Problemi najmanjih kvadrata i najmanjih apsolutnih odstupanja s primjenama

Rudolf Scitovski

Odjel za matematiku, Sveučilište u Osijeku

Rijeka, 24. veljače 2011

- (w_i, x_i, y_i), $i = 1, \ldots, m$ – podaci
- $x \mapsto f(x; a)$ – funkcija model
- $a \in \mathcal{P} \subseteq \mathbb{R}^n$ – prostor parametara
- $y_i = f(x_i; a) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$
- $r(a)$, $r_i(a) = y_i - f(x_i; a)$ – vektor reziduala

Postoji li $a^\star \in \mathcal{P}$, takav da bude

$$F(a^\star) = \inf_{a \in \mathcal{P}} F(a), \quad F(a) = \|r(a)\|$$

$$F(a) = \|r(a)\|_2^2 = \sum_{i=1}^{m} w_i (y_i - f(t_i; a))^2$$

$$\text{grad } F(a) = J^T(a) r(a)$$

$$H_F(a) = J^T(a) J(a) + \sum_{k=1}^{m} r_k(a) H_k(a)$$

$$J(a) = \begin{bmatrix} \frac{\partial r_1(a)}{\partial a_1} & \cdots & \frac{\partial r_1(a)}{\partial a_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_m(a)}{\partial a_1} & \cdots & \frac{\partial r_m(a)}{\partial a_n} \end{bmatrix}, \quad (H_k(a))_{ij} = \frac{\partial^2 r_k(a)}{\partial a_i \partial a_j}$$

$$y_i = f(x_i + \delta_i; a) + \varepsilon_i, \qquad i = 1, \dots, m$$

δ_i – nepoznata aditivna pogreška nezavisne varijable x_i,

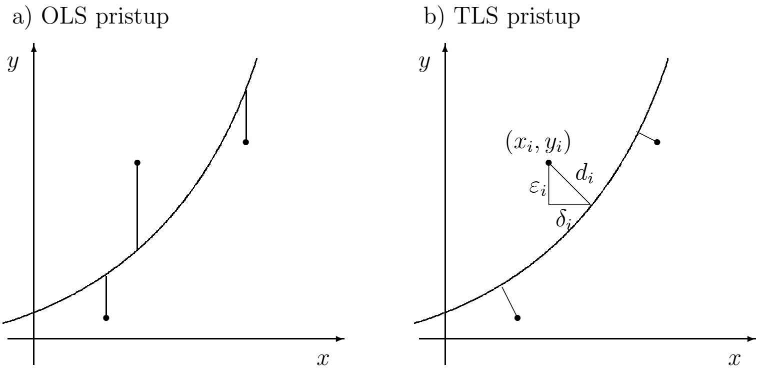
ε_i–nepoznata aditivna pogreška zavisne varijable y_i.

$$F(a, \delta) = \sum_{i=1}^{m} p_i \Big[\big(f(x_i + \delta_i; a) - y_i \big)^2 + \delta_i^2 \Big], \qquad \delta := (\delta_1, \dots, \delta_m)^T.$$

TLS-problem:

Postoji li točka $(a^\star, \delta^\star) \in \mathcal{P} \times \mathbb{R}^m$, takva da je

$$F(a^\star, \delta^\star) = \inf_{(a,\delta) \in \mathcal{P} \times \mathbb{R}^m} F(a, \delta) \ ?$$



D. Jukić , R. Scitovski and H. Späth, Partial linearization of one class of the nonlinear total least squares problem by using the inverse model function, Computing 62(1999), 163-178.

Problem najmanjih apsolutnih odstupanja (Least Absolute Deviations – LAD)

$$F(a) = \|r(a)\|_1 = \sum_{i=1}^{m} w_i |y_i - f(t_i; a)|$$

$$F(a) = \|r(a)\|_p^p = \sum_{i=1}^{m} w_i |y_i - f(t_i; a)|^p, \quad 1 \le p < +\infty$$

Josip Ruđer Bošković (1711-1787), hrvatskoi znanstvenik (matematičar, fizičar, astronom i filozof) rođenom u Dubrovniku (Dodge, 1987), (Schöbel, 2003).

Serija konferencija posvećena J. R. Boškoviću: "Statistical Data Analysis Based on the L_1-Norm and Related Methods (Y. Dodge, Neuchâtel, Switzerland)

Marquis Pierre-Simon de Laplace (1749–1827), francuski matematičar i astronom: *Mécanique Céleste* (1798–1825), I-V: Dostupno na engleskom (N. Bewditch, 1829): `http://www.archive.org/details/mcaniquecles02laplrich`

- Y. Dodge, *An introduction to L_1-norm based statistical data analysis*, Computational Statistics & Data Analysis **5**(1997), 239–253
- A. Schöbel, *Anchored hyperplane location problems*, Discrete & Computational Geometry **29**(2003), 229–238

- R. Scitovski, Some special nonlinear least squares problems, Radovi matematički 4(1988), 279-298.
- R. Scitovski and D. Jukić, Total least squares problem for exponential function, Inverse Problems 12(1996), 341-349.
- D. Jukić and R. Scitovski, Existence of optimal solution for exponential model by least squares, J. Comput. Appl. Math. 78(1997), 317-328.
- D. Jukić , T. Marosević and R. Scitovski, Discrete total l_p-norm approximation problem for exponential function, Appl. Math. Comput. 94(1998), 137-143.
- D. Jukić and R. Scitovski, The best least squares approximation problem for a 3-parametric exponential regression model, ANZIAM J. 42(2000), 254-266.

Primjene:

- R. Scitovski and D. Jukić, Analysis of solutions of the least squares problem, Mathematical Communications 4(1999), 53-61.
- R. Scitovski, S.Kosanović, Rate of change in economics research, Economics analysis and workers management 19(1985), 65-75.

$$\frac{dy(t)}{dt} = c\,y(t)(A - y(t))$$

- B.Schön, R. Scitovski, Pokušaj prilagođavanja logističke funkcije zadanom empirijskom statističkom materijalu, Statistička revija 27(1977), 178-194.
- R. Scitovski, Searching method and existence of solution of special nonlinear least squares problems, Glasnik matematički 20(40), 1985, 451-466.
- R. Scitovski, A special nonlinear least squares problem, Journal of Computational and Applied Mathematics 53(1994), 323 - 331.
- D. Jukić and R. Scitovski, Solution of the least squares problem for logistic function, J. Comput. Appl. Math. 156(2003), 159-177.

$$\frac{dy(t)}{dt} = c\,y(t)\left(1 - \left(\frac{y(t)}{A}\right)^{\gamma}\right), \qquad A, c, \gamma > 0.$$

- D. Jukić and R. Scitovski, The existence of the optimal parameters of the generalized logistic function, Appl. Math. Comput. 77(1996), 281-294.
- D. Jukić and R. Scitovski Existence results for special nonlinear total least squares problem, J. Math. Anal. Appl. 226(1998), 348-363.

Primjene:

- G.Kralik, R. Scitovski, D.Senčić, Application of asymmetric S-function for analysis and valuation of the growth of boars, Stočarstvo 47(1993), 425-433.
- G.Kralik, R. Scitovski, Istraživanje značajki rasta brojlera pomoću asimetrične S-funkcije, Stočarstvo 47(1993), 207-213.
- G. Kralik, G. Kušec, R. Scitovski, Z. Škrtić, A. Petričević, Growth and carcass quality of broilers, Czech J. Animal Science, 44(1999), 233-237.

