) - c\|^2$$

Proof: Because of  $f(\{y_{i_0}\}) = 0$  and  $c(A^+) = \frac{p}{p+1}c(A) + \frac{1}{p+1}y_{i_0}$ ,

$$\begin{aligned} f(A^+) &= f(A \cup \{y_{i_0}\}) = f(A) + p\|c(A) - c(A^+)\|^2 + 1 \cdot \|y_{i_0} - c(A^+)\|^2 \\ &= f(A) + \frac{p}{(p+1)^2}\|c(A) - y_{i_0}\|^2 + \frac{p^2}{(p+1)^2}\|c(A) - y_{i_0}\|^2 \\ &= f(A) + \frac{p}{p+1}\|c(A) - y_{i_0}\|^2. \end{aligned}$$

Because of  $c(B) = \frac{q-1}{q}c(B^-) + \frac{1}{q}y_{i_0}$  and  $c(B^-) = \frac{q}{q-1}c(B) - \frac{1}{q-1}y_{i_0}$ ,

$$f(B) = f(B^- \cup \{y_{i_0}\}) = f(B^-) + (q-1)\|c(B^-) - c(B)\|^2 + 1 \cdot \|y_{i_0} - c(B)\|^2$$

$$f(A \cup B) = f(A) + f(B) + p\|c(A) - c\|^2 + q\|c(B) - c\|^2$$

Proof: Because of  $f(\{y_{i_0}\}) = 0$  and  $c(A^+) = \frac{p}{p+1}c(A) + \frac{1}{p+1}y_{i_0}$ ,

$$\begin{aligned} f(A^+) &= f(A \cup \{y_{i_0}\}) = f(A) + p\|c(A) - c(A^+)\|^2 + 1 \cdot \|y_{i_0} - c(A^+)\|^2 \\ &= f(A) + \frac{p}{(p+1)^2}\|c(A) - y_{i_0}\|^2 + \frac{p^2}{(p+1)^2}\|c(A) - y_{i_0}\|^2 \\ &= f(A) + \frac{p}{p+1}\|c(A) - y_{i_0}\|^2. \end{aligned}$$

Because of  $c(B) = \frac{q-1}{q}c(B^-) + \frac{1}{q}y_{i_0}$  and  $c(B^-) = \frac{q}{q-1}c(B) - \frac{1}{q-1}y_{i_0}$ ,

$$\begin{aligned} f(B) &= f(B^- \cup \{y_{i_0}\}) = f(B^-) + (q-1)\|c(B^-) - c(B)\|^2 + 1 \cdot \|y_{i_0} - c(B)\|^2 \\ &= f(B^-) + \frac{q-1}{(q-1)^2}\|c(B) - y_{i_0}\|^2 + \|c(B) - y_{i_0}\|^2 \end{aligned}$$

$$f(A \cup B) = f(A) + f(B) + p\|c(A) - c\|^2 + q\|c(B) - c\|^2$$

Proof: Because of  $f(\{y_{i_0}\}) = 0$  and  $c(A^+) = \frac{p}{p+1}c(A) + \frac{1}{p+1}y_{i_0}$ ,

$$\begin{aligned} f(A^+) &= f(A \cup \{y_{i_0}\}) = f(A) + p\|c(A) - c(A^+)\|^2 + 1 \cdot \|y_{i_0} - c(A^+)\|^2 \\ &= f(A) + \frac{p}{(p+1)^2}\|c(A) - y_{i_0}\|^2 + \frac{p^2}{(p+1)^2}\|c(A) - y_{i_0}\|^2 \\ &= f(A) + \frac{p}{p+1}\|c(A) - y_{i_0}\|^2. \end{aligned}$$

Because of  $c(B) = \frac{q-1}{q}c(B^-) + \frac{1}{q}y_{i_0}$  and  $c(B^-) = \frac{q}{q-1}c(B) - \frac{1}{q-1}y_{i_0}$ ,

$$\begin{aligned} f(B) &= f(B^- \cup \{y_{i_0}\}) = f(B^-) + (q-1)\|c(B^-) - c(B)\|^2 + 1 \cdot \|y_{i_0} - c(B)\|^2 \\ &= f(B^-) + \frac{q-1}{(q-1)^2}\|c(B) - y_{i_0}\|^2 + \|c(B) - y_{i_0}\|^2 \\ &= f(B^-) + \frac{q}{q-1}\|c(B) - y_{i_0}\|^2. \end{aligned}$$

## Theorem

Let  $A = \{x_1, \dots, x_p\}$  and  $B = \{y_1, \dots, y_q\}$ ,  $q > 1$ , be two different sets in  $\mathbb{R}^n$ ,  $y_{i_0} \in B$ ,  $A^+ = A \cup \{y_{i_0}\}$ ,  $B^- = B \setminus \{y_{i_0}\}$ . Then

$$F(\{A, B\}) - F(\{A^+, B^-\}) = \frac{q}{q-1} \|c(B) - y_{i_0}\|^2 - \frac{p}{p+1} \|c(A) - y_{i_0}\|^2$$

## Theorem

Let  $A = \{x_1, \dots, x_p\}$  and  $B = \{y_1, \dots, y_q\}$ ,  $q > 1$ , be two different sets in  $\mathbb{R}^n$ ,  $y_{i_0} \in B$ ,  $A^+ = A \cup \{y_{i_0}\}$ ,  $B^- = B \setminus \{y_{i_0}\}$ . Then

$$F(\{A, B\}) - F(\{A^+, B^-\}) = \frac{q}{q-1} \|c(B) - y_{i_0}\|^2 - \frac{p}{p+1} \|c(A) - y_{i_0}\|^2$$

