

Kvazimetičke funkcije

Department of Mathematics, University of Osijek

1 Problem grupiranja podataka-Klasteri

Razmatramo problem grupiranja elemenata nekog skupa $\mathcal{A}$ s $m \geq 2$ elemenata u disjunktne podskupove $\pi_1, \dots, \pi_k$, $1 \leq k \leq m$, takve da vrijedi

$$\bigcup_{i=1}^k \pi_i = \mathcal{A}, \quad \pi_i \cap \pi_j = \emptyset, \quad i \neq j, \quad m_j := |\pi_j| \geq 1, \quad j = 1, \dots, k, \quad (1)$$

na osnovi jednog ili više obilježja uz korištenje raznih kriterijskih funkcija cilja. Rastave skupa $\mathcal{A}$ na podskupove $\pi_1, \dots, \pi_k$ koji zadovoljavaju (1) označavat ćemo s $\Pi = \{\pi_1, \dots, \pi_k\}$ i zvati *particija skupa* $\mathcal{A}$, a skupove $\pi_1, \dots, \pi_k$ zvat ćemo *klasteri*. Skup svih particija skupa $\mathcal{A}$ sastavljenih od k klastera koje zadovoljavaju (1) označit ćemo s $\mathcal{P}(\mathcal{A}, k)$. Nadalje, kad god budemo govorili o particiji skupa $\mathcal{A}$, podrazumijevat ćemo da je ona sastavljena od podskupova skupa $\mathcal{A}$, koji zadovoljavaju (1). Na taj način svjesno smo iz razmatranja isključili particije, koje sadržavaju prazan skup ili skup $\mathcal{A}$.

Može se pokazati (Veljan, 2001) da je broj svih particija skupa $\mathcal{A}$, koje zadovoljavaju (1) jednak Stirlingovom broju druge vrste

$$|\mathcal{P}(\mathcal{A}, k)| = \frac{1}{k!} \sum_{j=1}^k (-1)^{k-j} \binom{k}{j} j^m. \quad (2)$$

Primjer 1. Broj svih particija skupa $\mathcal{A}$ određenih s (1) specijalno za $m = 10, 50, 10^3, 10^6$ i $k = 2, 3, 5, 8, 10$ iznosi

$ \mathcal{P}(\mathcal{A}, 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}

Iz navedenog primjera vidi se da traženje optimalne particije općenito neće biti moguće provesti pretraživanjem čitavog skupa $\mathcal{P}(\mathcal{A}, k)$. Odmah teba reći da problem traženja optimalne particije spada u NP-teške probleme (vidi (Gan et al., 2007)) nekonveksne optimizacije općenito nediferencijabilne funkcije više varijabli, koja najčešće posjeduje značajan broj stacionarnih točaka. U znanstvenoj i stručnoj literaturi ovaj problem često nalazimo pod nazivom *cluster analysis*.

Namjera nam je na što jednostavniji način približiti ovo znanstveno područje što širem krugu stručnjaka jer problem klasifikacije i rangiranja podataka u posljednje vrijeme postaje sve interesantnije područje interesa raznim znanstvenicima i stručnjacima, ali također i donositeljima raznih odluka, primjerice u tijelima državne i lokalne administracije.

Postoji više međunarodno poznatih i priznatih specijaliziranih časopisa koji prate ovo područje, primjerice: *Clustering and Classification*, *Pattern Recognition*, *Journal of Classification*, *Journal of Machine Learning Research*. Također, redovito se održavaju specijalizirani znanstveni i stručni skupovi s ovom problematikom, primjerice: *SIAM International Conference on Data Mining*, *IEEE International Conference on Data Mining*, *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, a na brojnim međunarodnim pretraživačima i bazama podataka postoji ogroman broj podataka o ovom području.

U novijoj znanstvenoj i stručnoj literaturi mogu se pronaći brojne primjene u poljoprivredi (primjerice, razvrstavanje oranica prema plodnosti zemljišta), u biologiji (primjerice, klasifikacija kukaca u grupe), u medicini (Everitt et al., 2001; Littau, 2006), u prometu (Iyigun, 2007), kod klasifikacije teksta i kompjuterskih pretraživača baza podataka (Kogan et al., 2007; Su and Kogan, 2010), kod razumijevanja klimatskih kretanja i problema lokacije objekata (Divéki and Csanád, 2010; Gan et al., 2007), kod satne prognoze potrošnje prirodnog plina i drugih energenata (Sabo et al., 2011). Također, često se pojavljuju razne primjene u upravljanju (primjerice, rangiranje gradova i općina za potrebe financijske podrške) (Hudec and Vujošević, 2010), u poslovanju (Everitt et al., 2001), u društvenim znanostima i psihologiji, kod definiranja izbornih sustava (Dhillon et al., 2004; Kogan, 2007) itd.

