Alkalmazott algebra - SVD

Ivanyos Gábor

2011 ősz

SVD - szinguláris értékek szerinti felbontás

Poz. szemidefinit mátrixok spektrálfelbontásának általánosítása nem feltétlenül négyzetes mátrixokra

"Közelítő" webkeresés egy modellje

- dokumentumok (weblapok) $\sim$ vektorok,
- a koordináták a (lényeges) szavaknak felelnek meg:
- A koordináták kiszámítására példák:
  - egyszerű incidencia: 0 vagy 1
  - fontosság szerinti súlyok
    Pl. relatív említési gyakoriságból kalkulált
    Számíthat az említés módja is: címben, hivatkozásban, szövegben
- keresőkérdés $\sim$ dokumentum-vektor
- Feladat: a kérdés vektorához minél jobban illeszkedő: "hasonló", "közeli" dokumentum-vektor keresése
- Nem pontos illeszkedés: lehet, hogy a legjobb találat nem is tartalmazza a kérdés egy szavát se, inkább
- A kérdés "jelentéséhez" van köze

LSI - mögöttes szemantikájú indexelés

- Latent Semantic Indexing
- m dokumentum (pl. weblap), n szó
- $A = (\alpha_{ij})$ $m \times n$-es, $\alpha_{ij} \sim j$. szó az i. dok-ban
- A sorai $\sim$ dokumentumok
- A oszlopai: szavak "profilja"
- **Feltevés:**
  - a háttérben (általunk nem ismert) fogalmak, jelentések állnak
  - a szavak jól közelíthetők fontosabb jelentések **elegyével**
  - a jelentéseket a dokumentumokból "tanulhatjuk" meg
  - a közelítés hibája: "zaj", "lényegtelen", "esetleges" dolgok

LSI

- $A = (\alpha_{ij})$ $m \times n$-es, $\alpha_{ij} \sim j$. szó az i. dok-ban
- **Formálisan:**
  - léteznek $f_1 = \begin{pmatrix} f_{11} \\ f_{21} \\ \vdots \\ f_{m1} \end{pmatrix}, \ldots, f_k = \begin{pmatrix} f_{1k} \\ f_{2k} \\ \vdots \\ f_{mk} \end{pmatrix} \in \mathbb{R}^m$ oszlopvektorok (a jelentések profiljai), hogy
  - A oszlopai jól közelíthetők $f_1, \ldots, f_k$ **lineáris kombinációival**, azaz létezik olyan B $m \times k$-as mátrix, hogy

$$A \approx FB,$$

ahol $F = (f_{ij})$.
B i-edik sora az A-nak az i-edik oszlopát közelítő kombináció együtthatóiból áll

Alacsony rangú közelítés

- **Áll.:** Legyen A' egy $m \times n$-es mátrix. Ekkor A' rangja $\leq k \Leftrightarrow$ léteznek olyan B $k \times n$-es és F $m \times k$-as mátrixok, amelyekre

$$A' = FB.$$

Biz.: A' rangja $\leq k \Leftrightarrow$ létezik olyan F $m \times k$-as, hogy A' oszlopai megkaphatók F oszlopainak lineáris kombinációiként Emellett B i-edik sora az A' i-edik oszlopát előállító kombináció együtthatóiból áll.

- **Tehát:** a "lényeg" $\sim$ A alacsony rangú közelítése

Eredet: főkomponensanalízis

- K. Pearson 1901
- $A = (\alpha_{ij})$ $m \times n$-es, $\alpha_{ij} \sim j$. adat értéke az i. mérés során

  Pl. kérdőív kérdéseire adott válaszok, skálázva
- Oszlopok: $\sim$ val. változók mért adatai $a^j = \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{mj} \end{pmatrix}$
- Sorok: egy-egy mérésre az adatok sora $a_i = (\alpha_{i1} \quad \dots \quad \alpha_{in})$
- Feltesszük: $\sum_{i=1}^m \alpha_{ij} = 0 \sim$ 0-ra normált empirikus átlagok
- $|a_j|^2 = \sum_{i=1}^m \alpha_{ij}^2$: $\sim$ empirikus szórásnégyzet
- a legfontosabb "rejtett összetevő": abban az irányban vett adatok, amerre a leginkább szórnak.