Proof: Since

$$f(A^+) = f(A) + \frac{p}{p+1} \|c(A) - y_{i_0}\|^2,$$

$$f(B^-) = f(B) - \frac{q}{q-1} \|c(B) - y_{i_0}\|^2,$$

## Theorem

Let  $A = \{x_1, \dots, x_p\}$  and  $B = \{y_1, \dots, y_q\}$ ,  $q > 1$ , be two different sets in  $\mathbb{R}^n$ ,  $y_{i_0} \in B$ ,  $A^+ = A \cup \{y_{i_0}\}$ ,  $B^- = B \setminus \{y_{i_0}\}$ . Then

$$F(\{A, B\}) - F(\{A^+, B^-\}) = \frac{q}{q-1} \|c(B) - y_{i_0}\|^2 - \frac{p}{p+1} \|c(A) - y_{i_0}\|^2$$

Proof: Since

$$f(A^+) = f(A) + \frac{p}{p+1} \|c(A) - y_{i_0}\|^2,$$

$$f(B^-) = f(B) - \frac{q}{q-1} \|c(B) - y_{i_0}\|^2,$$

$$F(\{A, B\}) - F(\{A^+, B^-\}) = (f(A) + f(B)) - (f(A^+) + f(B^-))$$

## Theorem

Let  $A = \{x_1, \dots, x_p\}$  and  $B = \{y_1, \dots, y_q\}$ ,  $q > 1$ , be two different sets in  $\mathbb{R}^n$ ,  $y_{i_0} \in B$ ,  $A^+ = A \cup \{y_{i_0}\}$ ,  $B^- = B \setminus \{y_{i_0}\}$ . Then

$$F(\{A, B\}) - F(\{A^+, B^-\}) = \frac{q}{q-1} \|c(B) - y_{i_0}\|^2 - \frac{p}{p+1} \|c(A) - y_{i_0}\|^2$$

Proof: Since

$$f(A^+) = f(A) + \frac{p}{p+1} \|c(A) - y_{i_0}\|^2,$$

$$f(B^-) = f(B) - \frac{q}{q-1} \|c(B) - y_{i_0}\|^2,$$

$$\begin{aligned} F(\{A, B\}) - F(\{A^+, B^-\}) &= (f(A) + f(B)) - (f(A^+) + f(B^-)) \\ &= (f(A) - f(A^+)) + (f(B) - f(B^-)) \end{aligned}$$

# First variation of a partition

## Definition

Let  $\Pi = \{\pi_1, \dots, \pi_k\}$  be some  $k$ -partition of the set  $S$  and let  $i, j \in \{1, \dots, k\}$ . Every partition  $\Pi'$  obtained by replacing only one vector  $\hat{a} \in \pi_i \in \Pi$  into the cluster  $\pi_j \in \Pi$  is called a first variation of a partition  $\Pi$ .

(9)

# First variation of a partition

## Definition

Let  $\Pi = \{\pi_1, \dots, \pi_k\}$  be some  $k$ -partition of the set  $S$  and let  $i, j \in \{1, \dots, k\}$ . Every partition  $\Pi'$  obtained by replacing only one vector  $\hat{a} \in \pi_i \in \Pi$  into the cluster  $\pi_j \in \Pi$  is called a first variation of a partition  $\Pi$ .

$$\Delta_{ij} := F(\Pi) - F(\Pi') = \frac{m_i}{m_i - 1} \|c(\pi_i) - \hat{a}\|^2 - \frac{m_j}{m_j + 1} \|c(\pi_j) - \hat{a}\|^2, \quad (8)$$

(9)

# First variation of a partition

## Definition

Let  $\Pi = \{\pi_1, \dots, \pi_k\}$  be some  $k$ -partition of the set  $S$  and let  $i, j \in \{1, \dots, k\}$ . Every partition  $\Pi'$  obtained by replacing only one vector  $\hat{a} \in \pi_i \in \Pi$  into the cluster  $\pi_j \in \Pi$  is called a first variation of a partition  $\Pi$ .

$$\Delta_{ij} := F(\Pi) - F(\Pi') = \frac{m_i}{m_i - 1} \|c(\pi_i) - \hat{a}\|^2 - \frac{m_j}{m_j + 1} \|c(\pi_j) - \hat{a}\|^2, \quad (8)$$

$$\Delta_{ij} > 0 \Leftrightarrow \frac{m_i}{m_i - 1} \|c(\pi_i) - \hat{a}\|^2 > \frac{m_j}{m_j + 1} \|c(\pi_j) - \hat{a}\|^2. \quad (9)$$

# Improved $k$ -means algorithm

$$\Delta_{ij} := F(\Pi) - F(\Pi') = \frac{m_i}{m_i - 1} \|c(\pi_i) - \hat{a}\|^2 - \frac{m_j}{m_j + 1} \|c(\pi_j) - \hat{a}\|^2,$$

$$\hat{\Delta}_{ij} := \|c(\pi_i) - \hat{a}\|^2 - \|c(\pi_j) - \hat{a}\|^2$$

$$\Delta_{ij} - \hat{\Delta}_{ij} = \frac{1}{m_i - 1} \|c(\pi_i) - \hat{a}\|^2 + \frac{1}{m_j + 1} \|c(\pi_j) - \hat{a}\|^2 \geq 0 \quad \Rightarrow \quad \Delta_{ij} \geq \hat{\Delta}_{ij}.$$

# Improved $k$ -means algorithm

$$\Delta_{ij} := F(\Pi) - F(\Pi') = \frac{m_i}{m_i - 1} \|c(\pi_i) - \hat{a}\|^2 - \frac{m_j}{m_j + 1} \|c(\pi_j) - \hat{a}\|^2,$$

$$\hat{\Delta}_{ij} := \|c(\pi_i) - \hat{a}\|^2 - \|c(\pi_j) - \hat{a}\|^2$$

$$\Delta_{ij} - \hat{\Delta}_{ij} = \frac{1}{m_i - 1} \|c(\pi_i) - \hat{a}\|^2 + \frac{1}{m_j + 1} \|c(\pi_j) - \hat{a}\|^2 \geq 0 \quad \Rightarrow \quad \Delta_{ij} \geq \hat{\Delta}_{ij}.$$

Improved  $k$ -means algorithm:

- $k$ -means algorithm

**Step 5:** If there exists a vector  $\hat{a} \in S$  from some cluster  $\pi_i$ ,  $|\pi_i| > 1$ , such that there exists a cluster  $\pi_j$ ,  $j \neq i$ , such that  $\|c(\pi_i) - \hat{a}\|^2 \geq \|c(\pi_j) - \hat{a}\|^2$ , let us then make the first variation of the partition  $\Pi$  by replacing vector  $\hat{a}$  from cluster  $\pi_i$  into cluster  $\pi_j$ .

