Odjel za matematiku
Sveučilište J. J. Strossmayera u Osijeku

Algoritmi i strukture podataka

II. Reprezentant

Rudolf Scitovski, Martina Briš Alić

Sadržaj

1 Reprezentant podataka iz $\mathbb{R}$

Zadani su podaci $y_1, y_2, \ldots, y_m \in \mathbb{R}$.
Treba odrediti realni broj $c^\star \in \mathbb{R}$ (reprezentant podataka) koji će što bolje reprezentirati podatke.

Definicija 1. Funkciju $d\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$, koja ima svojstvo *pozitivne definitnosti*

$$\forall x, y \in \mathbb{R} \quad d(x,y) \geq 0 \quad \& \quad d(x,y) = 0 \quad \Leftrightarrow \quad x = y,$$

zovemo *kvazimetrička funkcija* (funkcija sličnosti, funkcija različitosti)

Primjer 1. *Dvije najčešće korištene kvazimetričke funkcije:*

(a) $d_{LS}(x,y) = (x-y)^2$ – *Least Squares (LS) kvazimetrička funkcija*

(b) $d_1(x,y) = |x-y|$ – l_1 *metrička funkcija (Manhattan metrika)*

Primijetite da u $\mathbb{R}$ *vrijedi* $d_1(x,y) = d_2(x,y) = d_\infty(x,y) = d_p(x,y), \quad p \geq 1$

Zadatak 1. *Pokažite da funkcija* d_{LS} *iz prethodnog primjera nije metrika na* $\mathbb{R}$*, a da je funkcija* d_1 *metrika na* $\mathbb{R}$*.*

Definicija 2. Neka je $d\colon \mathbb{R}\times\mathbb{R}\to\mathbb{R}_+$ kvazimetrička funkcija. Kažemo da je $c^\star\in\mathbb{R}$ *najbolji reprezentant* podataka $y_1,y_2,\ldots,y_m\in\mathbb{R}$ u odnosu na kvazimetričku funkciju d onda ako je

$$c^\star=\operatorname*{argmin}_{c\in\mathbb{R}}\sum_{i=1}^{m}d(c,y_i), \tag{1}$$

tj ako je $c^\star\in\mathbb{R}$ točka globalnog minimuma funkcionala $F\colon\mathbb{R}\to\mathbb{R}_+$

$$F(c)=\sum_{i=1}^{m}d(c,y_i). \tag{2}$$

Primjer 2. *Za LS-kvazimetričku funkciju najbolji reprezentant podataka* $y_1,y_2,\ldots,y_m\in\mathbb{R}$ *je običnа* aritmetička sredina

$$c^\star_{LS}=\operatorname*{argmin}_{c\in\mathbb{R}}\sum_{i=1}^{m}d_{LS}(c,y_i)=\tfrac{1}{m}\sum_{i=1}^{m}y_i,$$

a odgovarajući funkcional glasi

$$F_{LS}(c)=\sum_{i=1}^{m}(y_i-c)^2.$$

Za l_1*-metričku funkciju najbolji reprezentant podataka* $y_1,y_2,\ldots,y_m\in\mathbb{R}$ *je običnі* medijan

$$c^\star_1=\operatorname*{argmin}_{c\in\mathbb{R}}\sum_{i=1}^{m}d_1(c,y_i)=\operatorname*{med}_i y_i,$$

a odgovarajući funkcional glasi

$$F_1(c)=\sum_{i=1}^{m}|y_i-c|.$$

Definicija 3. Neka je $d\colon\mathbb{R}\times\mathbb{R}\to\mathbb{R}_+$ kvazimetrička funkcija. Kažemo da je $c^\star\in\mathbb{R}$ *najbolji reprezentant* podataka $y_1,y_2,\ldots,y_m\in\mathbb{R}$ s težinama $w_1,\ldots,w_m>0$ u odnosu na kvazimetričku funkciju d onda ako je

$$c^\star=\operatorname*{argmin}_{c\in\mathbb{R}}\sum_{i=1}^{m}w_i d(c,y_i), \tag{3}$$

tj ako je $c^\star\in\mathbb{R}$ točka globalnog minimuma funkcionala $F\colon\mathbb{R}\to\mathbb{R}_+$