Definicija 1. Neka je $d: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+$ funkcija, koja zadovoljava svojstva

- 1) $d(x, y) \geq 0$, $x, y \in \mathbb{R}^n$ i $d(x, y) = 0 \Leftrightarrow x = y$.
- 2) Za svaki $y \in \mathbb{R}^n$, funkcija $x \mapsto d(x, y)$ je konveksna na $\mathbb{R}^n$.
- 3) Za svaki $y \in \mathbb{R}^n$ vrijedi $\lim_{\|x\| \rightarrow \infty} d(x, y) = +\infty$.

Funkciju d sa tim svojstvom zovemo *kvazimetrička funkcija* na $\mathbb{R}^n$.

Pretpostavimo da je zadana kvazimetrička funkcija d na $\mathbb{R}^n \times \mathbb{R}^n$ te da je $\pi_j \in \Pi$ klaster. Promatrajmo sljedeći minimizacijski problem

$$\sum_{a_i \in \pi_j} d(x, a_i) \rightarrow \min_{x \in \mathbb{R}^n}. \quad (3)$$

Pokažimo da problem (3) ima rješenje.

Kako je prema svojstvu 1) postoji

$$I_0 := \inf_{x \in \mathbb{R}^n} \sum_{a_i \in \pi_j} d(x, a_i) \geq 0.$$

Sukladno Zadatku 6 iz domaće zadaće, zbog neprekidnosti funkcije $x \mapsto \sum_{a_i \in \pi_j} d(x, a_i)$ slijedi da postoji niz $(x_r) \in \mathbb{R}^n$, takav da je

$$\lim_{r \rightarrow \infty} \sum_{a_i \in \pi_j} d(x_r, a_i) = I_0. \quad (4)$$

Pokažimo da je niz (x_r) omeđen. Pretpostavimo suprotno, odnosno da je (x_r) takav da je $\lim_{r \rightarrow \infty} \|x_r\| \rightarrow \infty$. Prema svojstvu 3) funkcije $x \mapsto d(x, y)$ slijedi da je

$$\lim_{r \rightarrow \infty} \sum_{a_i \in \pi_j} d(x_r, a_i) = +\infty,$$

sto je u suprotnosti s (4). Prema tome niz (x_r) je omeđen te ima konvergentni podniz (x_{r_s}) takav da je $\lim_{s \rightarrow \infty} x_{r_s} = x^*$, odakle je

$$I_0 = \inf_{x \in \mathbb{R}^n} \sum_{a_i \in \pi_j} d(x, a_i) = \lim_{s \rightarrow \infty} \sum_{a_i \in \pi_j} d(x_{r_s}, a_i) = \sum_{a_i \in \pi_j} d(x^*, a_i),$$

odnosno problem (3) ima rješenje. Primijetimo dodatno da ako je funkcije $x \mapsto d(x, y)$ strogo konveksna onda problem (3) ima jedinstveno rješenje.

Ako je d kvazimetrička funkcija te $\pi_j \in \Pi$ klaster, onda odgovarajuće rješenje problema (3) zovemo **centar klastera** te označavamo s $c_j = c_j(\pi_j)$.

Nadalje, ako na skupu svih particija $\mathcal{P}(\mathcal{A}, k)$ skupa $\mathcal{A}$ sastavljenih od k klastera definiramo kriterijsku funkciju cilja $\mathcal{F} : \mathcal{P}(\mathcal{A}, k) \rightarrow \mathbb{R}_+$,

$$\mathcal{F}(\Pi) = \sum_{j=1}^k \sum_{a_i \in \pi_j} d(c_j, a_i), \quad (5)$$

onda d -optimalnu particiju Π^* tražimo rješavanjem sljedećeg optimizacijskog problema

$$\mathcal{F}(\Pi^*) = \min_{\Pi \in \mathcal{P}(\mathcal{A}, k)} \mathcal{F}(\Pi). \quad (6)$$

Primijetite da na taj način optimalna particija Π^* ima svojstvo da je suma “rasipanja” (suma odstupanja) elemenata klastera oko svog centra minimalna. Na taj način nastojimo postići što bolju unutrašnju kompaktnost i separiranost klastera.