# LS-optimal partition and $k$ -means optimal partition

## Theorem

*Each LS-optimal partition is a  $k$ -means optimal partition.*

# LS-optimal partition and $k$ -means optimal partition

## Theorem

*Each LS-optimal partition is a  $k$ -means optimal partition.*

Proof: For  $m = k = 1$  the assumption is obvious.

# LS-optimal partition and $k$ -means optimal partition

## Theorem

*Each LS-optimal partition is a  $k$ -means optimal partition.*

Proof: For  $m = k = 1$  the assumption is obvious.

Let  $1 < k < m$  and let  $\Pi^{(o)} = \{\pi_1^o, \dots, \pi_k^o\}$  be an LS-optimal partition. Assume that  $\Pi^{(o)}$  is not a  $k$ -means optimal partition.

# LS-optimal partition and $k$ -means optimal partition

## Theorem

*Each LS-optimal partition is a  $k$ -means optimal partition.*

Proof: For  $m = k = 1$  the assumption is obvious.

Let  $1 < k < m$  and let  $\Pi^{(o)} = \{\pi_1^o, \dots, \pi_k^o\}$  be an LS-optimal partition.

Assume that  $\Pi^{(o)}$  is not a  $k$ -means optimal partition.

Then there exists  $\hat{a}$  from some cluster  $\pi_i^o$  ( $m_i > 1$ ), which is closer to the center  $c(\pi_j^o)$ ,  $j \neq i$ ,

$$\|c(\pi_i^o) - \hat{a}\| > \|c(\pi_j^o) - \hat{a}\|.$$

# LS-optimal partition and $k$ -means optimal partition

## Theorem

*Each LS-optimal partition is a  $k$ -means optimal partition.*

Proof: For  $m = k = 1$  the assumption is obvious.

Let  $1 < k < m$  and let  $\Pi^{(o)} = \{\pi_1^o, \dots, \pi_k^o\}$  be an LS-optimal partition.

Assume that  $\Pi^{(o)}$  is not a  $k$ -means optimal partition.

Then there exists  $\hat{a}$  from some cluster  $\pi_i^o (m_i > 1)$ , which is closer to the center  $c(\pi_j^o)$ ,  $j \neq i$ ,

$$\|c(\pi_i^o) - \hat{a}\| > \|c(\pi_j^o) - \hat{a}\|.$$

$$\text{Since: } \frac{m_i}{m_i - 1} > 1 \quad \text{and} \quad 0 < \frac{m_j}{m_j + 1} < 1,$$

$$\frac{m_i}{m_i - 1} \|c(\pi_i^o) - \hat{a}\|^2 > \frac{m_j}{m_j + 1} \|c(\pi_j^o) - \hat{a}\|^2 \quad \Rightarrow \quad \Delta_{ij} > 0.$$

# LS-optimal partition and $k$ -means optimal partition

## Theorem

*Each LS-optimal partition is a  $k$ -means optimal partition.*

Proof: For  $m = k = 1$  the assumption is obvious.

Let  $1 < k < m$  and let  $\Pi^{(o)} = \{\pi_1^o, \dots, \pi_k^o\}$  be an LS-optimal partition.

Assume that  $\Pi^{(o)}$  is not a  $k$ -means optimal partition.

Then there exists  $\hat{a}$  from some cluster  $\pi_i^o (m_i > 1)$ , which is closer to the center  $c(\pi_j^o)$ ,  $j \neq i$ ,

$$\|c(\pi_i^o) - \hat{a}\| > \|c(\pi_j^o) - \hat{a}\|.$$

$$\text{Since: } \frac{m_i}{m_i - 1} > 1 \quad \text{and} \quad 0 < \frac{m_j}{m_j + 1} < 1,$$

$$\frac{m_i}{m_i - 1} \|c(\pi_i^o) - \hat{a}\|^2 > \frac{m_j}{m_j + 1} \|c(\pi_j^o) - \hat{a}\|^2 \Rightarrow \Delta_{ij} > 0.$$

Hence,  $\Pi^{(o)}$  is not an LS-optimal partition.

### Theorem

*Let  $1 \leq k < m$  and let  $\Pi$  and  $\Pi'$  be LS-optimal partitions of lengths  $k$  and  $k + 1$ , respectively. Then it holds*

$$F(\Pi') \leq F(\Pi).$$

### Theorem

*Let  $1 \leq k < m$  and let  $\Pi$  and  $\Pi'$  be LS-optimal partitions of lengths  $k$  and  $k + 1$ , respectively. Then it holds*

$$F(\Pi') \leq F(\Pi).$$

Proof: Note first that for every partition  $F(\Pi) \geq 0$ .

### Theorem

*Let  $1 \leq k < m$  and let  $\Pi$  and  $\Pi'$  be LS-optimal partitions of lengths  $k$  and  $k + 1$ , respectively. Then it holds*

$$F(\Pi') \leq F(\Pi).$$

Proof: Note first that for every partition  $F(\Pi) \geq 0$ .