Főkomponensanalízis

- u **irányú adatok:** Legyen $u = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$ egységvektor ($u^T u = 1$).

  Az i-edik mérés u irányú adata a_i^T vektor u irányú (u-val párhuzamos) komponensének az együtthatója, azaz $a_i u = \sum_{j=1}^{n} \alpha_{ij}\mu_j$. Az m darab u irányú adat együtt: Au.
- fő irány: az az u, amelyre $|Au|^2 = \sum_{i=1}^{m}(a_i u)^2$ maximális
- a megfelelő "rejtett változó" $\sim$ az Au oszlopvektor
- a következő fő irány: az u-ra merőlegesek közül a legnagyobb vetületet adó irány, stb...

Főkomponensanalízis és szinguláris értékek

- $A = (\alpha_{ij})$ tetszőleges $m \times n$-es mátrix ($\sum_{i=1}^{m} \alpha_{ij} = 0$ nem kell.)
- $u_1 \in \mathbb{R}^n$ olyan egységvektor, amelyre
$$|Au_1|^2 \text{ a lehető legnagyobb}$$
- Ha $u_1, \ldots, u_s$ már megvan, u_{s+1} egy olyan egységvektor $\{u_1, \ldots, u_s\}^{\perp}$-ből, amelyre
$$|Au_{s+1}|^2 \text{ a lehető legnagyobb}$$
- Mi lehet $u_1, \ldots, u_n$?

Főkomponensanalízis és szinguláris értékek

- **Észrevétel:** $|Au|^2 = u^T A^T A u$, és $A^T A$ poz. szemidefinit
  utóbbi biz.: $(A^T A)^T = A^T A$, tehát $A^T A$ szimm.
  $v^T(A^T A)v = (Av)^T(Av) \geq 0$. (Egyenlőség v ker A-ra).
- **Köv.:** $\exists$ M ortog., hogy $M^T A^T A M = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$. Feltehető, hogy $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$.
- (Elnevezés: **szinguláris értékek**)
- **Áll.:** $|u| = 1$ esetén $|Au| \leq \sigma_1$ és ha $|Au| = \sigma_1$, akkor u sajátvektora $A^T A$-nak σ_1^2 sajátértékkel. (zh091030, 5. fa.)

Biz.: Legyen M mit fent és $v := M^T u = \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_n \end{pmatrix}$. Ekkor $u = Mv$ és

$|Au|^2 = |AMv|^2 = v^T M^T A^T A M v = \sum_{i=1}^n \nu_i^2 \sigma_i^2 \leq \sum_{i=1}^n \nu_i^2 \sigma_1^2 = \sigma_1^2$, és ha $\nu_i \neq 0$ olyan i-re, amelyre $\sigma_i < \sigma_1$, akkor $<$ áll. Különben $u = Mv$ az M olyan oszlopainak lineáris kombinációja, amelyek sajátvektorok σ_1^2 sajátértékkel.

Főkomponensanalízis és szinguláris értékek

- **Áll.:** Tfh. $|u| = 1$ és $u \perp M$ első k oszlopára. Ekkor $|Au| \leq \sigma_{k+1}$ és ha $|Au| = \sigma_k$, akkor u sajátvektora A^TA-nak σ_k^2 sajátértékkel.

  Biz.: Legyen $v = M^Tu = \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_n \end{pmatrix}$. A merőlegességi feltétel miatt
  $\nu_1 = \ldots = \nu_k = 0$, Ekkor $Au = AMv$ és
  $|Au|^2 = \sum_{i=k+1}^{n} \nu_i^2 \sigma_i^2 \leq \sum_{i=k+1}^{n} \nu_i^2 \sigma_{k+1}^2 = \sigma_{k+1}^2$. Az egyenlőség vizsgálata is hasonló a $k = 1$ esethez.