$$F(c)=\sum_{i=1}^{m}w_i d(c,y_i). \tag{4}$$

Primjer 3. *Za LS-kvazimetričku funkciju najbolji reprezentant podataka* $y_1,y_2,\ldots,y_m\in\mathbb{R}$ *s težinama* $w_1,\ldots,w_m>0$ *je* težinska aritmetička sredina

$$c^\star_{LS}=\operatorname*{argmin}_{c\in\mathbb{R}}\sum_{i=1}^{m}w_i d_{LS}(c,y_i)=\tfrac{1}{W}\sum_{i=1}^{m}w_i y_i,\quad W=\sum_{i=1}^{m}w_i$$

a odgovarajući funkcional glasi

$$F_{LS}(c) = \sum_{i=1}^{m} w_i (y_i - c)^2.$$

Za l_1-kvazimetričku funkciju najbolji reprezentant podataka $y_1, y_2, \ldots, y_m \in \mathbb{R}$ s težinama $w_1, \ldots, w_m > 0$ je težinski medijan

$$c_1^\star = \operatorname*{argmin}_{c \in \mathbb{R}} \sum_{i=1}^{m} w_i d_1(c, y_i) = \operatorname*{med}_i (w_i, y_i),$$

a odgovarajući funkcional glasi

$$F_1(c) = \sum_{i=1}^{m} w_i |y_i - c|.$$

1.1 Udaljenost na kružnici

- Proteklo vrijeme na satu s 12 oznaka:

$$d(t_1, t_2) = \begin{cases} t_2 - t_1, & \text{ako } t_1 \leq t_2 \\ 12 + (t_2 - t_1), & \text{ako } t_1 > t_2 \end{cases},$$

  Udaljenost $d(t_1, t_2)$ predstavlja proteklo vrijeme na satu od trenutka "t_1" do trenutka "t_2".
  Primjerice: $d(2, 7) = 5$, ali $d(7, 2) = 12 + (-5) = 7$
  Primijetite da ova funkcija nije simetrična.

- Duljina vremenskog intervala na satu s 12 oznaka:

$$d(t_1, t_2) = \begin{cases} |t_1 - t_2|, & \text{ako } |t_1 - t_2| \leq 6 \\ 12 - |t_1 - t_2|, & \text{ako } |t_1 - t_2| > 6 \end{cases},$$

  Ova udaljenost $d(t_1, t_2)$ predstavlja duljina vremenskog intervala na satu s 12 oznaka od trenutka "t_1" do trenutka "t_2".
  Primjerice: $d(2, 9) = 12 - 7 = 5$, ali $d(2, 7) = 7 - 2 = 5$
  Primijetite da je ova funkcija simetrična.

- Udaljenost na jediničnoj kružnici:

$$d(t_1, t_2) = \begin{cases} |t_1 - t_2|, & \text{ako } |t_1 - t_2| \leq \pi \\ 2\pi - |t_1 - t_2|, & \text{ako } |t_1 - t_2| > \pi \end{cases}$$

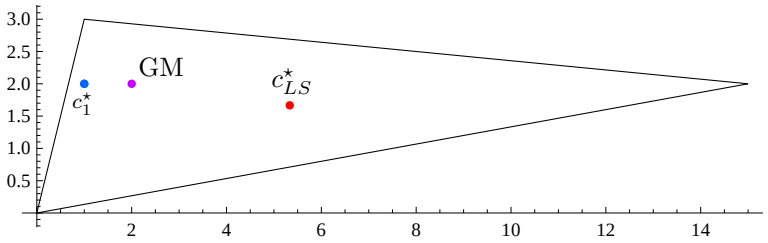
  Primjerice: $d(0, \frac{\pi}{4}) = \frac{\pi}{4}$, ali $d(\frac{\pi}{4}, \frac{3\pi}{2}) = 2\pi - \left(\frac{6\pi}{4} - \frac{\pi}{4}\right) = \frac{3\pi}{4}$
  Ova udaljenost $d(t_1, t_2)$ predstavlja duljinu luka jedinične kružnice od t_1 do t_2 radijana.