Assume first that  $k + 1 = m$ . In that case  $F(\Pi') \leq F(\Pi)$ .

### Theorem

*Let  $1 \leq k < m$  and let  $\Pi$  and  $\Pi'$  be LS-optimal partitions of lengths  $k$  and  $k + 1$ , respectively. Then it holds*

$$F(\Pi') \leq F(\Pi).$$

Proof: Note first that for every partition  $F(\Pi) \geq 0$ .

Assume first that  $k + 1 = m$ . In that case  $F(\Pi') \leq F(\Pi)$ .

Assume now that  $k + 1 < m$ .

Let  $\Pi = \{\pi_1, \dots, \pi_k\}$  and  $\Pi'$  be optimal partitions of lengths  $k$  and  $k + 1$ , respectively.

### Theorem

*Let  $1 \leq k < m$  and let  $\Pi$  and  $\Pi'$  be LS-optimal partitions of lengths  $k$  and  $k + 1$ , respectively. Then it holds*

$$F(\Pi') \leq F(\Pi).$$

Proof: Note first that for every partition  $F(\Pi) \geq 0$ .

Assume first that  $k + 1 = m$ . In that case  $F(\Pi') \leq F(\Pi)$ .

Assume now that  $k + 1 < m$ .

Let  $\Pi = \{\pi_1, \dots, \pi_k\}$  and  $\Pi'$  be optimal partitions of lengths  $k$  and  $k + 1$ , respectively.

Because of  $k + 1 < m$ , there exists  $1 \leq j_0 \leq m$ , such that  $m_{j_0} = |\pi_{j_0}| > 1$ .

## Theorem

Let  $1 \leq k < m$  and let  $\Pi$  and  $\Pi'$  be LS-optimal partitions of lengths  $k$  and  $k + 1$ , respectively. Then it holds

$$F(\Pi') \leq F(\Pi).$$

Proof: Note first that for every partition  $F(\Pi) \geq 0$ .

Assume first that  $k + 1 = m$ . In that case  $F(\Pi') \leq F(\Pi)$ .

Assume now that  $k + 1 < m$ .

Let  $\Pi = \{\pi_1, \dots, \pi_k\}$  and  $\Pi'$  be optimal partitions of lengths  $k$  and  $k + 1$ , respectively.

Because of  $k + 1 < m$ , there exists  $1 \leq j_0 \leq m$ , such that  $m_{j_0} = |\pi_{j_0}| > 1$ .

Let  $\hat{a} \in \pi_{j_0}$  and let  $\hat{\pi}_{j_0} = \pi_{j_0} \setminus \{\hat{a}\} \neq \emptyset$ .

$$f(\hat{\pi}_{j_0}) = f(\pi_{j_0}) - \frac{m_{j_0}}{m_{j_0} - 1} \|c(\pi_{j_0}) - \hat{a}\|^2 \leq f(\pi_{j_0}) \quad \& \quad f(\hat{\pi}_{j_0}) = f(\pi_{j_0}) \Leftrightarrow \hat{a} = c(\pi_{j_0}).$$

$\Pi'' = \{\pi_1, \dots, \pi_{j_0-1}, \hat{\pi}_{j_0}, \pi_{j_0+1}, \dots, \pi_k, \{\hat{a}\}\}$  a partition of the length  $k + 1$ .

$$F(\Pi') \leq F(\Pi'') = \sum_{j=1}^k f(\pi_j'') \leq \sum_{j=1}^k f(\pi_j) \leq F(\Pi).$$

# Agglomerative Nesting

- B. S. Everitt, S. Landau, M. Leese, *Cluster analysis*, Wiley, London, 2001.
- G. Gan, C Ma, J. Wu, *Data clustering: theory, algorithms, and applications*, SIAM, Philadelphia, 2007.

# Agglomerative Nesting

- B. S. Everitt, S. Landau, M. Leese, *Cluster analysis*, Wiley, London, 2001.
- G. Gan, C Ma, J. Wu, *Data clustering: theory, algorithms, and applications*, SIAM, Philadelphia, 2007.

Dissimilarity criteria:

$$d_{\min}(\pi_i, \pi_j) = \min_{x \in \pi_i, x' \in \pi_j} d(x, x')$$

$$d_{\max}(\pi_i, \pi_j) = \max_{x \in \pi_i, x' \in \pi_j} d(x, x')$$

$$d_c(\pi_i, \pi_j) = d(c_i, c_j), \quad c_i, c_j \text{ centroids of clusters } \pi_i, \pi_j$$

$$d_{\text{avg}}(\pi_i, \pi_j) = \frac{1}{m_i m_j} \sum_{x \in \pi_i} \sum_{x' \in \pi_j} d(x, x')$$

## Algorithm

**(Agglomerative Nesting - AGNES)**

## Algorithm

### (Agglomerative Nesting - AGNES)

**Step 1:** Input  $1 \leq k \leq m$ ,  $(x_i, i = 1, \dots, m)$ , and set  $\mu = 0$  and  $\Pi^{(m)} = \{\{x_1\}, \dots, \{x_m\}\}$ ;

## Algorithm

### (Agglomerative Nesting - AGNES)

**Step 1:** Input  $1 \leq k \leq m$ ,  $(x_i, i = 1, \dots, m)$ , and set  $\mu = 0$  and  $\Pi^{(m)} = \{\{x_1\}, \dots, \{x_m\}\}$ ;

**Step 2:** Construct a dissimilarity matrix of the partition  $\Pi^{(m-\mu)}$ ,  $R_\mu \in \mathbb{R}^{(m-\mu) \times (m-\mu)}$  with elements  $r_{ij} = d(\pi_i, \pi_j)$ , where  $d(\pi_i, \pi_j)$  represents dissimilarity of clusters  $\pi_i$  and  $\pi_j$ .