- **Köv.:** A fő irányok A^TA sajátvektorai, azaz (lényegében) M oszlopai.

Szinguláris értékek

A $m \times n$-es **valós** mátrix **Megj:** (a komplex eset hasonlóan működik)

- **Def.:** A **szinguláris értékei** $A^T A$ (nem nulla) sajátértékeinek $\sqrt{}$-e
- **Áll.:** $A^T A$ rangja $=$ A rangja
  Biz.:
  Észrevétel: $\ker A^T \cap A\mathbb{R}^n = 0$.
  Észrv. biz.: Tfh. $v \in \ker A^T \cap A\mathbb{R}^n$: $A^T v = 0$ és $v = Au$.
  Ekkor $u^T A^T A u = u^T A^T v = 0$, így $Au = 0$.
  Dimenziótétel A^T-nek $A\mathbb{R}^n$-re való megszorítására+Észrv.:
  $\dim A^T A\mathbb{R}^n = \dim A\mathbb{R}^n - 0 = \dim A\mathbb{R}^n$, azaz

  $A^T A$ rangja $=$ A rangja.
- **Köv.:** (poz.) szinguláris értékek száma $=$ rang

A és A^T szing. értékei

- **Áll.:** Ha σ (poz.) szing. értéke A-nak, akkor A^T-nak is (és viszont).
- **Részletesebben:** Ha v sajátvektora $A^T A$-nak $\sigma^2 \neq 0$ sajátértékkel, akkor Av sajátvektora AA^T-nak szintén σ^2 sajátértékkel.
  Biz.: Tfh. $0 \neq v$ és $A^T Av = \sigma^2 v$. Ekkor
  $(AA^T)(Av) = A(A^T A)v = A\sigma^2 v = \sigma^2 Av$.

  $v = \frac{1}{\sigma^2} A^T Av$, így $Av \neq 0$.
- **Megj.** A (poz.) szing. értékek multiplicitásai is ugyanazok.
  Biz.: A $\frac{1}{\sigma}A$ az $A^T A$ σ^2-sajátalterét az AA^T σ^2-sajátalterébe viszi. A $\frac{1}{\sigma}A^T$ pedig fordítva.

  A két leképezés megszorítása a σ^2-sajátalterekre inverzei egymásnak.

SVD - A fő lemma

Lemma: Tetsz. A $m \times n$-es valós mátrixhoz léteznek M $n \times n$-es és M' $m \times m$-es ortogonális mátrixok, hogy:

$$M'^{T} A M = \Sigma' = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ & & & \end{pmatrix} \text{ illetve } \begin{pmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_n & \end{pmatrix}$$

Az első az $m > n$ esetet, a második az $m < n$ esetet ábrázolja

$\sigma_1 \geq \ldots \geq \sigma_n$: $A^T A$ sajátértékeinek nemnegatív $\sqrt{\ }$-e

SVD - a fő lemma bizonyítása

- Legyen M olyan ortog., amelyre

$$M^T A^T A M = \operatorname{diag}(\sigma_1^2, \dots, \sigma_n^2)$$

- **Megj:** főkomponensanalízisnél M oszlopai: a fő irányok, AM oszlopai: a fő irányok mentén vett értékek vektorai
- Legyenek AM oszlopai $w_1, \dots, w_n$.
- $(AM)^T(AM) = \operatorname{diag}(\sigma_1^2, \dots, \sigma_n^2)$ miatt
- $w_i^T w_j = \begin{cases} \sigma_i^2, & \text{ha } i = j \\ 0 & \text{ha } i \neq j \end{cases}$