2 Reprezentant podataka iz $\mathbb{R}^2$

2.1 Fermat – Torricelli – Weberov problem

Neka su $A_1, A_2, A_3 \in \mathbb{R}^2$ tri nekolinearne točke u ravnini. Točka $c^\star_{LS} \in \mathbb{R}^2$, za koju je suma kvadrata euklidskih udaljenosti do vrhova trokuta minimalna zove se **centroid** ili **Steinerova točka** (povezano s pojmom centra masa u fizici). Naka je

$$A_1 = (x_1, y_1), \quad A_2 = (x_2, y_2), \quad A_3 = (x_3, y_3).$$



Slika 1: Fermat – Torricelli – Weberov problem (GM-geometrijski medijan, $c^\star_{LS}$ - centroid, $c^\star$ - medijan)

Centroid $c^\star_{LS} = (x_c, y_c)$ točaka A_1, A_2, A_3 je rjesenje optimizacijskog problema

$$\underset{T \in \mathbb{R}^2}{\operatorname{argmin}} \sum_{i=1}^{3} d_2^2(T, A_i),$$

tj. točka u kojoj se postiže minimum funkcije (suma kvadrata euklidskih l_2 udaljenosti do tocaka A_1, A_2, A_3)

$$F_{LS}(x, y) = \sum_{i=1}^{3} [(x - x_i)^2 + (y - y_i)^2],$$

Dobivamo

$$x_c = \frac{1}{3} \sum_{i=1}^{3} x_i, \qquad y_c = \frac{1}{3} \sum_{i=1}^{3} y_i.$$

Primjer 4. $A_1 = (0, 0), \quad A_2 = (1, 3.5), \quad A_3 = (14, 2.5)$
$c^\star_{LS} = (5, 2)$ – *centroid*

Točka $c^\star_1 = (x^\star, y^\star)$ koja je rješenje optimizacijskog problema

$$\underset{T \in \mathbb{R}^2}{\operatorname{argmin}} \sum_{i=1}^{3} d_1(T, A_i),$$

tj. točka u kojoj se postiže minimum funkcije (suma l_1 udaljenosti do točaka A_1, A_2, A_3)

$$F_1(x, y) = \sum_{i=1}^{3} (|x - x_i| + |y - y_i|)$$

naziva se medijan točaka $A_1, A_2, A_3 \in \mathbb{R}^2$. Lako dobivamo

$$c_1^\star = (\operatorname{med} x_i, \operatorname{med} y_i)\,.$$

Primjer 5. $A_1 = (0,0), \quad A_2 = (1, 3.5), \quad A_3 = (14, 2.5)$
$c_1^\star = (1, 2.5)$ – *median*

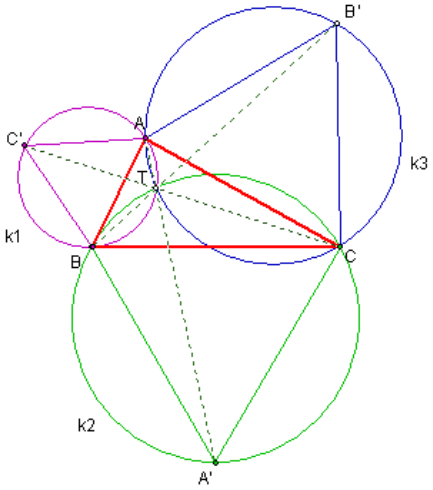
Točka $M = (x_{GM}, y_{GM})$ za koju se postiže minimum funkcionala

$$\Psi(x,y) = \sum_{i=1}^{3} d_2(T(x,y), A_i(x_i,y_i)) = \sum_{i=1}^{3} \sqrt{(x-x_i)^2 + (y-y_i)^2} \to \min_{(x,y)\in\mathbb{R}^2}$$

naziva se **Geometrijski medijan** točaka $A_1, A_2, A_3 \in \mathbb{R}^2$ u smislu l_2-norme i ne može se eksplicitno izraziti (vidi Sliku 1).