## Algorithm

### (Agglomerative Nesting - AGNES)

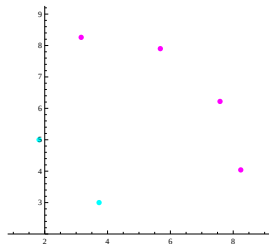
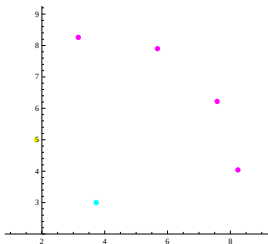
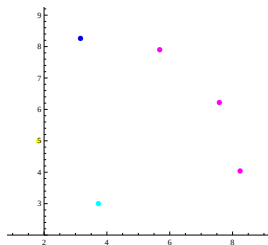
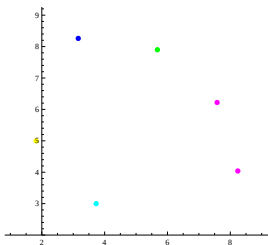
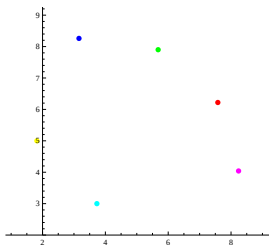
- Step 1:** Input  $1 \leq k \leq m$ ,  $(x_i, i = 1, \dots, m)$ , and set  $\mu = 0$  and  $\Pi^{(m)} = \{\{x_1\}, \dots, \{x_m\}\}$ ;
- Step 2:** Construct a dissimilarity matrix of the partition  $\Pi^{(m-\mu)}$ ,  $R_\mu \in \mathbb{R}^{(m-\mu) \times (m-\mu)}$  with elements  $r_{ij} = d(\pi_i, \pi_j)$ , where  $d(\pi_i, \pi_j)$  represents dissimilarity of clusters  $\pi_i$  and  $\pi_j$ .
- Step 3:** Find the smallest element  $r_{i_0 j_0}^*$  of the matrix  $R_\mu$ , and form a partition  $\Pi^{(m-\mu-1)} = (\Pi^{(m-\mu)} \setminus \{\pi_{i_0}, \pi_{j_0}\}) \cup \{\pi_{i_0} \cup \pi_{j_0}\}$ .

## Algorithm

### (Agglomerative Nesting - AGNES)

- Step 1:** Input  $1 \leq k \leq m$ ,  $(x_i, i = 1, \dots, m)$ , and set  $\mu = 0$  and  $\Pi^{(m)} = \{\{x_1\}, \dots, \{x_m\}\}$ ;
- Step 2:** Construct a dissimilarity matrix of the partition  $\Pi^{(m-\mu)}$ ,  $R_\mu \in \mathbb{R}^{(m-\mu) \times (m-\mu)}$  with elements  $r_{ij} = d(\pi_i, \pi_j)$ , where  $d(\pi_i, \pi_j)$  represents dissimilarity of clusters  $\pi_i$  and  $\pi_j$ .
- Step 3:** Find the smallest element  $r_{i_0 j_0}^*$  of the matrix  $R_\mu$ , and form a partition  $\Pi^{(m-\mu-1)} = (\Pi^{(m-\mu)} \setminus \{\pi_{i_0}, \pi_{j_0}\}) \cup \{\pi_{i_0} \cup \pi_{j_0}\}$ .
- Step 4:** If  $\mu < m - k$ , set  $\mu = \mu + 1$  and GoTo Step 2; Else STOP;

# Agglomerative Nesting



# Spectral Relaxation

$S = \{x_i \in \mathbb{R}^n : i = 1, \dots, m\} \subseteq \mathbb{R}^n$  (a set of data vectors)

$A = [a_1, \dots, a_m] \in \mathbb{R}^{n \times m} \in \mathbb{R}^{n \times m}$  (the matrix)

# Spectral Relaxation

$S = \{x_i \in \mathbb{R}^n : i = 1, \dots, m\} \subseteq \mathbb{R}^n$  (a set of data vectors)

$A = [a_1, \dots, a_m] \in \mathbb{R}^{n \times m} \in \mathbb{R}^{n \times m}$  (the matrix)

$\Pi(S) = \{\pi_1, \dots, \pi_k\}$  (partition of the set  $S$ )

$\Pi(A) = [A_1, \dots, A_k], \quad A_j = [a_1^{(j)}, \dots, a_{s_j}^{(j)}]$  (partition of the matrix  $A$ )

# Spectral Relaxation

$S = \{x_i \in \mathbb{R}^n : i = 1, \dots, m\} \subseteq \mathbb{R}^n$  (a set of data vectors)

$A = [a_1, \dots, a_m] \in \mathbb{R}^{n \times m} \in \mathbb{R}^{n \times m}$  (the matrix)

$\Pi(S) = \{\pi_1, \dots, \pi_k\}$  (partition of the set  $S$ )

$\Pi(A) = [A_1, \dots, A_k], \quad A_j = [a_1^{(j)}, \dots, a_{s_j}^{(j)}]$  (partition of the matrix  $A$ )

$\Pi(A) = AE, \quad E \in \mathbb{R}^{m \times m}$  (permutation matrix)

$s_1 + \dots + s_k = m$

## Example

$$A = \begin{bmatrix} 2 & 3 & 4 & 5 & 1 \\ -3 & -5 & -4 & -2 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$AE = \begin{bmatrix} 3 & 2 & 5 & 4 & 1 \\ -5 & -3 & -2 & -4 & 0 \end{bmatrix} = [A_1, A_2],$$

where

$$A_1 = [a_2, a_1] = \begin{bmatrix} 3 & 2 \\ -5 & -3 \end{bmatrix} = [a_1^{(1)}, a_2^{(1)}], \quad s_1 = 2,$$

$$A_2 = [a_4, a_3, a_5] = \begin{bmatrix} 5 & 4 & 1 \\ -2 & -4 & 0 \end{bmatrix} = [a_1^{(2)}, a_2^{(2)}, a_3^{(2)}], \quad s_2 = 3.$$