Obrnuto, za dani skup centara $c_1, \dots, c_k \in \mathbb{R}^n$, uz primjenu *principa minimalnih udaljenosti* možemo definirati particiju $\Pi = \{\pi_1, \dots, \pi_k\}$ skupa $\mathcal{A}$ na sljedeći način:

$$\pi_j = \{a \in \mathcal{A} : d(c_j, a) \leq d(c_s, a), \forall s = 1, \dots, k\}, \quad j = 1, \dots, k, \quad (7)$$

pri čemu treba voditi računa o tome da svaki element skupa $\mathcal{A}$ pripadne samo jednom klasteru. Primijetimo da smo (7) mogli napisati u sljedećem obliku

$$\pi_j = \{a \in \mathcal{A} : c_j = \operatorname{argmin}_{j'} d(c_{j'}, a_i)\}, \quad j = 1, \dots, k.$$

Zato se problem traženja optimalne particije skupa $\mathcal{A}$ može svesti na sljedeći optimizacijski problem

$$\min_{c_1, \dots, c_k \in \mathbb{R}^n} F(c_1, \dots, c_k), \quad F(c_1, \dots, c_k) = \sum_{i=1}^m \min_{j=1, \dots, k} d(c_j, a_i), \quad (8)$$

gdje je $F : \mathbb{R}^{nk} \rightarrow \mathbb{R}_+$. Općenito, ova funkcija nije konveksna ni diferencijabilna, a može imati više lokalnih minimuma (Gan et al., 2007; Iyigun and Ben-Israel, 2010; Teboulle, 2007). Optimizacijski problem (8) u literaturi se može naći pod nazivom *k-median problem*, a najčešće se rješava raznim metaheurističkim metodama (Domínguez and Muñoz, 2005) ili uz primjenu cjelobrojnog programiranja (Pan and Chen, 2007; Schöbel and Scholz, 2010; Rodríguez-Chia et al., 2010). Pregled radova iz ovog područja do 2006. godine može se vijeti kod (Reese, 2006).

2 Standardni k -means algoritam

Uz pretpostavku da smo na neki način dobro procijenili početnu aproksimaciju centara klastera ili dobru početnu particiju, niže navedenim algoritmom možemo dobiti particiju dosta blisku optimalnoj (Kogan, 2007; Teboulle, 2007). Algoritam ćemo napisati dovoljno općenito uz korištenje kvazimetričke funkcije $d: X \times X \rightarrow \mathbb{R}_+$, gdje je X prostor podataka.

(Standardni k -means algoritam)

Učitati m, k , skup $\mathcal{A}$ i izabrati početne centre $c_1, \dots, c_k$;

repeat

Postaviti $\pi_j = \emptyset$, $j = 1, \dots, k$.

Separiranje

for $i = 1$ to m **do**

$\pi_j = \pi_j \cup a_i$

gdje je $j = j^*(a_i) = \operatorname{argmin}_{1 \leq j' \leq k} d(c_{j'}, a_i)$

end for

Određivanje centara

for $j = 1$ to k **do**

$c_j = c(\pi_j) := \operatorname{argmin}_{x \in \mathbb{R}^n} \sum_{a_i \in \pi_j} d(x, a_i)$

end for

until convergence

Može se pokazati da primjena algoritma smanjuje vrijednost kriterijske funkcije $\mathcal{F}$, ali općenito ne dovodi do točke globalnog minimuma funkcije.

3 Primjeri centra i separatora za neke kvazimetričke funkcije

Primjer 2. Funkcija $d: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_+$ zadana s

$$d_{LS}(x, y) = (x - y)^2.$$

je kvazimetrička funkcija, koju zovemo LS kvazimetrička funkcija. Prepostavimo da su zadani podaci $\mathcal{A} = \{a_i : i = 1, \dots, m\} \subseteq \mathbb{R}$. Za LS kvazimetričku funkciju riješit ćemo probleme separiranja i određivanje centara koji su pojavljuju u standardnom k -means algoritmu.

