Karakterizacija rješenja LAD problema

1 LAD problem

Neka su $\mathbf{X} \in \mathbb{R}^{m\times n}$ te $\mathbf{y} \in \mathbb{R}^m$. Označimo s $\mathbf{r}(\mathbf{a}) = \mathbf{X}\mathbf{a} - \mathbf{y} \in \mathbb{R}^m$. Treba odrediti vektor $\mathbf{a} \in \mathbb{R}^n$ takav da je

$$\|\mathbf{r}(\mathbf{a})\|_1 \to \min_{\mathbf{a}\in\mathbb{R}^n} . \tag{1}$$

Problem (1) zovemo LAD (least absolute deviation) problem.

Primjer 1 *Zadani su podaci (x_i, y_i), $i = 1,\ldots,m$ $x_i \neq x_j$, $i \neq j$, za koje treba odrediti pravac oblika $y = ax + b$ tako da vrijedi*

$$\sum_{i=1}^{m} |ax_i + b - y_i| \to \min_{(a,b)\in\mathbb{R}^2} .$$

Pravac sa tim svojstvom zovemo LAD pravac. Problem određivanja LAD pravca specijalni je slučaj LAD problema (1) uz oznake $\mathbf{X} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{pmatrix}$, $\mathbf{y} = (y_1,\ldots,y_m)^T$ *te* $\mathbf{a} = (a,b)^T$.

2 Svođenje LAD problema na problem linearnog programiranja

Neka je

$$\|\mathbf{X}\mathbf{a} - \mathbf{y}\|_1 = \|\mathbf{r}(\mathbf{a})\|_1 = \sum_{i=1}^{m} |r_i|, \quad r_i := r_i(\mathbf{a}).$$

Uvedimo nove varijable $u_i, v_i \geq 0$ tako da je $r_i = u_i - v_i$, $i = 1,\ldots,m$ te $c_j, d_j \geq 0$ sa svojstvom $a_j = c_j - d_j$, $j = 1,\ldots,n$. Označimo s $\mathbf{e} = (1,\ldots,1)^T \in \mathbb{R}^m$. Promatrajmo sljedeći problem linearnog programiranja:

$$\begin{aligned} \sum_{i=1}^{m}(u_i + v_i) = \mathbf{e}^T\mathbf{u} + \mathbf{e}^T\mathbf{v} + \mathbf{0}^T\mathbf{c} + \mathbf{0}^T\mathbf{d} &\to \min_{\mathbf{u},\mathbf{v},\mathbf{c},\mathbf{d}} \\ -\mathbf{u} + \mathbf{v} + \mathbf{X}(\mathbf{c} - \mathbf{d}) &= \mathbf{y} \\ \mathbf{u},\mathbf{v},\mathbf{c},\mathbf{d} &\geq \mathbf{0}, \end{aligned}$$

odnosno u matričnom obliku

$$\begin{aligned} (\mathbf{e}^T, \mathbf{e}^T, \mathbf{0}^T, \mathbf{0}^T) \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{c} \\ \mathbf{d} \end{pmatrix} &\to \min_{\mathbf{u},\mathbf{v},\mathbf{c},\mathbf{d}} \\ (-\mathbf{I} \quad \mathbf{I} \quad \mathbf{X} \quad -\mathbf{X}) \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{c} \\ \mathbf{d} \end{pmatrix} &= \mathbf{y} \\ \mathbf{u},\mathbf{v},\mathbf{c},\mathbf{d} &\geq \mathbf{0}. \end{aligned} \tag{2}$$

Uočimo da za svaki $i \in \{1,\ldots,m\}$ varijable u_i i v_i u sustavu uvjeta (2) stoje uz dva linearno zavisna stupca matrice $(-\mathbf{I} \quad \mathbf{I} \quad \mathbf{X} \quad -\mathbf{X})$ pa oba ne mogu biti ispovremeno bazične dopustive varijable odnosno mora vrijediti $u_i v_i = 0$, $i = 1,\ldots,m$. Prema tome $|u_i - v_i| = u_i + v_i$, $i = 1,\ldots,m$, što znači da je problem linearnog programiranja (2) ekvivalentan LAD problemu (1).