# Spectral relaxation

$$c_j = \frac{1}{s_j} \sum_{i=1}^{s_j} a_i^{(j)} = \frac{1}{s_j} A_j e, \quad e = [1, \dots, 1]^T \in \mathbb{R}^{s_j} \quad (\text{center of the cluster } A_j)$$

# Spectral relaxation

$$c_j = \frac{1}{s_j} \sum_{i=1}^{s_j} a_i^{(j)} = \frac{1}{s_j} A_j e, \quad e = [1, \dots, 1]^T \in \mathbb{R}^{s_j} \quad (\text{center of the cluster } A_j)$$

$$f(A_j) = \sum_{i=1}^{s_j} \|a_i^{(j)} - c_j\|^2 \quad (\text{objective function on the cluster } A_j)$$

# Spectral relaxation

$$c_j = \frac{1}{s_j} \sum_{i=1}^{s_j} a_i^{(j)} = \frac{1}{s_j} A_j e, \quad e = [1, \dots, 1]^T \in \mathbb{R}^{s_j} \quad (\text{center of the cluster } A_j)$$

$$f(A_j) = \sum_{i=1}^{s_j} \|a_i^{(j)} - c_j\|^2 \quad (\text{objective function on the cluster } A_j)$$

$$F(AE) = \sum_{j=1}^k f(A_j) \quad (\text{objective function on the partition } AE = [A_1, \dots, A_k])$$

# Spectral relaxation

$$c_j = \frac{1}{s_j} \sum_{i=1}^{s_j} a_i^{(j)} = \frac{1}{s_j} A_j e, \quad e = [1, \dots, 1]^T \in \mathbb{R}^{s_j} \quad (\text{center of the cluster } A_j)$$

$$f(A_j) = \sum_{i=1}^{s_j} \|a_i^{(j)} - c_j\|^2 \quad (\text{objective function on the cluster } A_j)$$

$$F(AE) = \sum_{j=1}^k f(A_j) \quad (\text{objective function on the partition } AE = [A_1, \dots, A_k])$$

- I. S. Dhillon, Y. Guan, B. Kulis, *Kernel k-means, spectral clustering and normalized cuts*, 2004.
- Stella X. Yu, Jianbo Shi, *Multiclass Spectral Clustering*, 2003.
- Hongyuan Zha, Xiaofeng He, Chris Ding, Horst Simon, Ming Gu, *Spectral Relaxation for K-means Clustering*, 2002.

# Objective function on the cluster $A_j$

$$f(A_j) = \sum_{i=1}^{s_j} \|a_i^{(j)} - c_j\|^2 = \sum_{i=1}^{s_j} (a_i^{(j)} - c_j)^T (a_i^{(j)} - c_j), \quad c_j = \frac{1}{s_j} \sum_{i=1}^{s_j} a_i^{(j)} = \frac{1}{s_j} A_j e$$

# Objective function on the cluster $A_j$

$$\begin{aligned} f(A_j) &= \sum_{i=1}^{s_j} \|a_i^{(j)} - c_j\|^2 = \sum_{i=1}^{s_j} (a_i^{(j)} - c_j)^T (a_i^{(j)} - c_j), & c_j &= \frac{1}{s_j} \sum_{i=1}^{s_j} a_i^{(j)} = \frac{1}{s_j} A_j e \\ &= \sum_{i=1}^{s_j} \text{tr} (a_i^{(j)} - c_j)(a_i^{(j)} - c_j)^T = \text{tr} \sum_{i=1}^{s_j} (a_i^{(j)} - c_j)(a_i^{(j)} - c_j)^T \end{aligned}$$

# Objective function on the cluster $A_j$

$$\begin{aligned} f(A_j) &= \sum_{i=1}^{s_j} \|a_i^{(j)} - c_j\|^2 = \sum_{i=1}^{s_j} (a_i^{(j)} - c_j)^T (a_i^{(j)} - c_j), & c_j &= \frac{1}{s_j} \sum_{i=1}^{s_j} a_i^{(j)} = \frac{1}{s_j} A_j e \\ &= \sum_{i=1}^{s_j} \text{tr} (a_i^{(j)} - c_j)(a_i^{(j)} - c_j)^T = \text{tr} \sum_{i=1}^{s_j} (a_i^{(j)} - c_j)(a_i^{(j)} - c_j)^T \\ &= \text{tr} \left( [a_1^{(j)} - c_j, a_2^{(j)} - c_j, \dots, a_{s_j}^{(j)} - c_j] [a_1^{(j)} - c_j, a_2^{(j)} - c_j, \dots, a_{s_j}^{(j)} - c_j]^T \right) \end{aligned}$$

## Objective function on the cluster $A_j$

$$\begin{aligned} f(A_j) &= \sum_{i=1}^{s_j} \|a_i^{(j)} - c_j\|^2 = \sum_{i=1}^{s_j} (a_i^{(j)} - c_j)^T (a_i^{(j)} - c_j), \quad c_j = \frac{1}{s_j} \sum_{i=1}^{s_j} a_i^{(j)} = \frac{1}{s_j} A_j e \\ &= \sum_{i=1}^{s_j} \text{tr} (a_i^{(j)} - c_j)(a_i^{(j)} - c_j)^T = \text{tr} \sum_{i=1}^{s_j} (a_i^{(j)} - c_j)(a_i^{(j)} - c_j)^T \\ &= \text{tr} \left( [a_1^{(j)} - c_j, a_2^{(j)} - c_j, \dots, a_{s_j}^{(j)} - c_j] [a_1^{(j)} - c_j, a_2^{(j)} - c_j, \dots, a_{s_j}^{(j)} - c_j]^T \right) \\ &= \text{tr} (A_j - c_j e^T)(A_j - c_j e^T)^T = \|A_j - c_j e^T\|_F^2 = \|A_j \left( I_{s_j} - \frac{e e^T}{s_j} \right)\|_F^2. \end{aligned}$$