- 1) **problem separiranja** Neka su $c_1, c_2, \dots, c_k$ zadani centri takvi da je $c_1 \leq c_2 \leq \dots \leq c_k$. Treba odrediti klastere π_j , $j = 1, \dots, k$ takve da je

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

Kako su prema pretpostavci $c_1 \leq c_2 \leq \dots \leq c_k$ za određivanje klastera π_j dovoljno je promatrati susjedne centre c_{j-1}, c_j i c_{j+1} , $j = 2, \dots, k-1$, odnosno u klasteru π_j nalazit će se oni elementi $a \in \mathcal{A}$ za koje je $d_{LS}(c_j, a) \leq d_{LS}(c_{j-1}, a)$ i $d_{LS}(c_j, a) \leq$

$d_{LS}(c_{j+1}, a)$, odakle lako slijedi da je $a \geq \frac{c_{j-1}+c_j}{2}$ i $a \leq \frac{c_j+c_{j+1}}{2}$. Prema tome separator dva klastera π_j i π_{j+1} je aritmetička sredina centara c_j i c_{j+1} . Pri tome treba paziti da elementi na polovici intervala pripadnu točno jednom klasteru.

2) problem određivanja centra

Prepostavimo da je zadan klaster π_j . Problem određivanja centra $c_j = c(\pi_j)$ klastera π_j sastoji se od rješavanja optimizacijskog problema

$$c_j = c(\pi_j) = \operatorname{argmin}_{x \in \mathbb{R}} \sum_{a_i \in \pi_j} d_{LS}(x, a_i) = \operatorname{argmin}_{x \in \mathbb{R}} \sum_{a_i \in \pi_j} (x - a_i)^2,$$

odakle slijedi da je $c_j = c(\pi_j) = \frac{1}{|\pi_j|} \sum_{a_i \in \pi_j} a_i$.

Primjer 3. Funkcija $d : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_+$ zadana s

$$d_1(x, y) = |x - y|.$$

je kvazimetrička funkcija, koja dodatno zadovoljava i svojstva simetričnosti i nejednakost trokuta pa je ujedno i metrika. Ovu metriku zovemo L_1 metrička funkcija. Prepostavimo da su zadani podaci $\mathcal{A} = \{a_i : i = 1, \dots, m\} \subseteq \mathbb{R}$. Za L_1 metričku funkciju riješit ćemo probleme separiranja i određivanje centara.

1) **problem separiranja** Kako su prema pretpostavci $c_1 \leq c_2 \leq \dots \leq c_k$ za određivanje klastera π_j dovoljno je promatrati susjedne centre c_{j-1}, c_j i c_{j+1} , $j = 2, \dots, k-1$, odnosno u klasteru π_j nalazit će se oni elementi $a \in \mathcal{A}$ za koje je $d_1(c_j, a) \leq d_1(c_{j-1}, a)$ i $d_1(c_j, a) \leq d_1(c_{j+1}, a)$, odakle lako slijedi da je $a \geq \frac{c_{j-1}+c_j}{2}$ i $a \leq \frac{c_j+c_{j+1}}{2}$. Prema tome separator dva klastera π_j i π_{j+1} je aritmetička sredina centara c_j i c_{j+1} . Pri tome treba paziti da elementi na polovici intervala pripadnu točno jednom klasteru.

2) problem određivanja centra

Prepostavimo da je zadan klaster π_j . Problem određivanja centra $c_j = c(\pi_j)$ klastera π_j sastoji se od rješavanja optimizacijskog problema

$$c_j = c(\pi_j) = \operatorname{argmin}_{x \in \mathbb{R}} \sum_{a_i \in \pi_j} d_1(x, a_i) = \operatorname{argmin}_{x \in \mathbb{R}} \sum_{a_i \in \pi_j} |x - a_i|,$$

odakle slijedi da je $c_j = c(\pi_j) = \operatorname{med}_{a_i \in \pi_j}(a_i)$.

Primjer 4. Prepostavimo da su zadani podaci $\mathcal{A} = \{a_i : i = 1, \dots, m\} \subseteq \mathbb{R}^n$ te kvazimetrička funkciju $d_{LS} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+$, zadana su $d_{LS}(x, y) = \|x - y\|_2^2$. Nadalje pretpostavimo da su zadani centri $c_1, \dots, c_k \in \mathbb{R}^n$. Treba odrediti klastere π_j , $j = 1, \dots, k$ tako da vrijedi

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

Vrijedi $d_{LS}(c_j, a) \leq d_{LS}(c_s, a)$ onda i samo onda ako je $\|c_j - a\|_2^2 \leq \|c_s - a\|_2^2$, $s = 1, \dots, k$, odnosno onda i samo onda ako je $(c_s - c_j)^T a \leq \frac{1}{2}(\|c_s\|_2^2 - \|c_j\|_2^2)$, $s = 1, \dots, k$. Prema tome separator između dva centra c_j i c_s je hiperravnina koja je okomita na spojnicu točaka c_j i c_s te prolazi točkom $\frac{1}{2}(c_j + c_s)$.