Razni modeli

- R. Scitovski, Krivulja magnetiziranja, Elektrotehnika 31(1988), 31-35.
- R. Scitovski, Analysis of parameter identification problem, Applied Mathematics and Computation, 82(1997), 39-55.
- R. Scitovski and M. Meler, Solving parameter estimation problem in New Product Diffusion Models, Applied Mathematics and Computation 127(2002), 45-63.
- D. Jukić , G. Kralik and R. Scitovski, Least squares fitting Gompertz curve, J. Comput. Appl. Math. 169(2004), 359-375.
- D. Jukić and R. Scitovski, Least squares fitting Gaussian type curve, Appl. Math. Comput. 167 (2005), 286-298.
- D. Jukić, K. Sabo, R. Scitovski, Total least squares fitting Michaelis-Menten enzyme kinetic model function, J. Comput. Appl. Math. 201(2007), 230-246.
- D. Jukić, K. Sabo, R. Scitovski, A review of existence criteria for parameter estimation of the Michaelis-Menten regression model, Annali dell'Universita' di Ferrara 53 (2007), 281-291.
- D.Jukić, M.Benšić, R.Scitovski, On the existence of the nonlinear weighted least squares estimate for a three-parameter Weibull distribution, Computational Statistics & Data Analysis 52 (2008) , 4502-4511

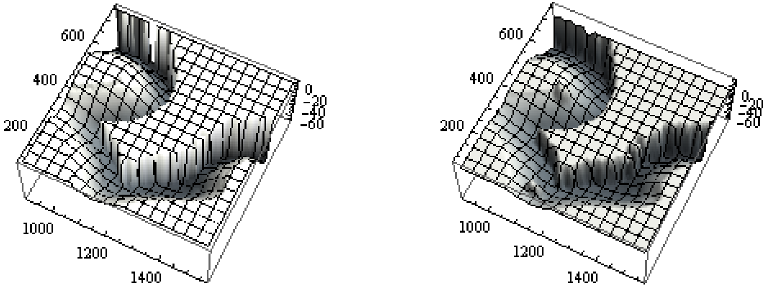
Pronaći "razumnu" vrijednost vektora parametara $a \in \mathbb{R}^n$, tako da uz optimalan izbor početnih uvjeta, rješenje $t \mapsto y(t)$ diferencijalne jednadžbe

$$\frac{d^2y}{dt^2} = f(t, y(t), y'(t); a)$$

"dobro" fituje eksperimentalne podatke (t_i, y_i), $i = 1, \ldots, m$.

- R. Scitovski and D. Jukić , A method for solving the parameter identification problem for ordinary differential equations of the second order, Appl. Math. Comput. 74 (1996), 273-291.
- R. Scitovski, D. Jukić and I. Urbiha, Solving the parameter identification problem by using TL_p spline,Mathematical Communications-Supplement 1(2001), 81-91.
- K.E.Nyarko and R.Scitovski, Solving the parameter identification problem of mathematical model using genetic algorithms, Applied Mathematics and Computation, 153(2004) 651-658.

Rekonstrukcija plohe



Pakleni kanal – Hvar

- R. Scitovski, Surface generating on the basis of experimental data, Mathematical Communications 2(1997), 59-67.
- R. Scitovski, Š. Ungar and D. Jukić, Approximating surfaces by moving total least squares method, Appl. Math. Comp. 93(1998), 219-232.

- R.Scitovski, G.Kralik, K.Sabo, T.Jelen, A mathematical model of controlling the growth of tissue in pigs, Applied Mathematics and Computation, 181(2006) 1126-1138
- T. Marošević, R.Scitovski, An application of a few inequalities among sequences in electoral systems, Appl. Math. Comput., 194(2007), 480-485.
- J.Cumin, B.Grizelj, R.Scitovski, Numerical solving of ballistic flight equations for big bore air rifle, Tehnički vjesnik. 16 (2009) , 41-46
- K.Scitovski, R.Scitovski, Rekonstrukcija meridijanske elipse na bazi eksperimentalnih mjerenja, Geodetski list 60(83)(2009) 19-35
- G. Blagojević, D. Jukić, K. Sabo, R. Scitovski, Matematički model endocitoze, Seminar za optimizaciju i primjene, Osijek, 2009.

$\mathbf{X}\mathbf{a} = \mathbf{z}$

$\mathbf{X} \in \mathbb{R}^{m\times n}$, $m \geq n$ – matrica punog ranga po stupcima,

$\mathbf{x}_i^T$ – i-ti redak matrica $\mathbf{X}$

$\mathbf{a} \in \mathbb{R}^n, \quad \mathbf{z} \in \mathbb{R}^m, \quad I = \{1, \ldots, m\}$

$$F(\mathbf{a}) = \|\mathbf{z} - \mathbf{X}\mathbf{a}\|_1 = \sum_{i\in I} |r_i(\mathbf{a})|, \quad r_i(\mathbf{a}) = z_i - \mathbf{a}^T\mathbf{x}_i,$$

$$\mathbf{a}^* = \operatorname*{argmin}_{a\in\mathbb{R}^n} F(a)$$

K.Sabo, R.Scitovski, I.Vazler, Searching for a best LAD-solution of an overdetermined system of linear equations motivated by searching for a best LAD-hyperplane on the basis of given data, JOTA (The Journal of Optimization Theory and Applications), DOI 10.1007/s10957-010-9791-1

http://www.mathos.hr/seminar/Software.html

- Procjena parametara LAD-hiperravnine $\mathbf{x} \mapsto \mathbf{a}^T\mathbf{x}$, $\mathbf{x} \in \mathbb{R}^n$, na bazi eksperimentalnih podataka $(\mathbf{x}_i, z_i)$, $\mathbf{x}_i = (x_1^{(i)}, \ldots, x_n^{(i)})^T \in \mathbb{R}^n$, $z_i \in \mathbb{R}$, $i = 1, \ldots, m$
- Procjena LAD-optimalnih parametara linearne regresije

Example 1

Promatramo sustav $\mathbf{X}\alpha = \mathbf{z}$, $\mathbf{X} \in \mathbb{R}^{m\times 1}$, $\mathbf{z} \in \mathbb{R}^m$.

$$\min_{\alpha\in\mathbb{R}} \sum_{i=1}^{m} |z_i - \alpha x_i| = \sum_{x_i=0} |z_i| + \min_{\alpha\in\mathbb{R}} \sum_{x_i\neq 0} |x_i| \left| \frac{z_i}{x_i} - \alpha \right|,$$

Lemma 1

Let (w_i, y_i), $i \in I = \{1,\dots,m\}$, $m \geq 2$, *be some data, where* $y_1 \leq y_2 \leq \ldots \leq y_m$ *are real numbers and* $w_i > 0$ *are the corresponding data weights. Then there exists a* $\mu \in I$, *such that* $y_\mu \in \mathrm{Med}(w,y)$. *Therefore, by denoting*

$$J := \Big\{\nu \in I : \sum_{i=1}^{\nu} w_i \leq \frac{W}{2}\Big\},$$

where $W := \sum_{i=1}^{m} w_i$, *the following holds:*

(a) *if* $J = \emptyset$, *then* $\mathrm{Med}(w,y) = \{y_1\}$;

(b) *if* $J \neq \emptyset$ *and* $\nu_0 := \max J$, *then*

  (i) *if* $\sum_{i=1}^{\nu_0} w_i < \frac{W}{2}$, *then* $\mathrm{Med}(w,y) = \{y_{\nu_0+1}\}$;

  (ii) *if* $\sum_{i=1}^{\nu_0} w_i = \frac{W}{2}$, *then* $\mathrm{Med}(w,y) = [y_{\nu_0}, y_{\nu_0+1}]$.

Corollary 1

Let $y_1 \le y_2 \le \cdots \le y_m$ be some real numbers. Then there exists a $\mu \in \{1, \ldots, m\}$, such that $y_\mu \in \operatorname{Med}(y)$. Consequently,

(i) *if m is odd ($m = 2k + 1$), then $\operatorname{Med}(y) = \{y_{k+1}\}$;*

(ii) *if m is even ($m = 2k$), then $\operatorname{Med}(y) = [y_k, y_{k+1}]$.*

- K.Sabo, R.Scitovski, The best least absolute deviations line - properties and two efficient methods for its derivation, ANZIAM J. 50(2008), 185-198
- I.Vazler, K.Sabo, R.Scitovski, Weighted median of the data in solving least absolute deviations problems, Communications in Statistics – Theory and Methods
- 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

LAD-line

$$G(a^*, b^*) = \min_{(a,b)\in\mathbb{R}^2} G(a, b), \qquad G(a, b) = \sum_{i=1}^{m} w_i \, |y_i - a x_i - b|.$$

Lemma 2

Let $\Lambda = \{T_i = (x_i, y_i) \in \mathbb{R}^2 : i \in I\}$, where $I = \{1, \dots, m\}$, $m \geq 2$, be a set of points in the plane such that $x_1 \leq \cdots \leq x_m$ and $x_1 < x_m$, and let $w_i > 0$ be the corresponding weights.