SVD - a fő lemma bizonyítása II

- $w_i^T w_j = \begin{cases} \sigma_i^2, & \text{ha } i = j \\ 0 & \text{ha } i \neq j \end{cases}$
- Tfh. $\sigma_1 \geq \ldots \geq \sigma_r > 0$, de $\sigma_{r+1} = \ldots = \sigma_n = 0$.
- **Megj.:** $i > r$-re $w_i = 0$ (mert $w_i^T w_i = \sigma_i^2 = 0$)
- Legyenek $i \leq r$-re $v_i = \frac{1}{\sigma_i} w_i$
- Ha $m > r$, egészítsük ki $v_{r+1}, \ldots, v_m$-mel $\mathbb{R}^m$ ortonormált bázisává.
- $v_i^T w_j = \begin{cases} \sigma_i, & \text{ha } i = j < r \\ 0 & \text{egyébként} \end{cases}$
- Legyenek M' oszlopai: $v_1, \ldots, v_m$. Ekkor M' ortog. és $M'^T AM = \Sigma'$

Teljes SVD

Tétel: Tetsz. A $m \times n$-es valós mátrixhoz léteznek M $n \times n$-es és M' $m \times m$-es ortogonális mátrixok, hogy

$$A = M'\Sigma' M^T, \qquad \text{ahol}$$

$$\Sigma' = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ & & & \\ & & & \end{pmatrix} \text{ illetve } \begin{pmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_n & & \end{pmatrix}$$

Biz.: a fő lemma egyenlőségét M'-vel ill. M^T-vel szorozzuk balról-jobbról.

Redukált SVD

Tétel: Tetsz. A $m \times n$-es valós, r rangú mátrixhoz vannak U' $m \times r$-es, U $n \times r$-es mátrixok, hogy

- $U^T U = U'^T U' = I_r$, azaz U, ill. U' oszlopai ortonormált rendszerek
- $A = U' \Sigma U^T$, ahol

$$\Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{pmatrix}$$

- $\sigma_1 > \ldots > \sigma_r > 0$ A (poz.) sztng. értékei.

Megj. (a komplex eset): $U^* U = U'^* U' = I_r$, $A = U' \Sigma U^*$.
(Ekkor $A^* A$ lesz poz. szemidef. önadj.)

Redukált SVD - bizonyítás

- Teljes SVD: $A = M'\Sigma' M^T$
- Legyenek $M = (U|T)$, $M' = (U'|T')$.

  U ill. U' az M ill. M' első r oszlopa. Nyilván $U^T U = U'^T U' = I_r$.
- $\Sigma' = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}$
- $A = (U'|T') \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U^T \\ T^T \end{pmatrix} = (U'\Sigma|0) \begin{pmatrix} U^T \\ T^T \end{pmatrix} = U'\Sigma U^T$.

SVD: "Egyértelműség" kérdése

- Tfh. $A = U'\Sigma U^T$, ahol $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$, $\sigma_1 \geq \ldots \geq \sigma_r > 0$, U' $m \times r$-es U $n \times r$-es, $U'^T U' = I_r$, $U^T U = I_r$.
- Ekkor $A^T A = U\Sigma U'^T U'\Sigma U^T = U\Sigma^2 U^T$, így $\Sigma^2 = U^T A^T A U$.
- Innen: U oszlopai az $A^T A$ mátrix nem 0 sajátértékeihez tartozó sajátaltereinek ortonormált bázisait adják.

  Egészítsük ki U-t egy $n \times n$-es M ortog. mátrixszá ker $A^T A$-ból vett oszlopvektorokkal. Erre $M^T A^T A M = \Sigma'^2$.
- U': fenti érvelés, $A \leftrightarrow A^T$ cserével.
- **Köv:** ha A pozitív szinguláris értékei páronként kül, akkor U és U' az oszlopok egységszerese (± 1) erejéig egyértelmű.