4 Bregmanove divergencije

Neka je $\phi : S \rightarrow \mathbb{R}$ strogo konveksna funkcija definirana na konveksnom skupu $S \subseteq \mathbb{R}^n$, takva da je ϕ diferencijabilna na $\text{int}S$. Funkciju $d_\phi : S \times \text{int}S \rightarrow [0, \infty)$, zadanu formulom

$$d_\phi(x, y) = \phi(x) - \phi(y) - (\nabla\phi(y))^T(x - y),$$

zovemo Bregmanova divergencija.

Zbog jednostavnosti pretpostavimo da je $n = 1$ te $x, y \in \mathbb{R}$. Pretpostavimo da smo u točki $(y, \phi(y))$ povukli tangentu t na graf funkcije ϕ , čija je jednačba $z_t(\xi) = \phi(y) + \phi'(y)(\xi - y)$. Onda je

$$d_\phi(x, y) = \phi(x) - z_t(x) = \phi(x) - \phi(y) - \phi'(y)(x - y).$$

Primjer 5. Ako je $\phi : \mathbb{R} \rightarrow \mathbb{R}$ zadana s $\phi(x) = x^2$, onda je

$$d_\phi(x, y) = x^2 - y^2 - 2y(x - y) = (x - y)^2 = d_{LS}.$$

Nije teško vidjeti da je Bregmanova divergencija kvazimetrička funkcija u smislu Definicije 1.

Primjer 6. Pretpostavimo da je $\phi(x) = -\ln x$, odakle je $d_\phi(x, y) = \frac{x}{y} - \ln \frac{x}{y} - 1 = d_{IS}(x, y)$. Ovu Bregmanovu divergenciju zovemo Itakura-Saito divergencija. Pretpostavimo da su zadani podaci $\mathcal{A} = \{a_i : i = 1, \dots, m\} \subseteq (0, \infty)$. Za kvazimetričku funkciju d_{IS} riješit ćemo probleme separiranja i određivanje centara.

- 1) **problem separiranja** Kako su prema pretpostavci $c_1 \leq c_2 \leq \dots \leq c_k$ za određivanje klastera π_j dovoljno je promatrati susjedne centre c_{j-1}, c_j i c_{j+1} , $j = 2, \dots, k-1$, odnosno u klasteru π_j nalazit će se oni elementi $a \in \mathcal{A}$ za koje je $d_{IS}(c_j, a) \leq d_{IS}(c_{j-1}, a)$ i $d_{IS}(c_j, a) \leq d_{IS}(c_{j+1}, a)$. Bez smanjenja općenitosti možemo pretpostaviti da je $c_{j-1} < c_j < c_{j+1}$. Odatle lako slijedi da je $a \geq \frac{c_j - c_{j-1}}{\ln c_j - \ln c_{j-1}}$ i $a \leq \frac{c_{j+1} - c_j}{\ln c_{j+1} - \ln c_j}$. Prema tome separator dva klastera π_j i π_{j+1} je broj

$$\frac{c_{j+1} - c_j}{\ln c_{j+1} - \ln c_j}.$$

- 2) **problem određivanja centra** Pretpostavimo da je zadan klaster π_j . Problem određivanja centra $c_j = c(\pi_j)$ klastera π_j sastoji se od rješavanja optimizacijskog problema

$$c_j = \operatorname{argmin}_{x \in \mathbb{R}} \sum_{a_i \in \pi_j} d_{IS}(x, a_i) = \operatorname{argmin}_{x \in \mathbb{R}^n} \sum_{a_i \in \pi_j} \left(\frac{x}{a_i} - \ln \frac{x}{a_i} - 1 \right).$$

Označimo s $g(x) = \sum_{a_i \in \pi_j} \left(\frac{x}{a_i} - \ln \frac{x}{a_i} - 1 \right)$. Deriviranjem slijedi da je

$$g'(x) = \sum_{a_i \in \pi_j} \left(\frac{1}{a_i} - \frac{1}{x} \right),$$

odakle slijedi da je stacionarna točka funkcije g jednaka

$$x = \frac{|\pi_j|}{\sum_{a_i \in \pi_j} \frac{1}{a_i}},$$

što je zbog konveksnosti funkcije g ujedno točka globalnog minimuma. Prema tome centar klastera u smislu ove Bregmanove divergencija je harmonijska sredina podataka.