Then for arbitrary $T_\mu = (x_\mu, y_\mu) \in \mathbb{R}^2$, there exists a $T_\nu = (x_\nu, y_\nu) \in \Lambda$, $x_\nu \neq x_\mu$, such that for $a^* = \frac{y_\nu - y_\mu}{x_\nu - x_\mu}$

$$\overline{G}(a; T_\mu) \geq \overline{G}(a^*; T_\mu), \quad \forall\, a \in \mathbb{R}, \quad \textit{where}$$

$$\overline{G}(a; T_\mu) = \sum_{i=1}^{m} w_i \, |y_i - a(x_i - x_\mu) - y_\mu|.$$

http://www.mathos.hr/seminar/Software.html

Lemma 3

(i) $\Lambda = \{T_i = (x_i, y_i, z_i) \in \mathbb{R}^3 : i \in I\}$ *a set of points in space with corresponding data weights* $w_i > 0$, *such that the set of points* $\mathcal{L} = \{P_i = (x_i, y_i) \in \mathbb{R}^2 : i \in I\}$ *does not lie on any line;*

(ii) $T_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$ *and* $T_\mu = (x_\mu, y_\mu, z_\mu) \in \Lambda \setminus \{T_0\}$ *such that* $(x_\mu, y_\mu) \neq (x_0, y_0)$;

(iii) $I_0 = \{i \in I : (y_i - y_\mu)(x_0 - x_\mu) - (x_i - x_\mu)(y_0 - y_\mu) = 0\}$.

Then there exists a third point $T_\nu \in \Lambda \setminus M(T_0, T_\mu)$ *such that a best LAD-plane containing* T_0 *and* T_μ *passes through these three points. Thereby,*

(a) *if* $x_0 \neq x_\mu$, *then a best LAD-plane is determined by the affine function*

$$f_1(x, y; \beta^*) = z_\mu + \left(\frac{z_0 - z_\mu}{x_0 - x_\mu} - \beta^* \frac{y_0 - y_\mu}{x_0 - x_\mu} \right) (x - x_\mu) + \beta^*(y - y_\mu),$$

where β^* *is obtained by solving the weighted median problem*

$$\sum_{i \in I \setminus I_0} w_i \left| z_i - z_\mu - \frac{z_0 - z_\mu}{x_0 - x_\mu}(x_i - x_\mu) - \beta \left(y_i - y_\mu - \frac{y_0 - y_\mu}{x_0 - x_\mu}(x_i - x_\mu) \right) \right| \longrightarrow \min_\beta,$$

which also defines the third point $T_\nu \in \Lambda \setminus M(T_0, T_\mu)$, *through which this best LAD-plane passes.*

Overdetermined system – degenerate case

Theorem 1

Let $\mathbf{X} \in \mathbb{R}^{m\times n}$, $m \geq n$, *be the matrix of full column rank,* $\mathbf{z} \in \mathbb{R}^m$, $\mathbf{B} = [\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_n}]^T \in \mathbb{R}^{n\times n}$, *a nonsingular square submatrix of the matrix* $\mathbf{X}$, $\mathbf{z}_B = (z_{i_1}, \ldots, z_{i_n})^T$, $I_B = \{i_1, \ldots, i_n\}$, *and let*

(i) $\hat{\mathbf{a}} \in \mathbb{R}^n$ *be the solution of the system* $\mathbf{B}\mathbf{a} = \mathbf{z}_B$, *and* $\mathbf{B}^{-1} = [\mathbf{d}_1, \ldots, \mathbf{d}_n]$,

(ii) $I_0 = \{i \in I : r_i(\hat{\mathbf{a}}) = 0\}$,

(iii) $h(\hat{\mathbf{a}}) = \sum_{i\in I\setminus I_0} \sigma_i(\hat{\mathbf{a}})\mathbf{x}_i \in \partial F(\hat{\mathbf{a}})$.

If there exists $j_0 \in I_B$ *such that* $|\mathbf{d}_{j_0}^T \mathbf{h}(\hat{\mathbf{a}})| > 1 + \sum_{i\in I_0\setminus I_B} |\mathbf{d}_{j_0}^T \mathbf{x}_i|$, *then*

$$\vartheta^* = \operatorname*{med}_{i\in I_0\cup J} (w_i, \rho_i) \neq 0, \quad w_i = |\mathbf{d}_{j_0}^T \mathbf{x}_i|, \quad \rho_i = \begin{cases} 0, & i \in I_0 \\ \frac{r_i(\hat{\mathbf{a}})}{\mathbf{d}_{j_0}^T \mathbf{x}_i}, & i \in J \end{cases},$$

where $J = \{i \in I \setminus I_B : \mathbf{d}_{j_0}^T \mathbf{x}_i \neq 0\}$, *and the following holds*

$$F(\hat{\mathbf{a}} + \vartheta^* \mathbf{d}_{j_0}) < F(\hat{\mathbf{a}}).$$

Overdetermined system – nondegenerate case

Theorem 2

By the assumption as in Theorem 1, *let* $I_0 = I_B$. *Then,*

I. *Functional* F *attains its global minimum for* $\hat{\mathbf{a}} = \mathbf{B}^{-1}\mathbf{z}_B$ *if and only if* $|\mathbf{d}_j^T\mathbf{h}(\hat{\mathbf{a}})| \le 1\ \forall j \in I_B$.

II. *If there exists* $j_0 \in I_B$ *such that* $|\mathbf{d}_{j_0}^T\mathbf{h}(\hat{\mathbf{a}})| > 1$, *then*

$$F(\hat{\mathbf{a}} + \vartheta^* \mathbf{d}_{j_0}) < F(\hat{\mathbf{a}}),$$

where ϑ^* *is given by*

$$\vartheta^* = \operatorname*{med}_{i \in I_0 \cup J} (w_i, \rho_i) \neq 0, \quad w_i = |\mathbf{d}_{j_0}^T\mathbf{x}_i|, \quad \rho_i = \begin{cases} 0, & i \in I_0 \\ \frac{r_i(\hat{\mathbf{a}})}{\mathbf{d}_{j_0}^T\mathbf{x}_i}, & i \in J \end{cases},$$

where $J = \{i \in I \setminus I_B : \mathbf{d}_{j_0}^T\mathbf{x}_i \neq 0\}$.

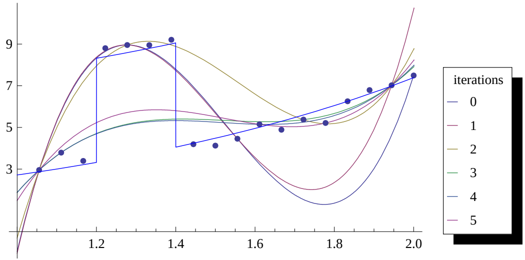
http://www.mathos.hr/seminar/Software.html

Example

For the function $f:[1,2]\to\mathbb{R}$, $f(x)=e^x+\begin{cases}5, & x\in(1.2,1.4)\\ 0, & x\notin(1.2,1.4)\end{cases}$ we will search for a best LAD-polynomial of the $(n-1)$-th degree $P_{n-1}(x)=\alpha_1+\alpha_2x+\cdots+\alpha_nx^{n-1}$ on the basis of given data (x_i,z_i), $i=1,\ldots,m$, where

$$x_i=1+\frac{i}{m},\qquad z_i=f(x_i)+\varepsilon_i,\quad \varepsilon_i\sim\mathcal{N}(0,\sigma^2),\quad \sigma=0.5.$$

For $n=4$ and $m=18$ the problem is reduced to searching for a best LAD-solution of the system $\mathbf{Xa}=\mathbf{z}$, where $\mathbf{X}_{ij}=x_i^{j-1}$, $i=1,\ldots,m$, $j=1,\ldots n$.



Clusters

- $\mathcal{A} = \{a_i \in \mathbb{R}^n : i = 1, \ldots, m\} \subset \mathbb{R}^n$
- $\Pi(\mathcal{A}) = \{\pi_1, \ldots, \pi_k\}$ – partition of the set $\mathcal{A}$, such that

$$\bigcup_{i=1}^{k} \pi_i = S \quad \pi_i \bigcap \pi_j = \emptyset, \qquad i \neq j \quad |\pi_j| \geq 1, \quad j = 1, \ldots, k$$

$|\mathcal{P}(\mathcal{A}, k)| = \frac{1}{k!} \sum_{j=1}^{k} (-1)^{k-j} \binom{k}{j} j^m$ – number of all partitions containing k clusters

$\lvert\mathcal{P}(\mathcal{A}, k)\rvert$	$k = 2$	$k = 3$	$k = 5$	$k = 8$	$k = 10$
$m = 10$	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}

Solving approaches: Algebraic and Combinatorial Optimization, Statistics, Graph Theory, Fuzzy Clustering

Clusters

- $\mathcal{A} = \{a_i \in \mathbb{R}^n : i = 1, \ldots, m\} \subset \mathbb{R}^n$
- $\Pi(\mathcal{A}) = \{\pi_1, \ldots, \pi_k\}$ – partition of the set $\mathcal{A}$, such that

$$\bigcup_{i=1}^{k} \pi_i = S \quad \pi_i \bigcap \pi_j = \emptyset, \qquad i \neq j \quad |\pi_j| \geq 1, \quad j = 1, \ldots, k$$