SVD - példa

- $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$
- $A^T A = \begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}$, $M^T A^T A M = \Sigma'^2 = \begin{pmatrix} 6 & 0 \\ 0 & 0 \end{pmatrix}$, $\Sigma' = \begin{pmatrix} \sqrt{6} & 0 \\ 0 & 0 \end{pmatrix}$,
- $M = \begin{pmatrix} +\frac{1}{\sqrt{2}} & +\frac{1}{\sqrt{2}} \\ +\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$,
- $AM = \begin{pmatrix} \sqrt{2} & 0 \\ \sqrt{2} & 0 \\ \sqrt{2} & 0 \end{pmatrix}$
- $M' = \begin{pmatrix} \frac{1}{\sqrt{3}} & c_{12} & c_{13} \\ \frac{1}{\sqrt{3}} & c_{22} & c_{23} \\ \frac{1}{\sqrt{3}} & c_{32} & c_{33} \end{pmatrix}$
- $\Sigma = \sqrt{6}$, $U = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$, $U' = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}$, tehát $A = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix} \sqrt{6} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$.

SVD szimmetrikus mátrixokra

$A^T = A$

- a szinguláris értékek A sajátértékeinek abszolút értékei (multiplicitással).
- Tfh. $A = U \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} U^T$,
  ahol U $n \times n$-es ortogonális; $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$.
- Ekkor A egy lehetséges teljes SVD-je

$$A = U' \begin{pmatrix} |\lambda_1| & & & \\ & |\lambda_2| & & \\ & & \ddots & \\ & & & |\lambda_n| \end{pmatrix} U^T,$$

ahol U' i-edik oszlopa U i-edik oszlopának ± 1-szerese, aszerint, hogy $\lambda_i \geq 0$ vagy $\lambda_i < 0$.

A Frobenius-norma

A $m \times n$-es valós

- **Def. (Frobenius-norma-négyzet):** $(=$ eukl. hossznégyzet$)$

$$\|A\|^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2.$$

- **Ugyanaz, mint:** sorok v. oszlopok hossznégyzet-összege
- **Elegáns alak**: $\|A\|^2 = \operatorname{tr} A^T A = \operatorname{tr} AA^T$, ahol
- **Emlék. (nyom, trace):** $B = (b_{ij})$ $n \times n$-esre $\operatorname{tr} B = \sum_{i=1}^{n} b_{ii}$
- **Megj. (komplex mátrixokra):**
  $\|A\|^2 = \sum \sum |a_{ij}|^2 = \operatorname{tr} A^* A = \operatorname{tr} AA^*$
- **Nyom tulajdonsága:** A $m \times n$-es, B $n \times m$-esre:
  $$\operatorname{tr} AB = \operatorname{tr} BA.$$

Normatartó szorzások:

Áll.: Tfh. A $m \times n$-es, B olyan $m' \times m$-es amelynek az oszlopai ortonormált rendszert alkotnak (azaz $B^T B = I_m$), C olyan $n \times n'$-es, amelynek a sorai ortonormált rendszert alkotnak (azaz $CC^T = I_n$). Ekkor

$$\|BA\|^2 = \|A\|^2 = \|AC\|^2.$$

Biz.: A oszlopai $a^1, \ldots, a^n$, BA oszlopai $Ba^1, \ldots, Ba^n$. B az $\mathbb{R}^m$ standard bázisát B képterének egy ortonormált bázisába viszi, tehát B egy izometria $\mathbb{R}^m$ és B képtere között.
Így $|Ba^j|^2 = |a^j|^2$ és $\|BA\|^2 = \sum |Ba^j|^2 = \sum |a^j|^2 = \|A\|$.
A másik $=$-ség innen $A \leftarrow A^T$, $B \leftarrow C^T$ helyettesítéssel.
Másik biz.:

$$\|BA\|^2 = \operatorname{tr} A^T B^T BA = \operatorname{tr} A^T I_m A = \operatorname{tr} A^T A = \|A\|^2.$$