$$M_2 = (x_{GM}, y_{GM}) = \operatorname{argmin} \Psi,$$

Primjer 6. $A_1 = (0,0), \quad A_2 = (1, 3.5), \quad A_3 = (14, 2.5)$
$\Psi(x,y) = \sqrt{x^2+y^2} + \sqrt{(x-1)^2 + (y-3.5)^2} + \sqrt{(x-14)^2 + (y-2.5)^2}$;
$M_2 = (1.51827, 2.5876)$ – *Weiszfeldov algoritam*



Slika 2: Fermat – Torricelli – Weberov problem - geometrijsko rješenje: T je geometrijski medijan

2.2 Kvazimetričke funkcije i reprezentanti

Zadan je skup točaka $A = \{T_i = (x_i, y_i) \in \mathbb{R}^2 \colon i = 1, \ldots, m\}$, odnosno
vektora $\mathcal{A} = \{a^i = (x_i, y_i)^T \in \mathbb{R}^2 \colon i = 1, \ldots, m\}$.
Treba odrediti točku $C^\star$ (odnosno vektor $c^\star$) koja će što bolje reprezentirati skup točaka A (odnosno skup vektora $\mathcal{A}$).

Definicija 4. Funkciju $d\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}_+$, koja ima svojstvo *pozitivne definitnosti*

$$\forall x, y \in \mathbb{R}^2 \quad d(x,y) \geq 0 \quad \& \quad d(x,y) = 0 \quad \Leftrightarrow \quad x = y,$$

zovemo *kvazimetrička funkcija* (funkcija sličnosti, funkcija različitosti) na $\mathbb{R}^2$.

Primjer 7. *Najčešći primjeri (Gan et al., 2007; Kogan, 2007; Späth, 1983):*

(a) $d_{LS}(x,y) = \|x-y\|_2^2 = (x-y)^T(x-y)$ – *Least Squares (LS) kvazimetrička funkcija*

(b) $d_2(x,y) = \|x-y\|_2 = \sqrt{(x-y)^T(x-y)}$ – l_2 *euklidska metrička funkcija*

(c) $d_1(x,y) = \|x-y\|_1$ – l_1 *metrička funkcija (Manhattan metrika)*

(d) $d_\infty(x,y) = \|x-y\|_\infty$ – l_∞ *Čebiševljeva metrička funkcija*

(e) $d_p(x,y) = \|x-y\|_p$, $p \geq 1$ – l_p *Minkowsky metrička funkcija*

(f) $d_M(x,y) = (x-y)^T S^{-1}(x-y)$ – *Mahalanobis kvazimetrička funkcija* ($S \in \mathbb{R}^{2\times 2}$ *je simetrična pozitivno definitna matrica)*

Definicija 5. Neka je $d\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}_+$ kvazimetrička funkcija. Kažemo da je $c^\star \in \mathbb{R}^2$ *najbolji reprezentant* (centroid) skupa $\mathcal{A}$ u odnosu na kvazimetričku funkciju d onda ako je

$$c^\star = \operatorname*{argmin}_{c\in\mathbb{R}^2} \sum_{i=1}^m d(c, a^i), \tag{5}$$

tj. ako je $c^\star \in \mathbb{R}^2$ točka globalnog minimuma funkcionala $F\colon \mathbb{R}^2 \to \mathbb{R}_+$

$$F(c) = \sum_{i=1}^m d(c, a^i). \tag{6}$$

Vrijedi:

(a) Za LS-kvazimetričku funkciju najbolji reprezentant skupa $\mathcal{A}$ je *centroid (težište) skupa*

$$c^\star_{LS} = \operatorname*{argmin}_{c\in\mathbb{R}^2} \sum_{i=1}^m d_{LS}(c, a^i) = \tfrac{1}{m}\sum_{i=1}^m a^i = \left(\tfrac{1}{m}\sum_{i=1}^m x_i, \tfrac{1}{m}\sum_{i=1}^m y_i\right),$$

a odgovarajući funkcional glasi

$$F_{LS}(c) = \sum_{i=1}^m \|c - a^i\|_2^2.$$

(b) Za l_1-metričku funkciju najbolji reprezentant skupa $\mathcal{A}$ je *medijan skupa*