$|\mathcal{P}(\mathcal{A}, k)| = \frac{1}{k!} \sum_{j=1}^{k} (-1)^{k-j} \binom{k}{j} j^m$ – number of all partitions containing k clusters

$\lvert\mathcal{P}(\mathcal{A}, k)\rvert$	$k = 2$	$k = 3$	$k = 5$	$k = 8$	$k = 10$
$m = 10$	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}

Solving approaches: Algebraic and Combinatorial Optimization, Statistics, Graph Theory, Fuzzy Clustering

Clusters

- $\mathcal{A} = \{a_i \in \mathbb{R}^n : i = 1, \ldots, m\} \subset \mathbb{R}^n$
- $\Pi(\mathcal{A}) = \{\pi_1, \ldots, \pi_k\}$ – partition of the set $\mathcal{A}$, such that

$$\bigcup_{i=1}^{k} \pi_i = S \quad \pi_i \bigcap \pi_j = \emptyset, \qquad i \neq j \quad |\pi_j| \geq 1, \quad j = 1, \ldots, k$$

$|\mathcal{P}(\mathcal{A}, k)| = \frac{1}{k!} \sum_{j=1}^{k} (-1)^{k-j} \binom{k}{j} j^m$ – number of all partitions containing k clusters

$\lvert\mathcal{P}(\mathcal{A}, k)\rvert$	$k = 2$	$k = 3$	$k = 5$	$k = 8$	$k = 10$
$m = 10$	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}

Solving approaches: Algebraic and Combinatorial Optimization, Statistics, Graph Theory, Fuzzy Clustering

Applications

The applications are important and very attractive:

- biology, classification of the plough-lands according to fertility, classification of insects into groups
- information retrieval, machine learning, ranking of municipalities for financial support, business, facility location problem
- medicine, pattern recognition, text classification, understanding the Earth's climate
- education, psychology, and other social sciences

- K.Sabo, R. Scitovski, I.Vazler, M.Zekić-Sušac, Mathematical models of natural gas consumption, Energy Conversion and Management, doi:10.1016/j.enconman.2010.10.037
- G. Gan, C Ma, J. Wu, *Data Clustering: Theory, Algorithms, and Applications*, SIAM, Philadelphia, 2007.
- J. Kogan, *Introduction to Clustering Large and High-Dimensional Data*, Cambridge University Press, 2007.
- J. Kogan, C. Nicholas, M. Teboulle (eds), *Grouping Multidimensional Data: Recent Advances in Clustering*, Springer-Verlag, New York, 2006.

Applications

The applications are important and very attractive:

- biology, classification of the plough-lands according to fertility, classification of insects into groups
- information retrieval, machine learning, ranking of municipalities for financial support, business, facility location problem
- medicine, pattern recognition, text classification, understanding the Earth's climate
- education, psychology, and other social sciences

- K.Sabo, R. Scitovski, I.Vazler, M.Zekić-Sušac, Mathematical models of natural gas consumption, Energy Conversion and Management, doi:10.1016/j.enconman.2010.10.037
- G. Gan, C Ma, J. Wu, *Data Clustering: Theory, Algorithms, and Applications*, SIAM, Philadelphia, 2007.
- J. Kogan, *Introduction to Clustering Large and High-Dimensional Data*, Cambridge University Press, 2007.
- J. Kogan, C. Nicholas, M. Teboulle (eds), *Grouping Multidimensional Data: Recent Advances in Clustering*, Springer-Verlag, New York, 2006.

Centers of clusters

$d\colon \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty\rangle$ – some distance-like function

Applying the *minimal distance condition*, we can associate to each cluster $\pi_j \in \Pi$ its center c_j, defined by

$$c_j = c(\pi_j) := \underset{x \in \mathcal{C}_j}{\operatorname{argmin}} \sum_{a_i \in \pi_j} d(x, a_i), \qquad \mathcal{C}_j = \operatorname{conv}(\pi_j)$$

Objective function: $\mathcal{F}\colon \mathcal{P}(\mathcal{A}, k) \to [0, +\infty\rangle$,

$$\mathcal{F}(\Pi) = \sum_{j=1}^{k} \sum_{i=1}^{m} d(c_j, a_i),$$

Optimal partition $\Pi^{(o)}$: $\mathcal{F}(\Pi^{(o)}) = \min_{\Pi \in \mathcal{P}(\mathcal{A},k)} \mathcal{F}(\Pi)$

Centers of clusters

$d\colon \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty\rangle$ – some distance-like function

Applying the *minimal distance condition*, we can associate to each cluster $\pi_j \in \Pi$ its center c_j, defined by

$$c_j = c(\pi_j) := \operatorname*{argmin}_{x \in \mathcal{C}_j} \sum_{a_i \in \pi_j} d(x, a_i), \qquad \mathcal{C}_j = \operatorname{conv}(\pi_j)$$

Objective function: $\mathcal{F}\colon \mathcal{P}(\mathcal{A}, k) \to [0, +\infty\rangle$,

$$\mathcal{F}(\Pi) = \sum_{j=1}^{k} \sum_{i=1}^{m} d(c_j, a_i),$$

Optimal partition $\Pi^{(o)}$: $\mathcal{F}(\Pi^{(o)}) = \min_{\Pi \in \mathcal{P}(\mathcal{A}, k)} \mathcal{F}(\Pi)$

Centers of clusters

$d\colon \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty\rangle$ – some distance-like function

Applying the *minimal distance condition*, we can associate to each cluster $\pi_j \in \Pi$ its center c_j, defined by

$$c_j = c(\pi_j) := \operatorname*{argmin}_{x \in \mathcal{C}_j} \sum_{a_i \in \pi_j} d(x, a_i), \qquad \mathcal{C}_j = \operatorname{conv}(\pi_j)$$

Objective function: $\mathcal{F}\colon \mathcal{P}(\mathcal{A}, k) \to [0, +\infty\rangle$,

$$\mathcal{F}(\Pi) = \sum_{j=1}^{k} \sum_{i=1}^{m} d(c_j, a_i),$$

Optimal partition $\Pi^{(o)}$: $\mathcal{F}(\Pi^{(o)}) = \min_{\Pi \in \mathcal{P}(\mathcal{A},k)} \mathcal{F}(\Pi)$

Centers of clusters

$d\colon \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty\rangle$ – some distance-like function

Applying the *minimal distance condition*, we can associate to each cluster $\pi_j \in \Pi$ its center c_j, defined by

$$c_j = c(\pi_j) := \operatorname*{argmin}_{x \in \mathcal{C}_j} \sum_{a_i \in \pi_j} d(x, a_i), \qquad \mathcal{C}_j = \operatorname{conv}(\pi_j)$$

Objective function: $\mathcal{F}\colon \mathcal{P}(\mathcal{A}, k) \to [0, +\infty\rangle$,

$$\mathcal{F}(\Pi) = \sum_{j=1}^{k} \sum_{i=1}^{m} d(c_j, a_i),$$

Optimal partition $\Pi^{(o)}$: $\mathcal{F}(\Pi^{(o)}) = \min\limits_{\Pi \in \mathcal{P}(\mathcal{A}, k)} \mathcal{F}(\Pi)$

Centers of clusters

Conversely, for a given set of center-vectors $c_1, \ldots, c_k \in \mathbb{R}^n$, applying the minimal distance condition, we can define the partition $\Pi = \{\pi_1, \ldots, \pi_k\}$,

$$\pi_j = \{a \in \mathcal{A} : d(c_j, a) \le d(c_s, a),\ \forall s = 1, \ldots, k)\}, \qquad j = 1, \ldots, k.$$

Therefore, the problem of finding an optimal partition of the set $\mathcal{A}$ can be reduced to the optimization problem for the functional $F \colon \mathbb{R}^n \times \mathbb{R}^k \to [0, +\infty\rangle$,

$$\min_{c_1,\ldots,c_k \in \mathbb{R}^n} F(c_1, \ldots, c_k), \qquad F(c_1, \ldots, c_k) = \sum_{i=1}^{m} \min_{j=1,\ldots,k} d(c_j, a_i)$$

In general, this functional is not differentiable and may have several local minima.