Az Eckart–Young-tétel

- Eckart–Young 1936 (Psychometrika)
- Korábban: Schmidt approximációs tétele 1907 (függvények kontextusában)
- **Redukált SVD:** $A = U'\Sigma U^T$.
- $k \leq r$-re legyenek
  - $U^{(k)}$: U első k oszlopa
  - $U'^{(k)}$: U' első k oszlopa
  - $\Sigma^{(k)} = \mathrm{diag}(\sigma_1, \ldots, \sigma_k)$: Σ bal felső $k \times k$-as része
  - $A^{(k)} := U'^{(k)}\Sigma^{(k)}{U^{(k)}}^T$
- **Tétel:** A-nal $A^{(k)}$ az (egyik) legjobb $\leq k$ rangú közelítése a Frobenius-normában

  (A hiba: $\left\|A - A^{(k)}\right\|^2 = \sum_{i=k+1}^{r} \sigma_i^2$)
- **Megj.:** Mirsky 1960: és még sok más fontos normában is ez a legjobb

Eckart–Young teljes SVD-vel

- **Teljes SVD:** $A = M'\Sigma' M^T$, ahol
- Σ' $m \times n$-es diagonális, $\sigma_1, \ldots, \sigma_r$ nemnulla átlós elemekkel
- $\Sigma'^{(k)}$: írjunk Σ'-ben $\sigma_{k+1}, \ldots, \sigma_r$ helyébe 0-t
- $A^{(k)} = M'\Sigma'^{(k)} M^T$

Eckart–Young bizonyítása

- Az U'^T-vel való balról, U-val jobbról szorzás tartja a Frobenius-normát:
- $\left\|A - A^{(k)}\right\|^2 = \left\|U'^T(A - A^{(k)})U\right\|^2 = \left\|\Sigma - \widetilde{\Sigma^{(k)}}\right\|^2 = \sigma_{k+1}^2 + \ldots + \sigma_r^2$,
  ahol $\widetilde{\Sigma^{(k)}} = \mathrm{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0)$ $r \times r$-es
- Legyen A' egy $\leq k$ rangú $m \times n$-es
- $\|A - A'\|^2 = \left\|U'^T(A - A')U\right\|^2 = \left\|U'^T AU - U'^T A'U\right\|^2 = \|\Sigma - C\|^2$,
  ahol $C := U'^T A'U \leq k$ rangú $r \times r$-es
- Elég tehát a következő spec. esetet igazolni:
- **Spec. eset:** tetsz. $C \leq k$ rangú $r \times r$-esre:

$$\|\mathrm{diag}(\sigma_1, \ldots, \sigma_r) - C\|^2 \geq \sigma_{k+1}^2 + \ldots + \sigma_r^2.$$

Eckart–Young spec. eset bizonyítása

Spec. eset: tetsz. $C \leq k$ rangú $r \times r$-esre:

$$\|\mathrm{diag}(\sigma_1, \ldots, \sigma_r) - C\|^2 \geq \sigma_{k+1}^2 + \ldots + \sigma_r^2.$$

- Legyen $W = \ker C$ egy $r - k$ dimenziós altere
- U_1 $r \times (r-k)$-as oszlopai: W egy ortonormált bázisa
- U_2 $r \times k$-as oszlopai: $W^\perp$ egy ortonormált bázisa
- $(U_1|U_2)$ $r \times r$-es ortogonális
- $D := \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$

Eckart–Young spec eset bizonyítása II

$$
\begin{aligned}
\|D-C\|^2 &= \left\|(U_1|U_2)^T(D-C)(U_1|U_2)\right\|^2 = \left\|\begin{pmatrix} U_1^T \\ U_2^T \end{pmatrix}(D-C)(U_1|U_2)\right\|^2 \\
&= \left\|\begin{pmatrix} U_1^T(D-C) \\ U_2^T(D-C) \end{pmatrix}(U_1|U_2)\right\|^2 = \left\|\left(\begin{array}{c|c} U_1^T(D-C)U_1 & U_1^T(D-C)U_2 \\ \hline U_2^T(D-C)U_1 & U_2^T(D-C)U_2^T \end{array}\right)\right\|^2 \\
&\geq \left\|U_1^T(D-C)U_1\right\|^2 = \left\|U_1^T DU_1 - U_1^T CU_1\right\|^2 = \left\|U_1^T DU_1\right\|^2
\end{aligned}
$$