# Objective function on the cluster $A_j$

Since  $\left(I_{s_j} - \frac{ee^T}{s_j}\right)$  is symmetric projector and  $(e^T e = s_j)$ ,

$$\left(I_{s_j} - \frac{ee^T}{s_j}\right)^2 = \left(I_{s_j} - 2\frac{ee^T}{s_j} + \frac{e(e^T e)e^T}{s_j^2}\right) = \left(I_{s_j} - \frac{ee^T}{s_j}\right),$$

# Objective function on the cluster $A_j$

Since  $\left(I_{s_j} - \frac{ee^T}{s_j}\right)$  is symmetric projector and  $(e^T e = s_j)$ ,

$$\left(I_{s_j} - \frac{ee^T}{s_j}\right)^2 = \left(I_{s_j} - 2\frac{ee^T}{s_j} + \frac{e(e^T e)e^T}{s_j^2}\right) = \left(I_{s_j} - \frac{ee^T}{s_j}\right),$$

$$\begin{aligned} f(A_j) &= \|A_j \left(I_{s_j} - \frac{ee^T}{s_j}\right)\|_F^2 \\ &= \text{tr} \left( A_j \left(I_{s_j} - \frac{ee^T}{s_j}\right) \left(I_{s_j} - \frac{ee^T}{s_j}\right)^T A_j^T \right) = \text{tr} \left( A_j \left(I_{s_j} - \frac{ee^T}{s_j}\right) A_j^T \right) \end{aligned}$$

# Objective function on the cluster $A_j$

Since  $\left(I_{s_j} - \frac{ee^T}{s_j}\right)$  is symmetric projector and  $(e^T e = s_j)$ ,

$$\left(I_{s_j} - \frac{ee^T}{s_j}\right)^2 = \left(I_{s_j} - 2\frac{ee^T}{s_j} + \frac{e(e^T e)e^T}{s_j^2}\right) = \left(I_{s_j} - \frac{ee^T}{s_j}\right),$$

$$\begin{aligned} f(A_j) &= \|A_j \left(I_{s_j} - \frac{ee^T}{s_j}\right)\|_F^2 \\ &= \text{tr} \left( A_j \left(I_{s_j} - \frac{ee^T}{s_j}\right) \left(I_{s_j} - \frac{ee^T}{s_j}\right)^T A_j^T \right) = \text{tr} \left( A_j \left(I_{s_j} - \frac{ee^T}{s_j}\right) A_j^T \right) \\ &= \text{tr} \left( \left(I_{s_j} - \frac{ee^T}{s_j}\right) A_j^T A_j \right) \end{aligned}$$

# Objective function on the partition $AE = [A_1, \dots, A_k]$

$$F(AE) = \sum_{j=1}^k f(A_j) = \sum_{j=1}^k \text{tr} \left( \left( I_{s_j} - \frac{ee^T}{s_j} \right) A_j^T A_j \right)$$

# Objective function on the partition $AE = [A_1, \dots, A_k]$

$$\begin{aligned} F(AE) &= \sum_{j=1}^k f(A_j) &= \sum_{j=1}^k \text{tr} \left( \left( I_{s_j} - \frac{ee^T}{s_j} \right) A_j^T A_j \right) \\ & &= \sum_{j=1}^k \text{tr} A_j^T A_j - \sum_{j=1}^k \text{tr} \left( \frac{ee^T}{s_j} A_j^T A_j \right) \end{aligned}$$

# Objective function on the partition $AE = [A_1, \dots, A_k]$

$$\begin{aligned} F(AE) &= \sum_{j=1}^k f(A_j) &= \sum_{j=1}^k \text{tr} \left( \left( I_{s_j} - \frac{ee^T}{s_j} \right) A_j^T A_j \right) \\ &= \sum_{j=1}^k \text{tr} A_j^T A_j - \sum_{j=1}^k \text{tr} \left( \frac{ee^T}{s_j} A_j^T A_j \right) \\ &= \sum_{j=1}^k \text{tr} A_j^T A_j - \sum_{j=1}^k \text{tr} \left( \frac{e^T}{\sqrt{s_j}} A_j^T A_j \frac{e}{\sqrt{s_j}} \right) \end{aligned}$$

# Objective function on the partition $AE = [A_1, \dots, A_k]$

$$\begin{aligned} F(AE) &= \sum_{j=1}^k f(A_j) &= \sum_{j=1}^k \operatorname{tr} \left( \left( I_{s_j} - \frac{ee^T}{s_j} \right) A_j^T A_j \right) \\ &= \sum_{j=1}^k \operatorname{tr} A_j^T A_j - \sum_{j=1}^k \operatorname{tr} \left( \frac{ee^T}{s_j} A_j^T A_j \right) \\ &= \sum_{j=1}^k \operatorname{tr} A_j^T A_j - \sum_{j=1}^k \operatorname{tr} \left( \frac{e^T}{\sqrt{s_j}} A_j^T A_j \frac{e}{\sqrt{s_j}} \right) \\ &= \operatorname{tr} \sum_{j=1}^k A_j^T A_j - \operatorname{tr} \sum_{j=1}^k \left( \frac{e^T}{\sqrt{s_j}} A_j^T A_j \frac{e}{\sqrt{s_j}} \right). \end{aligned}$$

Let  $e^{(j)} = [1, \dots, 1]^T \in \mathbb{R}^{s_j}$  and

$$X = \text{diag} \left( \frac{e^{(1)}}{\sqrt{s_1}}, \frac{e^{(2)}}{\sqrt{s_2}}, \dots, \frac{e^{(k)}}{\sqrt{s_k}} \right) \in \mathbb{R}^{m \times k}.$$