Centers of clusters

Conversely, for a given set of center-vectors $c_1, \dots, c_k \in \mathbb{R}^n$, applying the minimal distance condition, we can define the partition $\Pi = \{\pi_1, \dots, \pi_k\}$,

$$\pi_j = \{a \in \mathcal{A} : d(c_j, a) \leq d(c_s, a),\ \forall s = 1, \dots, k)\}, \qquad j = 1, \dots, k.$$

Therefore, the problem of finding an optimal partition of the set $\mathcal{A}$ can be reduced to the optimization problem for the functional
$F \colon \mathbb{R}^n \times \mathbb{R}^k \to [0, +\infty\rangle$,

$$\min_{c_1, \dots, c_k \in \mathbb{R}^n} F(c_1, \dots, c_k), \qquad F(c_1, \dots, c_k) = \sum_{i=1}^{m} \min_{j=1,\dots,k} d(c_j, a_i)$$

In general, this functional is not differentiable and may have several local minima.

One-dimensional l_1-clustering problem

The set $\mathcal{A} = \{a_i \in \mathbb{R} : i = 1, \ldots, m\} \subset I \subset \mathbb{R}$, $I = [\alpha, \beta]$, has to be divided into k disjoint subsets $\pi_1, \ldots, \pi_k$, $1 \leq k \leq m$, satisfying

$$\bigcup_{i=1}^{k} \pi_i = \mathcal{A}, \qquad \pi_i \cap \pi_j = \emptyset, \quad i \neq j, \qquad |\pi_j| \geq 1, \quad j = 1, \ldots, k$$

$d(x, y) = |x - y|$ – l_1-distance function

We consider the nonconvex and nonsmooth optimization problem for the functional $\Phi : I^k \to [0, +\infty\rangle$,

$$\min_{c_1, \ldots, c_k \in I} \Phi(c_1, \ldots, c_k), \qquad \Phi(c_1, \ldots, c_k) = \sum_{i=1}^{m} \min_{j=1,\ldots,k} |c_j - a_i|,$$

One-dimensional l_1-clustering problem

The set $\mathcal{A} = \{a_i \in \mathbb{R} : i = 1, \ldots, m\} \subset I \subset \mathbb{R}$, $I = [\alpha, \beta]$, has to be divided into k disjoint subsets $\pi_1, \ldots, \pi_k$, $1 \le k \le m$, satisfying

$$\bigcup_{i=1}^{k} \pi_i = \mathcal{A}, \qquad \pi_i \cap \pi_j = \emptyset, \quad i \neq j, \qquad |\pi_j| \geq 1, \quad j = 1, \ldots, k$$

$d(x, y) = |x - y|$ – l_1-distance function

We consider the nonconvex and nonsmooth optimization problem for the functional $\Phi : I^k \to [0, +\infty\rangle$,

$$\min_{c_1,\ldots,c_k \in I} \Phi(c_1, \ldots, c_k), \qquad \Phi(c_1, \ldots, c_k) = \sum_{i=1}^{m} \min_{j=1,\ldots,k} |c_j - a_i|,$$

One-dimensional l_1-clustering problem

The set $\mathcal{A} = \{a_i \in \mathbb{R} : i = 1, \ldots, m\} \subset I \subset \mathbb{R}$, $I = [\alpha, \beta]$, has to be divided into k disjoint subsets $\pi_1, \ldots, \pi_k$, $1 \leq k \leq m$, satisfying

$$\bigcup_{i=1}^{k} \pi_i = \mathcal{A}, \qquad \pi_i \cap \pi_j = \emptyset, \quad i \neq j, \qquad |\pi_j| \geq 1, \quad j = 1, \ldots, k$$

$d(x, y) = |x - y|$ – l_1-distance function

We consider the nonconvex and nonsmooth optimization problem for the functional $\Phi : I^k \to [0, +\infty\rangle$,

$$\min_{c_1,\ldots,c_k \in I} \Phi(c_1, \ldots, c_k), \qquad \Phi(c_1, \ldots, c_k) = \sum_{i=1}^{m} \min_{j=1,\ldots,k} |c_j - a_i|,$$

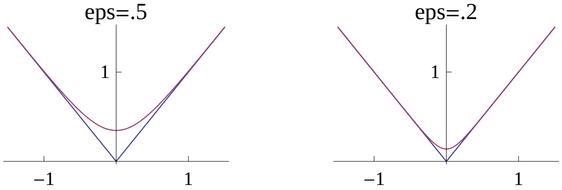
One-dimensional l_1-clustering problem

Note that the nondifferentiable function $f\colon \mathbb{R}^k \to \mathbb{R}$, $f(z) = \max\limits_{j=1,\ldots,k} z_j$ can be approximated by a differentiable function $f_\epsilon(z) = \epsilon \ln \sum\limits_{j=1}^{k} \exp\left(\frac{z_j}{\epsilon}\right)$

Example 1

$$|x| = \max\{-x, x\} = \lim_{\epsilon \to 0+} \epsilon \ln\left(e^{-\frac{x}{\epsilon}} + e^{\frac{x}{\epsilon}}\right) = \lim_{\epsilon \to 0+} \epsilon \ln 2 \operatorname{ch} \frac{x}{\epsilon},$$

$$|x| \approx \psi_\epsilon(x) = \epsilon \ln\left(e^{-\frac{x}{\epsilon}} + e^{\frac{x}{\epsilon}}\right) = \epsilon \ln 2 \operatorname{ch} \frac{x}{\epsilon}.$$



One-dimensional l_1-clustering problem

Instead of solving the optimization problem for the functional $\Phi : I^k \to [0, +\infty\rangle$,

$$\min_{c_1,\ldots,c_k \in I} \Phi(c_1, \ldots, c_k), \qquad \Phi(c_1, \ldots, c_k) = \sum_{i=1}^{m} \min_{j=1,\ldots,k} |c_j - a_i|,$$

we can solve the optimization problem for the functional $\Phi_\epsilon : I^k \to \mathbb{R}$

$$\min_{c_1,\ldots,c_k \in I} \Phi_\epsilon(c_1, \ldots, c_k), \qquad \Phi_\epsilon(c_1, \ldots, c_k) = -\epsilon \sum_{i=1}^{m} \ln \sum_{j=1}^{k} \exp\left(-\frac{|c_j - a_i|}{\epsilon}\right)$$

This is a continuous optimization problem, where the objective function does not have to be either convex or differentiable.

One-dimensional l_1-clustering problem

Instead of solving the optimization problem for the functional $\Phi : I^k \to [0, +\infty\rangle$,

$$\min_{c_1,\ldots,c_k \in I} \Phi(c_1, \ldots, c_k), \qquad \Phi(c_1, \ldots, c_k) = \sum_{i=1}^{m} \min_{j=1,\ldots,k} |c_j - a_i|,$$

we can solve the optimization problem for the functional $\Phi_\epsilon \colon I^k \to \mathbb{R}$

$$\min_{c_1,\ldots,c_k \in I} \Phi_\epsilon(c_1, \ldots, c_k), \qquad \Phi_\epsilon(c_1, \ldots, c_k) = -\epsilon \sum_{i=1}^{m} \ln \sum_{j=1}^{k} \exp\left(-\frac{|c_j - a_i|}{\epsilon}\right)$$

This is a continuous optimization problem, where the objective function does not have to be either convex or differentiable.

Properties of the functional Φ_ϵ

Theorem 3

Let $\mathcal{A} = \{a_i \in \mathbb{R} : i = 1, \dots, m\} \subset I \subset \mathbb{R}$, $I = [\alpha, \beta]$, be a given set and let $\epsilon > 0$. Then for all $\theta = (c_1, \dots, c_k)^T \in I^k$, the following inequalities hold

$$0 < \epsilon\, m \ln\left(1 + (k-1)e^{-\frac{1}{\epsilon}(\beta-\alpha)}\right) \le \Phi(\theta) - \Phi_\epsilon(\theta) \le \epsilon\, m \ln k. \qquad (1)$$

The functional Φ_ϵ is continuous, and according to Theorem 3, it is bounded below,

$$\Phi_\epsilon(\theta) \ge \Phi(\theta) - \epsilon\, m \ln k \ge -\epsilon\, m \ln k.$$

Therefore, since $I^k \subset \mathbb{R}^k$ is compact, Φ_ϵ attains its global minimum.

Lemma 4

For all $\theta_1, \theta_2 \in I^k$ there holds

$$|\Phi_\epsilon(\theta_2) - \Phi_\epsilon(\theta_1)| \le m\, \|\theta_2 - \theta_1\|_\infty. \qquad (2)$$

Properties of the functional Φ_ϵ

Theorem 3

Let $\mathcal{A} = \{a_i \in \mathbb{R} : i = 1, \ldots, m\} \subset I \subset \mathbb{R}$, $I = [\alpha, \beta]$, be a given set and let $\epsilon > 0$. Then for all $\theta = (c_1, \ldots, c_k)^T \in I^k$, the following inequalities hold

$$0 < \epsilon\, m \ln\left(1 + (k-1)e^{-\frac{1}{\epsilon}(\beta-\alpha)}\right) \le \Phi(\theta) - \Phi_\epsilon(\theta) \le \epsilon\, m \ln k. \quad (1)$$

The functional Φ_ϵ is continuous, and according to Theorem 3, it is bounded below,

$$\Phi_\epsilon(\theta) \ge \Phi(\theta) - \epsilon\, m \ln k \ge -\epsilon\, m \ln k.$$

Therefore, since $I^k \subset \mathbb{R}^k$ is compact, Φ_ϵ attains its global minimum.