$$c^\star_1 = \operatorname*{argmin}_{c\in\mathbb{R}^2} \sum_{i=1}^m d_1(c, a^i) = \operatorname*{med}_i a^i = \left(\operatorname*{med}_i x_i, \operatorname*{med}_i y_i\right),$$

a odgovarajući funkcional glasi

$$F_1(c) = \sum_{i=1}^m \|c - a^i\|_1.$$

Definicija 6. Neka je $d\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}_+$ kvazimetrička funkcija. Kažemo da je $c^\star \in \mathbb{R}^2$ *najbolji reprezentant* skupa $\mathcal{A}$ s težinama $w_1, \ldots, w_m > 0$ u odnosu na kvazimetričku funkciju d onda ako je

$$c^\star = \operatorname*{argmin}_{c\in\mathbb{R}^2} \sum_{i=1}^{m} w_i d(c, a^i), \tag{7}$$

tj. ako je $c^\star \in \mathbb{R}^2$ točka globalnog minimuma funkcionala $F\colon \mathbb{R}^2 \to \mathbb{R}_+$

$$F(c) = \sum_{i=1}^{m} w_i d(c, a^i). \tag{8}$$

Vrijedi:

(a) Za LS-kvazimetričku funkciju najbolji reprezentant skupa $\mathcal{A}$ s težinama $w_1, \ldots, w_m > 0$ je *težinski centroid (težište) skupa*

$$c^\star_{LS} = \operatorname*{argmin}_{c\in\mathbb{R}^2} \sum_{i=1}^{m} w_i d_{LS}(c, a^i) = \tfrac{1}{W}\sum_{i=1}^{m} w_i a^i, \quad W = \sum_{i=1}^{m} w_i, \quad \text{tj.}$$
$$c^\star_{LS} = \left(\tfrac{1}{W}\sum_{i=1}^{m} w_i x_i, \tfrac{1}{W}\sum_{i=1}^{m} w_i y_i\right),$$

a odgovarajući funkcional glasi

$$F_{LS}(c) = \sum_{i=1}^{m} w_i \|c - a^i\|_2^2.$$

(b) Za l_1-metričku funkciju najbolji reprezentant skupa $\mathcal{A}$ s težinama $w_1, \ldots, w_m > 0$ je *težinski medijan skupa*

$$c^\star_1 = \operatorname*{argmin}_{c\in\mathbb{R}^2} \sum_{i=1}^{m} w_i d_1(c, a^i) = \operatorname*{med}_i(w_i, a^i) = \left(\operatorname*{med}_i(w_i, x_i), \operatorname*{med}_i(w_i, y_i)\right),$$

a odgovarajući funkcional glasi

$$F_1(c) = \sum_{i=1}^{m} w_i \|c - a^i\|_1.$$

3 Reprezentant podataka iz $\mathbb{R}^n$

Zadan je skup točaka $A = \{T_i \in \mathbb{R}^n\colon i = 1, \ldots, m\}$, odnosno
vektora $\mathcal{A} = \{a^i = (x_1^i, \ldots, x_n^i)^T \in \mathbb{R}^n\colon i = 1, \ldots, m\}$.

Najbolji LS-reprezentant skupa $\mathcal{A}$ je **težinski centroid** skupa vektora $\mathcal{A} \subset \mathbb{R}^n$

$$c^\star_{LS} = \operatorname*{argmin}_{c\in\mathbb{R}^n} \sum_{i=1}^{m} w_i \|c - a^i\|_2^2 = \left(\tfrac{1}{W}\sum_{i=1}^{m} w_i a_1^{(i)}, \ldots, \tfrac{1}{W}\sum_{i=1}^{m} w_i a_n^{(i)}\right), \quad W = \sum_{i=1}^{m} w_i,$$

jer se na njemu postiže globalni minimum funkcionala

$$F_{LS}(c) = \sum_{i=1}^{m} w_i \|c - a^i\|_2^2.$$

Najbolji l_1-reprezentant skupa $\mathcal{A}$ je **težinski median** skupa vektora $\mathcal{A} \subset \mathbb{R}^n$

$$c_1^\star = \operatorname*{argmin}_{c \in \mathbb{R}^n} \sum_{i=1}^{m} w_i \|c - a^i\|_1 = \left(\operatorname*{med}_{i=1,\ldots,m} (w_j, a_1^{(i)}), \ldots, \operatorname*{med}_{j=1,\ldots,m} (w_j, a_n^{(i)}) \right)^T \tag{9}$$