Therefore

Let  $e^{(j)} = [1, \dots, 1]^T \in \mathbb{R}^{s_j}$  and

$$X = \text{diag} \left( \frac{e^{(1)}}{\sqrt{s_1}}, \frac{e^{(2)}}{\sqrt{s_2}}, \dots, \frac{e^{(k)}}{\sqrt{s_k}} \right) \in \mathbb{R}^{m \times k}.$$

Therefore

$$F(AE) = \text{tr } A^T A - \text{tr } (X^T A^T A X)$$

Let  $e^{(j)} = [1, \dots, 1]^T \in \mathbb{R}^{s_j}$  and

$$X = \text{diag} \left( \frac{e^{(1)}}{\sqrt{s_1}}, \frac{e^{(2)}}{\sqrt{s_2}}, \dots, \frac{e^{(k)}}{\sqrt{s_k}} \right) \in \mathbb{R}^{m \times k}.$$

Therefore

$$F(AE) = \text{tr } A^T A - \text{tr} (X^T A^T A X)$$

$$\underset{E}{\text{argmin}} F(AE) = \underset{X}{\text{argmax}} \text{tr} (X^T A^T A X)$$

Let  $e^{(j)} = [1, \dots, 1]^T \in \mathbb{R}^{s_j}$  and

$$X = \text{diag} \left( \frac{e^{(1)}}{\sqrt{s_1}}, \frac{e^{(2)}}{\sqrt{s_2}}, \dots, \frac{e^{(k)}}{\sqrt{s_k}} \right) \in \mathbb{R}^{m \times k}.$$

Therefore

$$F(AE) = \text{tr } A^T A - \text{tr} (X^T A^T A X)$$

$$\underset{E}{\text{argmin}} F(AE) = \underset{X}{\text{argmax}} \text{tr} (X^T A^T A X)$$

$$\min_E F(AE) = \text{tr } A^T A - \max_X \text{tr} (X^T A^T A X).$$

## Theorem

Let  $H \in \mathbb{R}^{n \times n}$  be a symmetric matrix with eigenvalues  $\lambda_1 \geq \dots \geq \lambda_n$  and corresponding eigenvectors  $U = [u_1, \dots, u_n]$ . Then

$$\lambda_1 + \dots + \lambda_k = \max_{X^T X = I_k} \text{tr}(X^T H X).$$

Moreover, the optimal  $X^*$  is given by  $X^* = [u_1, \dots, u_k]Q$  with  $Q$  an arbitrary orthogonal matrix.

$$\min_E F(AE) \geq \text{tr } A^T A - \max_X \text{tr}(X^T A^T A X) = \sum_{i=k+1}^{\min\{m,n\}} \sigma_i^2(A).$$

## Theorem

Let  $H \in \mathbb{R}^{n \times n}$  be a symmetric matrix with eigenvalues  $\lambda_1 \geq \dots \geq \lambda_n$  and corresponding eigenvectors  $U = [u_1, \dots, u_n]$ . Then

$$\lambda_1 + \dots + \lambda_k = \max_{X^T X = I_k} \text{tr}(X^T H X).$$

Moreover, the optimal  $X^*$  is given by  $X^* = [u_1, \dots, u_k]Q$  with  $Q$  an arbitrary orthogonal matrix.

$$\min_E F(AE) \geq \text{tr } A^T A - \max_X \text{tr}(X^T A^T A X) = \sum_{i=k+1}^{\min\{m,n\}} \sigma_i^2(A).$$

An application using pivoted QR-decomposition:

H. Zha, X. He, C. Ding, H. Simon, M. Gu, *Spectral Relaxation for k-means Clustering*, Advances in Neural Information Systems, 2002.

# A generalization

$G \in \mathbb{R}^{n \times n}$  – positive definite symmetric matrix

$$\|\cdot\|_G : \mathbb{R}^n \rightarrow [0, +\infty), \quad \|x\|_G^2 = x^T G x$$

$$d_G : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, +\infty), \quad d_G(x, y) = (x - y)^T G (x - y)$$

# A generalization

$G \in \mathbb{R}^{n \times n}$  – positive definite symmetric matrix

$$\|\cdot\|_G : \mathbb{R}^n \rightarrow [0, +\infty), \quad \|x\|_G^2 = x^T G x$$

$$d_G : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, +\infty), \quad d_G(x, y) = (x - y)^T G (x - y)$$

$$F_G(a) = \sum_{x_i \in \pi_j} d_G^2(a, x_i) = \sum_{x_i \in \pi_j} (x_i - a)^T G (x_i - a)$$

$$\operatorname{argmin} F_G = c_j = \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i$$

# A generalization

$G \in \mathbb{R}^{n \times n}$  – positive definite symmetric matrix

$$\|\cdot\|_G : \mathbb{R}^n \rightarrow [0, +\infty), \quad \|x\|_G^2 = x^T G x$$

$$d_G : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, +\infty), \quad d_G(x, y) = (x - y)^T G (x - y)$$

$$F_G(a) = \sum_{x_i \in \pi_j} d_G^2(a, x_i) = \sum_{x_i \in \pi_j} (x_i - a)^T G (x_i - a)$$

$$\operatorname{argmin} F_G = c_j = \frac{1}{m_j} \sum_{x_i \in \pi_j} x_i$$

$$F_G(\Pi) = \sum_{j=1}^k \sum_{x_i \in \pi_j} (x_i - c_j)^T G (x_i - c_j).$$

Other geometrical objects as centers of clusters:  
line, line segment, plane, circle

J. Kogan, *Introduction to Clustering Large and High-Dimensional Data*,  
Cambridge University Press, 2007.

H. Späth, *Clusterwise linear least absolute deviations regression*, Computing  
**37**(1986)