Lemma 4

For all $\theta_1, \theta_2 \in I^k$ there holds

$$|\Phi_\epsilon(\theta_2) - \Phi_\epsilon(\theta_1)| \le m\, \|\theta_2 - \theta_1\|_\infty. \quad (2)$$

Properties of the functional Φ_ϵ

Theorem 3

Let $\mathcal{A} = \{a_i \in \mathbb{R} : i = 1, \dots, m\} \subset I \subset \mathbb{R}$, $I = [\alpha, \beta]$, be a given set and let $\epsilon > 0$. Then for all $\theta = (c_1, \dots, c_k)^T \in I^k$, the following inequalities hold

$$0 < \epsilon\, m \ln\left(1 + (k-1)e^{-\frac{1}{\epsilon}(\beta-\alpha)}\right) \le \Phi(\theta) - \Phi_\epsilon(\theta) \le \epsilon\, m \ln k. \quad (1)$$

The functional Φ_ϵ is continuous, and according to Theorem 3, it is bounded below,

$$\Phi_\epsilon(\theta) \ge \Phi(\theta) - \epsilon\, m \ln k \ge -\epsilon\, m \ln k.$$

Therefore, since $I^k \subset \mathbb{R}^k$ is compact, Φ_ϵ attains its global minimum.

Lemma 4

For all $\theta_1, \theta_2 \in I^k$ there holds

$$|\Phi_\epsilon(\theta_2) - \Phi_\epsilon(\theta_1)| \le m\, \|\theta_2 - \theta_1\|_\infty. \quad (2)$$

A method for finding stationary points

Assuming that $\theta^{(t)} = (c_1^{(t)}, \ldots, c_k^{(t)}) \in I^k$ is known, we are going to look for the next iteration $\theta^{(t+1)} = (c_1^{(t+1)}, \ldots, c_k^{(t+1)})$ as the weighted median of real numbers $a_1, \ldots, a_m$

$$c_s^{(t+1)} = \operatorname*{med}_{i=1,\ldots,m} \left(w_i^{(s)}(\theta), a_i \right), \qquad s = 1, \ldots, k,$$

where $w_i^{(s)}(\theta) = \dfrac{\exp\left(-\frac{|c_s - a_i|}{\epsilon}\right)}{\sum\limits_{j=1}^{k} \exp\left(-\frac{|c_j - a_i|}{\epsilon}\right)}, \quad i = 1, \ldots, m, \quad s = 1, \ldots, k.$

This means that the next iteration $\theta^{(t+1)}$ is obtained as the solution of a sequence of one-dimensional weighted median problems

$$c_s^{(t+1)} = \operatorname*{argmin}_{\zeta \in \mathbb{R}} g_s(\zeta), \quad g_s(\zeta) = \sum_{i=1}^{m} w_i^{(s)}(\theta^{(t)}) |\zeta - a_i|, \quad s = 1, \ldots, k.$$

Weighted Median Problem

If we define the function $g(\,\cdot\,;\theta^{(t)})\colon I^k \to [0,+\infty\rangle$ by

$$g(\theta;\theta^{(t)}) = \sum_{s=1}^{k}\sum_{i=1}^{m} w_i^{(s)}(\theta^{(t)})|c_s - a_i|, \qquad \theta = (c_1,\dots,c_k), \tag{3}$$

then the iterative procedure can be written as

$$\theta^{(t+1)} = \operatorname*{argmin}_{\theta\in I^k} g(\theta;\theta^{(t)}). \tag{4}$$

Note that

- $\theta^{(t)} \in \mathcal{A}^k$, for all $t\in\mathbb{N}$;
- Because of symmetry properties of Φ and Φ_ϵ, if $\theta^\star = (c_1^\star,\dots,c_k^\star)$ minimizes the functional Φ_ϵ, and $\tilde{\theta}$ is an arbitrary permutation of $\theta^\star$, then also $\tilde{\theta}$ minimizes Φ_ϵ.
- For comparison, the same problem in (Kogan, 2007) is solved in the Least Squares sense by minimizing the functional ϕ_ϵ with $d(x,y)=(x-y)^2$. In this case the objective function is differentiable, and corresponding sequence $(\theta^{(t)})$ is obtained as a weighted arithmetic mean of the data.

Weighted Median Problem

If we define the function $g(\,\cdot\,;\theta^{(t)})\colon I^k \to [0,+\infty\rangle$ by

$$g(\theta;\theta^{(t)}) = \sum_{s=1}^{k}\sum_{i=1}^{m} w_i^{(s)}(\theta^{(t)})|c_s - a_i|, \qquad \theta = (c_1,\ldots,c_k), \tag{3}$$

then the iterative procedure can be written as

$$\theta^{(t+1)} = \operatorname*{argmin}_{\theta\in I^k} g(\theta;\theta^{(t)}). \tag{4}$$

Note that

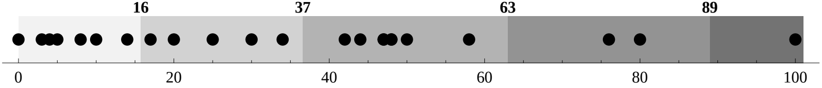
- $\theta^{(t)} \in \mathcal{A}^k$, for all $t \in \mathbb{N}$;
- Because of symmetry properties of Φ and Φ_ϵ, if $\theta^\star = (c_1^\star,\ldots,c_k^\star)$ minimizes the functional Φ_ϵ, and $\tilde{\theta}$ is an arbitrary permutation of $\theta^\star$, then also $\tilde{\theta}$ minimizes Φ_ϵ.
- For comparison, the same problem in (Kogan, 2007) is solved in the Least Squares sense by minimizing the functional ϕ_ϵ with $d(x,y) = (x-y)^2$. In this case the objective function is differentiable, and corresponding sequence $\big(\theta^{(t)}\big)$ is obtained as a weighted arithmetic mean of the data.

Ocjenjivanje studentskih kolokvija

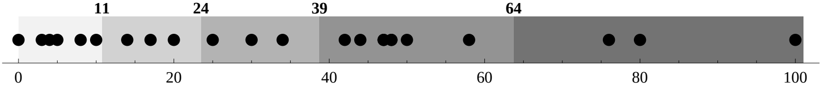
Bodovi postignuti na kolokviju za grupu od 23 studenta

$\mathcal{A} = \{0, 3, 4, 5, 8, 10, 14, 17, 20, 25, 30, 34, 42, 44, 47, 47, 48, 48, \; 50, 58, 76, 80, 100\}.$

LS-optimalna particija



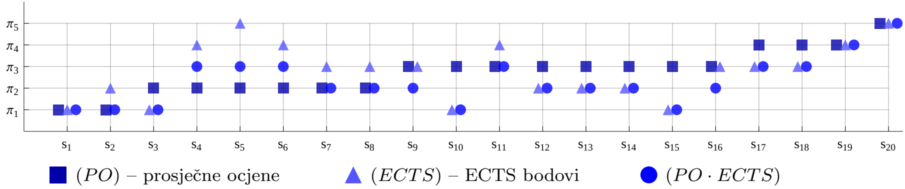
LAD-optimalna particija



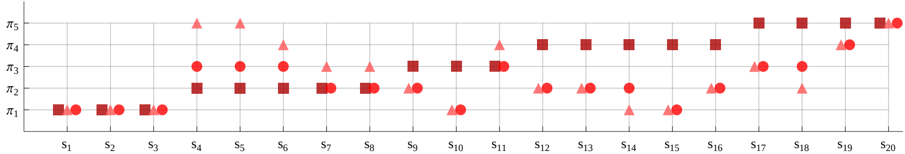
Prosječne ocjene (PO) i ECTS bodovi studenata

Student	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
PO	2.0	2.4	2.7	3.0	3.1	3.1	3.2	3.3	3.5	3.6
ECTS	45	48	47	57	60	55	52	51	50	47
Student	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}	s_{16}	s_{17}	s_{18}	s_{19}	s_{20}
PO	3.7	3.8	3.9	3.9	4.0	4.0	4.3	4.5	4.5	5.0
ECTS	54	49	49	48	46	50	51	50	54	60

