10. Singularni razcep

Bor Plestenjak

NLA

4. maj 2010

Singularni razcep

Izrek

Za vsako matriko $A \in \mathbb{R}^{m\times n}$, $m \geq n$, obstaja singularni razcep (SVD)

$$A = U\Sigma V^T,$$

kjer sta $U \in \mathbb{R}^{m\times m}$ in $V \in \mathbb{R}^{n\times n}$ ortogonalni matriki in je $\Sigma \in \mathbb{R}^{m\times n}$ oblike

$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & \\ & & \end{bmatrix},$$

kjer so $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ singularne vrednosti A.

Stolpci $U = [u_1 \ \cdots \ u_m]$ so levi, $V = [v_1 \ \cdots \ v_n]$ pa desni singularni vektorji.

V primeru $n < m$ dobimo SVD tako, da transponiramo SVD za A^T.

Lastnosti singularnega razcepa

Geometrijski pomen

A predstavlja preslikavo iz $\mathbb{R}^n$ v $\mathbb{R}^m$. Z ortogonalnima transformacijama baz U v $\mathbb{R}^m$ in V v $\mathbb{R}^n$ se A spremeni v diagonalno matriko, saj potem velja

$$Av_i = \sigma_i u_i, \quad i = 1, \ldots, n.$$

$$A = U\Sigma V^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T & V_2^T \end{bmatrix}, \qquad S = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$$

Lastnosti singularnega razcepa:

- r se ujema z $\mathrm{rang}(A)$,
- stolpci V_2 tvorijo bazo za $\ker(A)$,
- stolpci U_1 tvorijo bazo za $\mathrm{im}(A)$,
- stolpci U_2 tvorijo bazo za $\ker(A^T)$,
- stolpci V_1 tvorijo bazo za $\mathrm{im}(A^T)$.

Lastnosti singularnega razcepa

Velja:

1. Lastne vrednosti A^TA so $\sigma_1^2, \ldots, \sigma_n^2$. Lastni vektorji A^TA so desni singularni vektorji $v_1, \ldots, v_n$.
2. Lastne vrednosti AA^T so $\sigma_1^2, \ldots, \sigma_n^2, \underbrace{0, \ldots, 0}_{m-n}$. Lastni vektorji AA^T so levi singularni vektorji $u_1, \ldots, u_m$.
3. Če je $A = A^T$, potem se da A diagonalizirati kot $A = UDU^T$, $U^TU = I$. SVD za A je $A = U\Sigma V^T$ za $\sigma_i = |\lambda_i|$ in $v_i = \mathrm{sign}(\lambda_i)u_i$ (tu je $\mathrm{sign}(0) = 1$).

Skrčeni singularni razcep

$A = \widetilde{U}\widetilde{\Sigma}V^T$, kjer je $\widetilde{U}$ matrika $m \times n$ z ortonormiranimi stolpci, $\widetilde{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$, V pa je $n \times n$ ortogonalna matrika.

$\widetilde{U}$ se ujema s prvimi n stolpci matrike U, $\widetilde{\Sigma}$ pa je zgornji kvadrat Σ v singularnem razcepu $A = U\Sigma V^T$.

SVD in predoločeni sistemi

Lema

Če je $A \in \mathbb{R}^{m \times n}$, $m \geq n$, $\mathrm{rang}(A) = n$, *potem je minimum* $\|Ax - b\|_2$ *dosežen pri*

$$x = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i.$$

Psevdoinverz

Definicija

Za matriko $A \in \mathbb{R}^{m\times n}$, $m \geq n$, $\mathrm{rang}(A) = n$, *definiramo (Moore–Penroseov) psevdoinverz* $A^+ \in \mathbb{R}^{n\times m}$ *kot* $A^+ = (A^T A)^{-1} A^T$.

Rešitev predoločenega sistema polnega ranga $Ax = b$ po m.n.k. je $x = A^+ b$.

Definicija

Matrika X *je psevdoinverz matrike* A *natanko tedaj, ko izpolnjuje Moore–Penroseove pogoje:*

1) $AXA = A$,
2) $XAX = X$,
3) $(AX)^T = AX$,
4) $(XA)^T = XA$.