3 Karakterizacijski teorem

LAD problem (1) specijalni je slučaj općenitijeg problema aproksimacije, koji glasi:

Za zadanu matricu $\mathbf{X} \in \mathbb{R}^{m\times n}$ te vektor $\mathbf{y} \in \mathbb{R}^m$ treba odrediti vekotor $\mathbf{a} \in \mathbb{R}^n$ takav da je

$$\|\mathbf{r}(\mathbf{a})\| = \|\mathbf{X}\mathbf{a} - \mathbf{y}\| \to \min_{\mathbf{a}\in\mathbb{R}^n}, \tag{3}$$

pri čemu je $\|\cdot\| : \mathbb{R}^m \to \mathbb{R}$ bilo koja norma.

U ovoj točki navest ćemo teorem koji daje karakterizaciju rješenja problema (3). U tu svrhu trebat će nam nekoliko pomoćnih pojmova.

a) **Dualna norma**

Neka je $r \in \mathbb{R}^m$ te $\|\cdot\| : \mathbb{R}^m \to \mathbb{R}$ norma na $\mathbb{R}^m$. Definiramo

$$\|\mathbf{r}\|^D := \max_{\|\mathbf{v}\|=1} \mathbf{r}^T\mathbf{v}.$$

Može se pokazati da je funkcija $\|\cdot\|^D : \mathbb{R}^m \to \mathbb{R}$ također norma na $\mathbb{R}^m$ te je zovemo dualna norma norme $\|\cdot\|$. Također vrijede sljedeće tvrdnje:

i) $\|\mathbf{r}\|_1^D = \|\mathbf{r}\|_\infty$, $\|\mathbf{r}\|_\infty^D = \|\mathbf{r}\|_1$

ii) $\left(\|\mathbf{r}\|^D\right)^D = \|\mathbf{r}\|$

iii) $\|\mathbf{r}\|_p^D = \|\mathbf{r}\|_q$, $\frac{1}{p} + \frac{1}{q} = 1$, $p > 1$.

iv) $|\mathbf{r}^T\mathbf{v}| \le \|\mathbf{r}\|^D\|\mathbf{v}\|$, $\quad |\mathbf{r}^T\mathbf{v}| \le \|\mathbf{r}\|\|\mathbf{v}\|^D$

b) **Karakterizacijski skup**

Vektoru $\mathbf{r} \in \mathbb{R}^m$ i normi $\|\cdot\|$ pridružimo skup

$$V_{\|\cdot\|}(\mathbf{r}) = \{v \in \mathbb{R}^m : \|\mathbf{r}\| = \mathbf{r}^T\mathbf{v},\ \|\mathbf{v}\|^D = 1\}.$$

Skup $V_{\|\cdot\|}(\mathbf{r})$ zovemo karakterizacijski skup pridružen vektoru $\mathbf{r}$ i normi $\|\cdot\|$. Vrijede sljedeće tvrdnje

i) $V_{\|\cdot\|_1}(\mathbf{r}) = \left\{\mathbf{v} \in \mathbb{R}^m : v_i = \begin{cases} \text{sign}(r_i), & \text{ako je } r_i \ne 0 \\ \text{bilo koji broj iz intervala } [-1,1], & \text{inače} \end{cases}\right\}.$

ii) $V_{\|\cdot\|_2}(\mathbf{r}) = \frac{\mathbf{r}}{\|\mathbf{r}\|_2}$, $\mathbf{r} \ne \mathbf{0}$ te $V_{\|\cdot\|_2}(\mathbf{0}) = \{\mathbf{v} \in \mathbb{R}^m : v_1^2 + \ldots + v_m^2 = 1\}$

iii) $V_{\|\cdot\|_\infty}(\mathbf{r}) = \text{conv}\{\text{sign}(r_i)\mathbf{e_i} : |r_i| = \|\mathbf{r}\|_\infty\}$, pri čemu je $\mathbf{e_i}$ $i-$ti koordinatni vektor.

Vrijedi sljedeći karakterizacijski teorem

Teorem 1 *([1])Vektor* $\mathbf{a} \in \mathbb{R}^n$ *rješenje je problema (3) onda i samo onda ako postoji vektor* $\mathbf{v} \in V_{\|\cdot\|}(\mathbf{r}(\mathbf{a}))$ *takav da je* $\mathbf{X}^T\mathbf{v} = \mathbf{0}$.