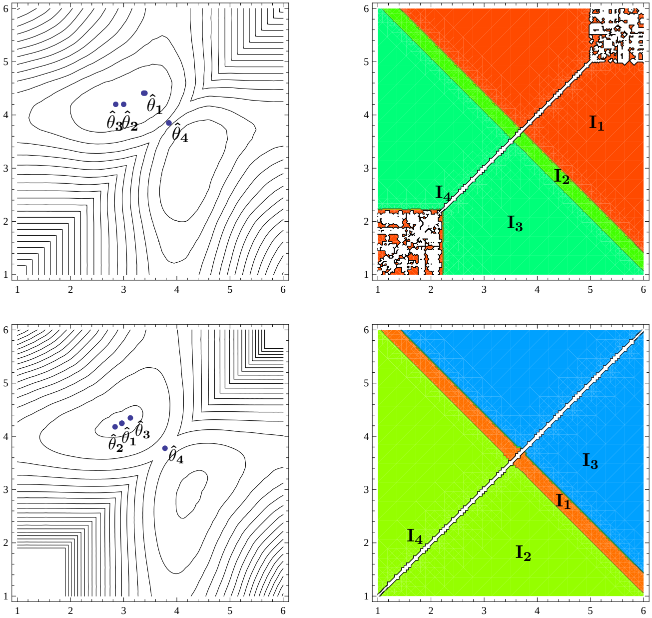
a) LS-optimalne particije



b) LAD-optimalne particije



PO studenata – dva klastera prema LAD i LS principu



Properties of the iterative process

Proposition 1

(i) *For every* $i = 1, \dots, m$ *and an arbitrary* $\theta \in \mathbb{R}^k$*, the sequence of weights* $w_i^{(s)}(\theta)$*,* $s = 1, \dots, k$*, satisfies*

$$0 < w_i^{(s)}(\theta) < 1, \qquad \sum_{s=1}^{k} w_i^{(s)}(\theta) = 1.$$

(ii) *For an arbitrary* $\theta^{(0)} \in \mathbb{R}^k$*, the sequence* $\left(\theta^{(t)}\right)$*, defined by iterative process* (4)*, remains in* I^k*, and hence it is bounded.*

Proposition 2

Let $\theta^{(0)} \in \mathbb{R}^k$ *be an arbitrary point, and let the sequence* $\left(\theta^{(t)}\right)$ *be given by the iterative process* (4)*.*

If $\theta^{(t+1)} \neq \theta^{(t)}$*, then* $\Phi_\epsilon(\theta^{(t+1)}) < \Phi_\epsilon(\theta^{(t)})$*.*

Convergence of the iterative process

Theorem 4

Let $\theta^{(0)} \in \mathbb{R}^k$ *be an arbitrary point, and let the sequence* $\left(\theta^{(t)}\right)$ *be defined by iterative process* (4)*. Then*

- (*i*) *The sequence* $\left(\theta^{(t)}\right)$ *has an accumulation point.*
- (*ii*) *The sequence* $\left(\Phi_\epsilon^{(t)}\right)$*, where* $\Phi_\epsilon^{(t)} := \Phi_\epsilon(\theta^{(t)})$*, converges.*
- (*iii*) *Every accumulation point* $\hat{\theta}$ *of the sequence* $\left(\theta^{(t)}\right)$ *is a stationary point of the functional* Φ_ϵ*, and it is obtained by iterative process* (4) *in finitely many steps, i.e. there exists a* $\mu \in \mathbb{N}$*, such that* $\theta^{(\mu+1)} = \theta^{(\mu)} = \hat{\theta}$*.*
- (*iv*) *If* $\hat{\theta}_1$ *and* $\hat{\theta}_2$ *are two accumulation points of the sequence* $\left(\theta^{(t)}\right)$*, then* $\Phi_\epsilon(\hat{\theta}_1) = \Phi_\epsilon(\hat{\theta}_2)$*.*

Algorithm: One-dimensional l_1-clustering

Step 1: Input $m \geq 1$, $1 \leq k \leq m$, $\epsilon > 0$, $\mathcal{A} = \{a_i \in \mathbb{R} : i = 1, \dots, m\}$, and choose an initial vector of centers $\theta^{(0)} = (c_1^{(0)}, \dots, c_k^{(0)})^T$;

Step 2: For all $s = 1, \dots, k$ define vectors $w^{(s)}$ with components

$$w_i^{(s)} = \frac{\exp\left(-\frac{|c_s^{(0)} - a_i|}{\epsilon}\right)}{\sum\limits_{j=1}^{k} \exp\left(-\frac{|c_j^{(0)} - a_i|}{\epsilon}\right)}, \quad i = 1, \dots, m;$$

Step 3: For all $s = 1, \dots, k$ solve the weighted median problem[1]

$$\mathrm{med}(w^{(s)}, a) =: c_s^{(1)},$$

and set $\theta^{(1)} = (c_1^{(1)}, \dots, c_k^{(1)})^T$;

Step 4: If $\theta^{(1)} = \theta^{(0)}$, set $\theta^{(0)} = \theta^{(1)}$ and go to Step 2; Else Go To Step 5;

Step 5: According to the minimal distance principle, define a partition $\Pi = \{\pi_1, \dots, \pi_k\}$ with centers $c_1^{(1)}, \dots, c_k^{(1)}$:

$$\pi_1 = \{a_i \in \mathcal{A} : |a_i - c_1^{(1)}| \leq |a_i - c_l^{(1)}|, \, l = 1, \dots, k\},$$

$$\pi_j = \{a_i \in \mathcal{A} \setminus \bigcup_{s=1}^{j-1} \pi_s : |a_i - c_j^{(1)}| \leq |a_i - c_l^{(1)}|, \, \forall\, l = 1, \dots, k\}, \quad j = 2, \dots, k.$$

Algorithm: One-dimensional l_1-clustering

Step 1: Input $m \geq 1$, $1 \leq k \leq m$, $\epsilon > 0$, $\mathcal{A} = \{a_i \in \mathbb{R} : i = 1, \ldots, m\}$, and choose an initial vector of centers $\theta^{(0)} = (c_1^{(0)}, \ldots, c_k^{(0)})^T$;

Step 2: For all $s = 1, \ldots, k$ define vectors $w^{(s)}$ with components

$$w_i^{(s)} = \frac{\exp\left(-\frac{|c_s^{(0)} - a_i|}{\epsilon}\right)}{\sum\limits_{j=1}^{k} \exp\left(-\frac{|c_j^{(0)} - a_i|}{\epsilon}\right)}, \quad i = 1, \ldots, m;$$

Step 3: For all $s = 1, \ldots, k$ solve the weighted median problem[1]

$$\mathrm{med}(w^{(s)}, a) =: c_s^{(1)},$$

and set $\theta^{(1)} = (c_1^{(1)}, \ldots, c_k^{(1)})^T$;

Step 4: If $\theta^{(1)} = \theta^{(0)}$, set $\theta^{(0)} = \theta^{(1)}$ and go to Step 2; Else Go To Step 5;

Step 5: According to the minimal distance principle, define a partition $\Pi = \{\pi_1, \ldots, \pi_k\}$ with centers $c_1^{(1)}, \ldots, c_k^{(1)}$:

$$\pi_1 = \{a_i \in \mathcal{A} : |a_i - c_1^{(1)}| \leq |a_i - c_l^{(1)}|, \, l = 1, \ldots, k\},$$

$$\pi_j = \{a_i \in \mathcal{A} \setminus \bigcup_{s=1}^{j-1} \pi_s : |a_i - c_j^{(1)}| \leq |a_i - c_l^{(1)}|, \, \forall\, l = 1, \ldots, k\}, \quad j = 2, \ldots, k.$$

Algorithm: One-dimensional l_1-clustering

Step 1: Input $m \geq 1$, $1 \leq k \leq m$, $\epsilon > 0$, $\mathcal{A} = \{a_i \in \mathbb{R} : i = 1, \dots, m\}$, and choose an initial vector of centers $\theta^{(0)} = (c_1^{(0)}, \dots, c_k^{(0)})^T$;

Step 2: For all $s = 1, \dots, k$ define vectors $w^{(s)}$ with components

$$w_i^{(s)} = \frac{\exp\left(-\frac{|c_s^{(0)} - a_i|}{\epsilon}\right)}{\sum_{j=1}^{k} \exp\left(-\frac{|c_j^{(0)} - a_i|}{\epsilon}\right)}, \quad i = 1, \dots, m;$$