Psevdoinverz in singularni razcep

Izrek

Naj bo $A \in \mathbb{R}^{m\times n}$, $m \geq n$, $\mathrm{rang}(A) = r$ *in* $A = U\Sigma V^T$*, kjer je*

$$U = \begin{array}{c} \begin{array}{cc} r & m-r \end{array} \\ \begin{pmatrix} U_1 & U_2 \end{pmatrix} \end{array}, \quad V = \begin{array}{c} \begin{array}{cc} r & n-r \end{array} \\ \begin{pmatrix} V_1 & V_2 \end{pmatrix} \end{array}, \quad \Sigma = \begin{array}{c|cc} & r & n-r \\ \hline r & S & 0 \\ m-r & 0 & 0 \end{array}$$

in $S = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$. *Potem je*

$$A^+ = V\Sigma^+ U^T,$$

kjer je

$$\Sigma^+ = \begin{array}{c|cc} & r & m-r \\ \hline r & S^{-1} & 0 \\ n-r & 0 & 0 \end{array}.$$

Psevdoinverz in defekten problem najmanjših kvadratov

Če A ni polnega ranga, pravimo, da ima *defekten rang*. V primeru defektnega ranga rešitev, ki minimizira $\|Ax - b\|_2$ ni enolična, saj lahko x prištejemo poljuben $z \in \ker A$. V takšnem primeru pravimo, da je izmed vseh vektorjev x, ki minimizirajo $\|Ax - b\|_2$, rešitev tisti, ki ima minimalno normo.

Trditev

Če je $A \in \mathbb{R}^{m \times n}$, $m \geq n$, $\mathrm{rang}(A) = r < n$ in $A = U\Sigma V^T$ singularni razcep, potem ima izmed vseh vektorjev x, ki minimizirajo normo $\|Ax - b\|_2$, najmanjšo normo $x = A^+ b$ oziroma

$$x = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i$$

in

$$\|Ax - b\|_2^2 = \sum_{i=r+1}^{m} (u_i^T b)^2.$$

Aproksimacija z matriko nižjega ranka

Če je $A = U\Sigma V^T$ singularni razcep A in je $\mathrm{rang}(A) = r$, potem velja

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T.$$

Izrek (Eckart–Young–Mirsky)

Naj bo $A = U\Sigma V^T$ *SVD matrike* $A \in \mathbb{R}^{m\times n}$, $m \geq n$, *in* $\mathrm{rang}(A) > k$. *Naj bo*

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T.$$

Potem velja

$$\min_{\mathrm{rang}(B)=k} \|B - A\|_2 = \|A_k - A\|_2 = \sigma_{k+1}$$

in

$$\min_{\mathrm{rang}(B)=k} \|B - A\|_F = \|A_k - A\|_F = \left(\sum_{i=k+1}^{n} \sigma_i^2\right)^{1/2}.$$

Uporaba: Aproksimacija slik

Sliko lahko predstavimo z matriko A, katere elementi predstavljajo nivo sivine.

Če namesto A vzamemo najboljšo aproksimacijo z matriko ranga k, potem namesto mn podatkov potrebujemo le $(m+n)k$ podatkov za $[u_1 \ \cdots \ u_k]$ in $[\sigma_1 v_1 \ \cdots \ \sigma_k v_k]$.

Drugačen pogled na predoločen sistem

Pri reševanju predoločenega sistema po m.n.k. iščemo x, ki minimizira $\|Ax - b\|_2$.

To si lahko predstavljamo tudi drugače:

Če označimo ostanek $r = Ax - b$, potem za vsak x velja $Ax = b + r$, torej $b + r$ leži v $\mathrm{im}(A)$ in x je točna rešitev $Ax = b + r$.

Iskanje rešitve po m.n.k. je torej ekvivalentno iskanju minimalne motnje δb desne strani, da je $b + \delta b \in \mathrm{im}(A)$.

Potem je rešitev linearnega sistema $Ax = b + \delta b$ iskana rešitev po m.n.k..

Totalni najmanjši kvadrati

Iskanje rešitve po m.n.k. je ekvivalentno iskanju minimalne motnje δb desne strani, da je $b + \delta b \in \mathrm{im}(A)$.

To je smiselno, če so podatki taki, da so napake prisotne le v vektorju b.

Če pa so napake lahko prisotne tudi v matriki A, je bolje iskati najbližji eksaktno rešljiv sistem. Pri totalnih najmanjših kvadratih iščemo matriko $\widetilde{A}$ in vektor $\widetilde{b}$, da je $\widetilde{b} \in \mathrm{im}(\widetilde{A})$ in je razlika $\left\| \begin{bmatrix} \widetilde{A} & \widetilde{b} \end{bmatrix} - \begin{bmatrix} A & b \end{bmatrix} \right\|_F$ minimalna.