Primjer 7. Zadana je konveksna funkcija $\phi : \mathbb{R}_+^n \rightarrow \mathbb{R}$ formulom $\phi(x) = \sum_{i=1}^n x^{(i)} \ln x^{(i)}$, odakle se lako dobiva odgovarajuća Bregmanova divergencija $d_\phi(x, y) = \sum_{i=1}^n x^{(i)} \ln \frac{x^{(i)}}{y^{(i)}} - \sum_{i=1}^n (x^{(i)} - y^{(i)})$. Ova se divergencija obično naziva **generalizirana I-divergencija**. Ako su c_1, c_2 zadani centri onda za skup svih $a \in \mathbb{R}_+^n$ takvih da je $d_\phi(c_1, a) = d_\phi(c_2, a)$ vrijedi

$$\sum_{i=1}^n c_1^{(i)} \ln \frac{c_1^{(i)}}{a^{(i)}} - \sum_{i=1}^n c_1^{(i)} = \sum_{i=1}^n c_2^{(i)} \ln \frac{c_2^{(i)}}{a^{(i)}} - \sum_{i=1}^n c_2^{(i)}. \quad (9)$$

Prema tome separator u smislu ove kvazimetričke funkcije je ploha čija je jednadžba zadana s (9). Za centre $c_1 = (1, 5)$ i $c_2 = (3, 1)$ odgovarajući separator u smislu ove kvazimetričke funkcije prikazan je na Slici 1.

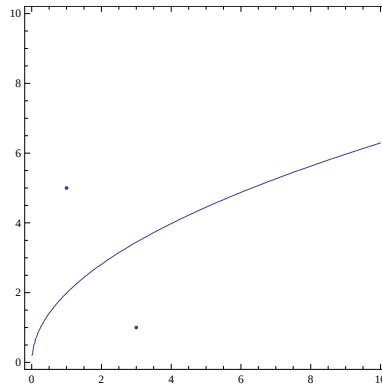


Figure 1: Separator u smislu generalizirane I-divergencije

References

- A. Ben-Israel, C. Iyigun, *Probabilistic D-clustering*, Journal of Classification **25**(2008), 5–26
- D. L. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.
- T. Calinski, J. Harabasz, *A dendrite method for cluster analysis*, Communications in Statistics, **3**(1974), 1–27
- D. Davies, D. Bouldin, *A cluster separation measure*, IEEE Transactions on Pattern Analysis and Machine Intelligence, textbf2(1979), 224–227

- I. S. Dhillon, Y. Guan, B. Kulis, *Kernel k -means, spectral clustering and normalized cuts*, Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), August 22–25, 2004, Seattle, Washington, USA, 551–556, 2004
- G. Divéki, I. Csanád, *Online facility location with facility movements*, CEJOR (2010), DOI 10.1007/s10100-010-0153-8
- E. Domínguez, J. Muñoz, *Applying bio-inspired techniques to the p -median problem*, IWANN 2005; Computational Intelligence Bioinspired Syst., 8th Int. Workshop Artificial Neural Networks, Lecture Notes in Computer Science, Springer-Verlag, Berlin, 2005, pp. 67 – 74
- Z. Drezner, *Facility Location: A Survey of Applications and Methods*, Springer-Verlag, Berlin, 2004.
- J. Dunn, *Well separated clusters and optimal fuzzy partitions*, Journal of Cybernetics, **4**(1974), 95–104
- B. S. Everitt, S. Landau, M. Leese, *Cluster analysis*, Wiley, London, 2001.
- D. E. Finkel, C. T. Kelley, *Additive scaling and the DIRECT algorithm*, J. Glob. Optim. **36**(2006), 597–608
- C. A. Floudas, C. E. Gounaris, *A review of recent advances in global optimization*, J. Glob. Optim. **45**(2009), 3–38
- G. Gan, C Ma, J. Wu, *Data Clustering: Theory, Algorithms, and Applications*, SIAM, Philadelphia, 2007.
- E. R. Hansen, G. W. Walster, *Global Optimization Using Interval Analysis*. Marcel Dekker, New York, Second Edition, Revised and Expanded, 2004.
- M. Hudec, M. Vujošević, *A fuzzy system for municipalities classification*, CEJOR **18**(2010), 171–180
- C. Iyigun, A. Ben-Israel, *A generalized Weiszfeld method for the multi-facility location problem*, Operations Research Letters **38**(2010), 207–214
- C. Iyigun, *Probabilistic Distance Clustering*, Dissertation, Graduate School – New Brunswick, Rutgers, 2007
- D. R. Jones, C. D. Perttunen, B. E. Stuckman, *Lipschitzian optimization without the Lipschitz constant*, JOTA **79**(1993), 157–181
- J. Kogan, *Introduction to Clustering Large and High-Dimensional Data*, Cambridge University Press, 2007.
- J. Kogan, C. Nicholas, M. Wiacek, *Hybrid Clustering of large high dimensional data*, In M. Castellanos and M. W. Berry (Eds.), Proceedings of the Workshop on Text Mining, SIAM, 2007.