Step 3: For all $s = 1, \dots, k$ solve the weighted median problem[1]

$$\mathrm{med}(w^{(s)}, a) =: c_s^{(1)},$$

and set $\theta^{(1)} = (c_1^{(1)}, \dots, c_k^{(1)})^T$;

Step 4: If $\theta^{(1)} = \theta^{(0)}$, set $\theta^{(0)} = \theta^{(1)}$ and go to Step 2; Else Go To Step 5;

Step 5: According to the minimal distance principle, define a partition $\Pi = \{\pi_1, \dots, \pi_k\}$ with centers $c_1^{(1)}, \dots, c_k^{(1)}$:

$$\pi_1 = \{a_i \in \mathcal{A} : |a_i - c_1^{(1)}| \leq |a_i - c_l^{(1)}|, \, l = 1, \dots, k\},$$

$$\pi_j = \{a_i \in \mathcal{A} \setminus \bigcup_{s=1}^{j-1} \pi_s : |a_i - c_j^{(1)}| \leq |a_i - c_l^{(1)}|, \, \forall\, l = 1, \dots, k\}, \quad j = 2, \dots, k.$$

Algorithm: One-dimensional l_1-clustering

Step 1: Input $m \geq 1$, $1 \leq k \leq m$, $\epsilon > 0$, $\mathcal{A} = \{a_i \in \mathbb{R} : i = 1, \ldots, m\}$, and choose an initial vector of centers $\theta^{(0)} = (c_1^{(0)}, \ldots, c_k^{(0)})^T$;

Step 2: For all $s = 1, \ldots, k$ define vectors $w^{(s)}$ with components

$$w_i^{(s)} = \frac{\exp\left(-\frac{|c_s^{(0)} - a_i|}{\epsilon}\right)}{\sum_{j=1}^{k} \exp\left(-\frac{|c_j^{(0)} - a_i|}{\epsilon}\right)}, \quad i = 1, \ldots, m;$$

Step 3: For all $s = 1, \ldots, k$ solve the weighted median problem[1]

$$\mathrm{med}(w^{(s)}, a) =: c_s^{(1)},$$

and set $\theta^{(1)} = (c_1^{(1)}, \ldots, c_k^{(1)})^T$;

Step 4: If $\theta^{(1)} = \theta^{(0)}$, set $\theta^{(0)} = \theta^{(1)}$ and go to Step 2; Else Go To Step 5;

Step 5: According to the minimal distance principle, define a partition $\Pi = \{\pi_1, \ldots, \pi_k\}$ with centers $c_1^{(1)}, \ldots, c_k^{(1)}$:

$$\pi_1 = \{a_i \in \mathcal{A} : |a_i - c_1^{(1)}| \leq |a_i - c_l^{(1)}|,\ l = 1, \ldots, k\},$$

$$\pi_j = \{a_i \in \mathcal{A} \setminus \bigcup_{s=1}^{j-1} \pi_s : |a_i - c_j^{(1)}| \leq |a_i - c_l^{(1)}|,\ \forall\, l = 1, \ldots, k\}, \quad j = 2, \ldots, k.$$

Algorithm: One-dimensional l_1-clustering

Step 1: Input $m \geq 1$, $1 \leq k \leq m$, $\epsilon > 0$, $\mathcal{A} = \{a_i \in \mathbb{R} : i = 1, \dots, m\}$, and choose an initial vector of centers $\theta^{(0)} = (c_1^{(0)}, \dots, c_k^{(0)})^T$;

Step 2: For all $s = 1, \dots, k$ define vectors $w^{(s)}$ with components

$$w_i^{(s)} = \frac{\exp\left(-\frac{|c_s^{(0)} - a_i|}{\epsilon}\right)}{\sum_{j=1}^{k} \exp\left(-\frac{|c_j^{(0)} - a_i|}{\epsilon}\right)}, \quad i = 1, \dots, m;$$

Step 3: For all $s = 1, \dots, k$ solve the weighted median problem[1]

$$\mathrm{med}(w^{(s)}, a) =: c_s^{(1)},$$

and set $\theta^{(1)} = (c_1^{(1)}, \dots, c_k^{(1)})^T$;

Step 4: If $\theta^{(1)} = \theta^{(0)}$, set $\theta^{(0)} = \theta^{(1)}$ and go to Step 2; Else Go To Step 5;

Step 5: According to the minimal distance principle, define a partition $\Pi = \{\pi_1, \dots, \pi_k\}$ with centers $c_1^{(1)}, \dots, c_k^{(1)}$:

$$\pi_1 = \{a_i \in \mathcal{A} : |a_i - c_1^{(1)}| \leq |a_i - c_l^{(1)}|, \, l = 1, \dots, k\},$$

$$\pi_j = \{a_i \in \mathcal{A} \setminus \bigcup_{s=1}^{j-1} \pi_s : |a_i - c_j^{(1)}| \leq |a_i - c_l^{(1)}|, \, \forall\, l = 1, \dots, k\}, \quad j = 2, \dots, k.$$

Algorithm: One-dimensional l_1-clustering

Step 1: Input $m \geq 1$, $1 \leq k \leq m$, $\epsilon > 0$, $\mathcal{A} = \{a_i \in \mathbb{R} : i = 1, \ldots, m\}$, and choose an initial vector of centers $\theta^{(0)} = (c_1^{(0)}, \ldots, c_k^{(0)})^T$;

Step 2: For all $s = 1, \ldots, k$ define vectors $w^{(s)}$ with components

$$w_i^{(s)} = \frac{\exp\left(-\frac{|c_s^{(0)} - a_i|}{\epsilon}\right)}{\sum_{j=1}^{k} \exp\left(-\frac{|c_j^{(0)} - a_i|}{\epsilon}\right)}, \quad i = 1, \ldots, m;$$

Step 3: For all $s = 1, \ldots, k$ solve the weighted median problem[1]

$$\mathrm{med}(w^{(s)}, a) =: c_s^{(1)},$$

and set $\theta^{(1)} = (c_1^{(1)}, \ldots, c_k^{(1)})^T$;

Step 4: If $\theta^{(1)} = \theta^{(0)}$, set $\theta^{(0)} = \theta^{(1)}$ and go to Step 2; Else Go To Step 5;

Step 5: According to the minimal distance principle, define a partition $\Pi = \{\pi_1, \ldots, \pi_k\}$ with centers $c_1^{(1)}, \ldots, c_k^{(1)}$:

$$\pi_1 = \{a_i \in \mathcal{A} : |a_i - c_1^{(1)}| \leq |a_i - c_l^{(1)}|, \ l = 1, \ldots, k\},$$

$$\pi_j = \{a_i \in \mathcal{A} \setminus \bigcup_{s=1}^{j-1} \pi_s : |a_i - c_j^{(1)}| \leq |a_i - c_l^{(1)}|, \ \forall\, l = 1, \ldots, k\}, \quad j = 2, \ldots, k.$$

Clustering High-Dimensional Data

$$\mathcal{A} = \{a_i \in \mathbb{R}^n : i = 1, \dots, m\} \subset \mathbb{R}^n \qquad \Pi(\mathcal{A}) = \{\pi_1, \dots, \pi_k\}$$

$$c_j = c(\pi_j) := \operatorname*{argmin}_{x \in \mathcal{C}_j} \sum_{a_i \in \pi_j} \|x - a_i\|_1, \qquad \mathcal{C}_j = \operatorname{conv}(\pi_j)$$

$$\Phi\colon R^k \times \mathbb{R}^n \to [0, +\infty\rangle, \quad \Phi(c_1, \dots, c_k) = \sum_{i=1}^{m} \min_{j=1,k} \|c_j - a_i\|_1$$

$$\Phi_\epsilon\colon R^k \times \mathbb{R}^n \to [0, +\infty\rangle, \quad \Phi_\epsilon(c_1, \dots, c_k) = -\epsilon \sum_{i=1}^{m} \ln \sum_{j=1}^{k} \exp\left(-\frac{\|c_j - a_i\|_1}{\epsilon}\right)$$

$$\theta^{(t+1)} = \operatorname*{argmin}_{\theta \in \mathcal{C}^k} g(\theta; \theta^{(t)}),$$

$$g(\theta; \theta^{(t)}) = \sum_{s=1}^{k} \sum_{i=1}^{m} w_i^{(s)}(\theta^{(t)}) \|c_s - a_i\|_1,$$

$$w_i^{(s)}(\theta) = \frac{\exp\left(-\frac{\|c_s - a_i\|_1}{\epsilon}\right)}{\sum_{j=1}^{k} \exp\left(-\frac{\|c_j - a_i\|_1}{\epsilon}\right)}, \quad i = 1, \dots, m, \quad s = 1, \dots, k.$$