$\geq$: a mátrix normája $\geq$ a mátrix bal felső blokkjának a normája
utolsó $=$: $U_1^T CU_1 = 0$ mert $CU_1 = 0$ (U_1 oszlopai C magjából)
Elég tehát: $D = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$-re és tetsz olyan U_1 $r \times (r-k)$-asra, amelyre $U_1^T U_1 = I_{r-k}$, teljesül

$$\left\|U_1^T DU_1\right\|^2 \geq \sigma_{k+1}^2 + \ldots + \sigma_r^2$$

Eckart–Young spec eset bizonyítása III

Utolsó lépés $D = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$, U_1 $r \times (r-k)$-as, amelyre $U_1^T U_1 = I_{r-k}$.

$$\left\| U_1^T D U_1 \right\|^2 = \mathrm{tr}\,(U_1^T D U_1)^T U_1^T D U_1 = \mathrm{tr}\, U_1^T D^T D U_1 = \mathrm{tr}\, U_1 U_1^T D^T D = \mathrm{tr}\, U_1 U_1^T D^2 = \sum_{i=1}^{r} \beta_{ii} \sigma_i^2,$$

ahol $U_1 U_1^T = (\beta_{ij})$.

- **Áll.:** $U_1 U_1^T$ egy $r-k$ rangú merőleges vetítés

  $(U_1 U_1^T)^T = U_1 U_1^T$ és $(U_1 U_1^T)^2 = U_1 U_1^T U_1 U_1^T = U_1 I_{n-k} U_1^T = U_1 U_1^T$
- **Innen:** $0 \le \beta_{ii} \le 1$ és $\sum_{i=1}^r \beta_{ii} = r - k$.
- Fenti feltételek mellett: $\sum_{i=1}^r \beta_{ii}\sigma_i^2 \to \min$, ha a legkisebb σ_i^2-eket vesszük a lehető legnagyobb súllyal:
- $\left\| U_1^T D U_1 \right\|^2 = \sum_{i=1}^r \beta_{ii}\sigma_i^2 \ge \sigma_{k+1}^2 + \ldots + \sigma_r^2$, épp ez kellett.

LSI - keresés záró megj.

- **Észrevétel:** Ha w A egy sora, az $A^{(k)}$ mátrixnak a megfelelő sora éppen w vetülete az $A^{(k)}$ sorai által generált altérre.

  Biz.: Ha nem az lenne, kicserélve a sort w vetületével jobb közelítést kapnánk
- Keresőkérdés $\sim$ dokumentum-vektor $\sim$ A sorai.
- **Eljárás:** A keresőkérdés vektorát levetítjük $A^{(k)}$ sorai által gen. altérre.

Az SVD kiszámításáról

- **Naiv módszer:** $A^T A$ sajátértékei a kar. pol.-ból, majd a sajátalterek homogén lin. egyenletrendszerekkel

  Gauss-elimináció
- **Gond:** a Gauss-elimináció numerikus hibákra instabil
- Hatékony **numerikus módszerek** vannak (később tárgyaljuk)

Négyzetgyök

Áll.: Ha A $n \times n$-es valós poz. szemidef., akkor $\exists!$ olyan B $n \times n$-es poz. szemidef., hogy $A = B^2$ (jel.: $B = \sqrt{A}$).