- J. Kogan, M. Teboulle, *Scaling clustering algorithms with Bregman distances*. In: M. W. Berry and M. Castellanos (Eds.), *Proceedings of the Workshop on Text Mining at the Sixth SIAM International Conference on Data Mining*, 2006.
- J. Kogan, C. Nicholas, M. Wiacek, *Hybrid clustering with divergences*. In: M. W. Berry and M. Castellanos (Eds.), *Survey of Text Mining: Clustering, Classification, and Retrieval*, Second Edition, Springer, 2007.
- C. Iyigun, A. Ben-Israel, *A generalized Weiszfeld method for the multi-facility location problem*, *Operations Research Letters* **38**(2010) 207–214
- D. Littau, D. L. Boley, *Clustering very large data sets with PDDP*. In J. Kogan, C. Nicholas, M. Teboulle (eds), *Grouping Multidimensional Data: Recent Advances in Clustering*, 99–126, Springer-Verlag, New York, 2006.
- S. Pan, J., S. Chen, *Two unconstrained optimization approaches for the Euclidean k -centrum location problem*, *Applied Mathematics and Computation* **189**(2007) 1368–1383
- С. А. Пиявский, Один алгоритм отыскания абсолютного экстремума функции, *Ж. вычисл. матем. и матем. физ.* **12**(1972), 888–896.
- J. Reese, *Solution methods for the p -median problem: an annotated bibliography*, Published online in Wiley InterScience, Wiley, 2006.
- A. M. Rodríguez-Chia, I. Espejo, Z. Drezner, *On solving the planar k -centrum problem with Euclidean distances*, *EJOR*, to appear
- K. Sabo, R. Scitovski, *The best least absolute deviations line – properties and two efficient methods*, *ANZIAM Journal* **50**(2008), 185–198
- K. Sabo, R. Scitovski, I. Vazler., M. Zekić-Sušac, *Mathematical models of natural gas consumption*, *Energy Conversion and Management*, prihvaćeno za objavljivanje
- A. Schöbel, D. Scholz, *The big cube small cube solution method for multidimensional facility location problems*, *Computers & Operations Research* **37**(2010), 115–122
- A. Schöbel, *Locating Lines and Hyperplanes: Theory and Algorithms*, Springer Verlag, Berlin, 1999.
- B. Shubert, *A sequential method seeking the global maximum of a function*, *SIAM Journal on Numerical Analysis*, **9**(1972), 379–388
- H. Späth, *Cluster-Formation und Analyse*, R. Oldenburg Verlag, München, 1983.
- Z. Su, J. Kogan, C. Nicholas, *Constrained clustering with k -means type algorithms*, In M.W. Berry, J. Kogan (eds), *Text Mining Applications and Theory*, 81–103, Willey, Chichester, 2010.
- M. Teboulle, *A unified continuous optimization framework for center-based clustering methods*, *Journal of Machine Learning Research* **8**(2007), 65–102

Vazler, I., Sabo, K., Scitovski, R.: Weighted median of the data in solving least absolute deviations problems. Comm. Statist. Theory Methods (to appear)

D. Veljan, *Kombinatorna i diskretna matematika*, Algoritam, Zagreb, 2001.