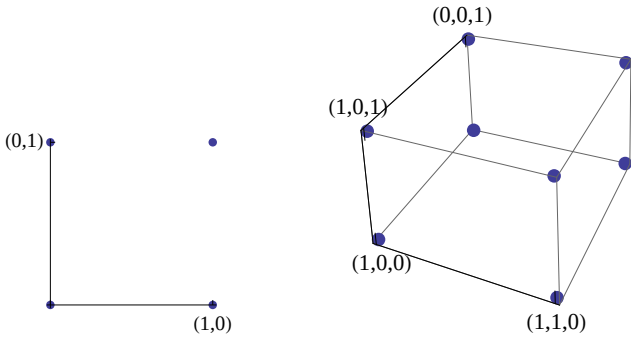
jer se na njemu postiže globalni minimum funkcionala

$$F_1(c) = \sum_{i=1}^{m} w_i \|c - a^i\|_1.$$

4 Prepoznavanje riječi

$\mathcal{A} = \{a^i = (x_1, \ldots, x_n)^T \in \{0,1\}^n : i = 1, \ldots, m\} \subset \mathbb{R}^n$
$d \colon \mathcal{A} \times \mathcal{A} \to \mathbb{R}_+$ – kvazimetrička funkcija, primjerice
d_{LS}, d_1, $d_c(x,y) = 1 - \frac{\langle x,y \rangle}{\|x\| \cdot \|y\|}$ (kosinus).

U nekom tekst prisutnost neke riječi kodira se s 1, a odsutnost te riječi iz teksta s 0. Postavlja se pitanje o sličnosti/različitosti dva teksta obzirom na prisutnost/odsutnost promatranih riječi. Tekst u kome je prisutno/odsutno $n \geq 1$ izabranih riječi prikazat ćemo vektorom iz $\mathbb{R}^n$ s komponentama 0 ili 1.



Slika 3: Skup $\mathcal{A}$ za $n = 2$ i $n = 3$

Primjer 8. *Promatramo tekstove u kojima se mogu pojaviti riječi: A,B,C. Neka je primjerice (vidi Sliku 1):*

$a^1 = (1,1,0)$: *tekst u kome se pojavljuju riječi A,B, a ne pojavljuje riječ C*

$a^2 = (1,0,0)$: *tekst u kome se pojavljuje riječ A, a ne pojavljuju riječi B,C*

$a^3 = (1,0,1)$: *tekst u kome se pojavljuju riječi A, C, a ne pojavljuje riječ B*

$a^4 = (0,0,1)$: *tekst u kome se pojavljuje riječ C, a ne pojavljuju riječi A, B*

U svrhu ispitivanja sličnosti/različitosti tekstova obzirom na prisutnost/odsutnost nekih riječi možemo pokušati iskoristiti ranije spomenute metričke funkcije d_1, d_2, d_∞. U znanstvenoj literaturi (vidi primjerice (Berry and Kogan, 2010; Zhang, 2009)) u tu svrhu koriste se neke tzv. *kvazimetričke funkcije*, kao što su

$$d_{LS}(x,y) = \|x-y\|^2 \quad \text{– Least Squares (LS) kvazimetrička funkcija}$$
$$d_c(x,y) = 1 - \frac{\langle x,y\rangle}{\|x\|\cdot\|y\|} \quad \text{– kosinus kvazimetrička funkcija}$$

Za prethodno spomenuti primjer dobivamo

$$\begin{array}{lll} d_{LS}(a^1,a^2)=1, & d_{LS}(a^1,a^3)=2, & d_{LS}(a^1,a^4)=3 \\ d_1(a^1,a^2)=1, & d_1(a^1,a^3)=2, & d_1(a^1,a^4)=3 \\ d_c(a^1,a^2)=1-\frac{\sqrt{2}}{2}=0.29, & d_c(a^1,a^3)=\frac{1}{2}, & d_c(a^1,a^4)=1 \end{array}$$

Prema LS-kvazimetričkoj funkciji (a također i prema l_1-metričkoj funkciji) tekstovi a^1 i a^2 su najsličniji (najbliži), a tekstovi a^1 i a^4 najrazličitiji (najudaljeniji) obzirom na pojavu riječi A,B,C.