4 Primjene Teorema 1 na LAD problem

Promatrajmo LAD problem (1). Neka su $I = \{1,\ldots,m\}$ te $Z(\mathbf{a}) = \{i \in I : r_i(\mathbf{a}) = 0\}$. Vrijedi sljedeći teorem.

Teorem 2 *Ako je rang*$(\mathbf{X}) = r$, *onda postoji rješenje LAD problema (1) sa svojstvom da* $Z(\mathbf{a})$ *sadrži barem* r *indeksa.*

Primjedba 1 *Primijenimo li Teorem 2 na Primjer 1 dobivamo poznatu tvrdnu da uvijek postoji najbolji LAD pravac* $y = a^*x + b^*$ *koji prolazi kroz barem dvije točke podataka* (x_μ, y_μ) *i* (x_ν, y_ν), $\mu,\nu \in I$, $I = \{1,\ldots,m\}$, $\mu \ne \nu$.

Nadalje vrijedi sljedeća karakterizacija:

Teorem 3 *Neka su* (x_i, y_i), $i = 1, \ldots, m$ *takvi da je* $x_i \neq x_j$ *za* $i \neq j$. *Pravac* $y = a^*x + b^*$ *koji prolazi točno dvijema točkama podataka* (x_μ, y_μ) *i* (x_ν, y_ν), $\mu, \nu \in I$, $I = \{1, \ldots, m\}$, $\mu \neq \nu$ *je najbolji LAD pravac onda i samo onda ako vrijedi*

$$\left| \sum_{i \in \setminus \{\mu,\nu\}} sign(a^*x_i + b^* - y_i) \frac{x_i - x_\mu}{x_\mu - x_\nu} \right| \leq 1 \ i \ \left| \sum_{i \in \setminus \{\mu,\nu\}} sign(a^*x_i + b^* - y_i) \frac{x_i - x_\nu}{x_\mu - x_\nu} \right| \leq 1, \tag{4}$$

ili u ekvivalentnom obliku

$$\max_{x \in \{x_\mu, x_\nu\}} \left| \sum_{i \in \setminus \{\mu,\nu\}} (x_i - x) sign(a^*x_i + b^* - y_i) \right| < |x_\mu - x_\nu|. \tag{5}$$

Teorem 4 *Neka su* (x_i, y_i, z_i), $i = 1, \ldots, m$ *podaci takvi da* (x_i, y_i) *ne leže na nekom pravcu. Ravnina* $z = a^*x + b^*y + c^*$ *koji prolazi točno trima točkama podataka* (x_μ, y_μ, z_μ) *i* (x_ν, y_ν, z_ν), (x_k, y_k, z_k) $\mu, \nu, k \in I$, $I = \{1, \ldots, m\}$, *je najbolja LAD ravnina onda i samo onda ako vrijede sljedeće tri nejednakosti*

$$\left| \sum_{i=1}^m sign(a^*x_i + b^*y_i + c^* - y_i) \big[(y_k - y_\nu)x_i + (x_\nu - x_k)y_i + (x_k y_\nu - x_\nu y_k) \big] \right| \leq |x_\mu(y_\nu - y_k) + x_\nu(y_k - y_\mu) + x_k(y_\mu - y_\nu)|$$

$$\left| \sum_{i=1}^m sign(a^*x_i + b^*y_i + c^* - y_i) \big[(y_k - y_\mu)x_i + (x_\mu - x_k)y_i + (x_k y_\mu - x_\mu y_k) \big] \right| \leq |x_\mu(y_\nu - y_k) + x_\nu(y_k - y_\mu) + x_k(y_\mu - y_\nu)|$$

$$\left| \sum_{i=1}^m sign(a^*x_i + b^*y_i + c^* - y_i) \big[(y_\nu - y_\mu)x_i + (x_\mu - x_{qnu})y_i + (x_\nu y_\mu - x_\mu y_\nu) \big] \right| \leq |x_\mu(y_\nu - y_k) + x_\nu(y_k - y_\mu) + x_k(y_\mu - y_\nu)|$$

Literatura

[1] G. E. Watson, *Approximation Theory and Numerical Methods*, John Wiley&Sons, New York, 1980.