SVD in totalni najmanjši kvadrati

Iščemo $\widetilde{A}$ in $\widetilde{b}$, da je $\widetilde{b} \in \mathrm{im}(\widetilde{A})$ in razlika $\left\| \begin{bmatrix} \widetilde{A} & \widetilde{b} \end{bmatrix} - \begin{bmatrix} A & b \end{bmatrix} \right\|_F$ minimalna.

Predpostavimo, da je A polnega ranga in $b \notin \mathrm{im}(A)$, kar pomeni, da je matrika $\begin{bmatrix} A & b \end{bmatrix}$ ranga $n+1$. Iz $\widetilde{b} \in \mathrm{im}(\widetilde{A})$ sledi, da ima matrika $\begin{bmatrix} \widetilde{A} & \widetilde{b} \end{bmatrix}$ rang n.

Naj bo $\begin{bmatrix} A & b \end{bmatrix} = U\Sigma V^T$ singularni razcep in $\sigma_n > \sigma_{n+1} > 0$. Matrika ranga n, ki je v Frobeniusovi normi najbližja $\begin{bmatrix} A & b \end{bmatrix}$, je enolična in enaka

$$\begin{bmatrix} \widetilde{A} & \widetilde{b} \end{bmatrix} = \begin{bmatrix} A & b \end{bmatrix} - \sigma_{n+1} u_{n+1} v_{n+1}^T.$$

Sistem $\widetilde{A}\widetilde{x} = \widetilde{b}$ je ekvivalenten $\begin{bmatrix} \widetilde{A} & \widetilde{b} \end{bmatrix} \begin{bmatrix} \widetilde{x} \\ -1 \end{bmatrix} = 0$. Vsi vektorji iz jedra $\begin{bmatrix} \widetilde{A} & \widetilde{b} \end{bmatrix}$ so oblike αv_{n+1} za $\alpha \in \mathbb{R}$. Tako lahko $\widetilde{x}$ izrazimo iz vektorja v_{n+1} kot

$$\widetilde{x} = \frac{-1}{v_{n+1,n+1}} \begin{bmatrix} v_{1,n+1} \\ \vdots \\ v_{n,n+1} \end{bmatrix}.$$

Rešitev je dobro definirana, če je $v_{n+1,n+1} \neq 0$.

Obstoj rešitve totalnih najmanjših kvadratov

Če so $\sigma'_1 \geq \cdots \geq \sigma'_n$ singularne vrednosti A in $\sigma_1 \geq \cdots \geq \sigma_{n+1}$ singularne vrednosti $\begin{bmatrix} A & b \end{bmatrix}$ se izkaže, da je

$$\sigma'_n > \sigma_{n+1}$$

potreben in zadosten pogoj za obstoj in enoličnost rešitve po metodi TLS. Velja namreč

$$\sigma'_n > \sigma_{n+1} \iff \sigma_n > \sigma_{n+1} \quad \text{in} \quad v_{n+1,n+1} \neq 0.$$

Uporaba: internetni iskalniki

V iskalnik vpišemo ključne besede in v kratkem času dobimo seznam ustreznih strani. Kako lahko iskalniki tako hitro poiščejo pravi seznam?

Podatki o straneh so shranjeni v ogromni $m \times n$ matriki A, kjer vrstice predstavljajo dokumente, stolpci pa ključne besede. V preprostem modelu element a_{ij} matrike A vsebuje število pojavitev j-te ključne besede v i-tem dokumentu.

Iz vnešenih ključnih besed sestavimo vektor $y \in \mathbb{R}^n$, ki ima na j-tem mestu 1, če iščemo j-to ključno besedo, sicer pa 0. Produkt $x = Ay$ je vektor iz $\mathbb{R}^m$. Element x_i predstavlja, kako dobro se i-ti dokument ujema z našo poizvedbo. Iskalnik kot rezultat vrne kratek seznam dokumentov, za katere so vrednosti x_i največje.

Matrika A je ogromna, saj je $m \approx 10^{11}$ in $n \approx 10^6$. Kljub razpršenosti je izračun produkta $x = Ay$ ozko grlo poizvedbe.

Rešitev: A aproksimiramo z matriko A_k majhnega ranga k (npr. $k = 20$) in kot približek za x vzamemo

$$x_k = A_k y = \sum_{i=1}^{k} \sigma_i (v_i^T y) u_i.$$