I prema kosinus-metričkoj funkciji d_c tekstovi a^1 i a^2 su najsličniji (najbliži), a tekstovi a^1 i a^4 potpuno različiti (maksimalno udaljeni) obzirom na pojavu riječi A,B,C.

Primjer 9. *Promatramo tekstove u kojima se mogu pojaviti riječi: A,B,C,D,E. Neka je primjerice:*

$a^1 = (1,0,0,0,1)$*: tekst u kome se pojavljuju riječi A, E, a ne pojavljuju riječi B,C,D*

$a^2 = (0,1,1,0,0)$*: tekst u kome se pojavljuju riječi B, C, a ne pojavljuju riječi A,D,E*

$a^3 = (1,0,0,0,0)$*: tekst u kome se pojavljuje riječe A, a ne pojavljuju riječi B,C,D, E*

$d_{LS}(a^i,a^j)$	a^1	a^2	a^3	$d_1(a^i,a^j)$	a^1	a^2	a^3	$d_c(a^i,a^j)$	a^1	a^2	a^3
a^1	0	4	1	a^1	0	4	1	a^1	0	1	0.29
a^2	4	0	3	a^2	4	0	3	a^2	1	0	1
a^3	1	3	0	a^3	1	3	0	a^3	0.29	1	0

Iz ovog primjera vidi se da kosinus-kvazimetrička funkcija puno bolje identificira sličnosti/različitosti tekstova obzirom na prisutnost/odsutnost riječi A,B,C,D,E (objasnite to na osnovi brojeva iz tablica!).

Literatura

M. W. Berry, J. Kogan, *Text Mining. Applications and Theory*, Wiley, 2010.

D. L. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

S. Butenko, W. A. Chaovalitwongse, P. M. Pardalos, *Clustering Challenges in Biological Networks*, World Scientific, 2009.

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

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.

C. Iyigun, A. Ben-Israel, *A generalized Weiszfeld method for the multi-facility location problem*, Operations Research Letters **38**(2010) 207–214

D. R. Jones, C. D. Perttunen, B. E. Stuckman, *Lipschitzian optimization without the Lipschitz constant*, JOTA **79**(1993), 157-181

D. Jukić, *Minimizacija najvećeg apsolutnog osdtupanja*, Osječka matematička škola **1**(2001) pp 118.

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

F. Leisch, *A toolbox for K-centroids cluster analysis*, Computational Statistics & Data Analysis **51**(2006), 526-544

A. Neumaier, *Complete search in continuous global optimization and constraint satisfaction*, Acta Numerica (2006), 271-369.

P. M. Pardalos, P. Hansen, *Data Mining and Mathematical Programming*, American Mathematical Society, Providence, 2008.

J. Reese, *Solution methods for the p-median problem: an annotated bibliography*, Published online in Wiley InterScience, Wiley, 2006.

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, *One-dimensional center-based l_1-clustering method*, Optimization Letters (accepted) DOI:10.1007/s11590-011-0389-9

A. Schöbel, D. Scholz, *The big cube small cube solution method for multidimensional facility location problems*, Computers & Operations Research **37**(2010), 115–122

R. Scitovski, *Numerička matematika*, Odjel za matematiku, Sveučilište u Osijeku, Osijek, 2004.

H. Späth, *Cluster-Formation und Analyse*, R. Oldenburg Verlag, München, 1983.

M. Teboulle, *A unified continuous optimization framework for center-based clustering methods*, Journal of Machine Learning Research **8**(2007), 65–102

I. Vazler, K. Sabo, R. Scitovski, *Weighted median of the data in solving least absolute deviations problems*, Communications in Statistics - Theory and Methods, to appear in 2011

H. Zhang, Statistical Clustering Analysis: An Introduction, in S. Butenko, W. A. Chaovalitwongse, and P. M. Pardalos (eds.), *Clustering Challenges in Biological Networks*, World Scientific, 2009, pp 101–126