Biz.: Legyen U ortog., hogy $D = U^T A U$ diag. (≥ 0 elemekkel),
$B' = \sqrt{D}$ (elemenként).
B' valós, poz. szemidef. és $B'^2 = D$.
$B = UB'U^T$ is poz. szemidef, továbbá

$$B^2 = UB'U^T UB'U^T = U{B'}^2 U^T = UDU^T = A.$$

Egyértelműség: Tfh $A = B^2$. Legyen U olyan ortog., hogy $U^T BU = B'$ diag. Ekkor $U^T AU = U^T B^2 U = B'^2$ diag.
Innen: $U^T AU$ és B' sajátalterei ugyanazok: $U^T AUv = \lambda v \Leftrightarrow B'v = \sqrt{\lambda} v$
Innen: $(v = Uu)$: $Au = \lambda u \Leftrightarrow UB'U^T u = \sqrt{\lambda} u \Leftrightarrow Bu = \sqrt{\lambda} u$. Így B nem lehet más, mint fenti konstrukció.

Poláris felbontás

Tétel: Tetsz. A $n \times n$-es valós mátrixra $A = PM$, ahol

- M $n \times n$-es ortog.
- P $n \times n$-es poz. szemidef.
- P egyértelmű, és ha A nemszing., akkor M is.

  Biz. Létezés: A teljes SVD-je: $A = U'\Sigma' U^T$, ahol U és U' $n \times n$-es ortog., Σ' pedig egy $n \times n$-es diag. nemneg. Legyen $P = U'\Sigma U'^T$ és $M = U'U^T$.

  P egyért.: $AA^T = PMM^TP = P^2$, így $P = \sqrt{AA^T}$. Ha A invertálható, akkor P is az és $M = P^{-1}A$.

Megj.: hasonlóan, $A = M'P'$, ahol M' unitér és P' poz. szemidef. $P' = P$ akkor és csak akkor, ha $A^TA = AA^T$, azaz ha A normális.
Megj: Komplex $A = PM$, ahol M unitér, P poz. szemidef önadj. 1-dim. komplex eset: komplex számok felbontása $re^{i\phi}$ alakban.

(teljes) SVD a poláris felbontásból

Négyzetes mátrixokra

- $A = PM$
- M ortog.
- P poz. szemidef: $M'^T PM' = \Sigma$ diag,
- $A = U'\Sigma U^T$, ahol $U' = M'^T$, $U = M^T M'^T$.

Homogén lineáris egyenletrendszerek megoldása SVD-vel

- **A feladat:** $Ax = 0$ megoldása, azaz $\ker A$ számítása.
- **Emlékeztető:** $A^T A$ rangja $=$ A rangja, így (pl. a dimenziótételből)

$$\ker A^T A = \ker A.$$

- **Teljes SVD:** $A = M' \Sigma' M^T$
  $\Sigma' = M'^T A M$, innen
  $\Sigma'^2 = \Sigma'^T \Sigma' = M^T A^T M' M'^T A M = M^T A^T A M$
- $\Sigma'^2 = M^T A^T A M$, azaz M oszlopai $A^T A$ sajátvektoraiból álló ortonormált bázis
- **Köv.:** M utolsó $n - r$ oszlopa: $\ker A^T A = \ker A$ egy ortonormált bázisa

Homogén lin. egyenletrendszerek megoldása legkisebb négyzetes hibával

- Elnevezés: "total least squares"
- $Ax = 0$ közelítő megoldása a legkisebb négyzetes hibával $|x| = 1$ mellett:

$$|Ax|^2 \longrightarrow \min, \text{feltéve } |x| = 1.$$

Feltehető: ker $A = 0$, így n poz. szing. érték van.

- SVD: $A = U'\Sigma U^T$ így $\Sigma^2 = U^T A^T A U$.
- legyenek $v_1, \ldots, v_n$ U oszlopai. Ortonormált bázis, $A^T A v_i = \sigma_i^2 v_i$.
- Legyen $x = \sum \alpha_i v_i$, ahol $|x|^2 = \sum |\alpha_i|^2 = 1$. Ekkor

$$\begin{aligned} |Ax|^2 &= (Ax, Ax) = (x, A^T Ax) = (\sum \alpha_i v_i, \sum \alpha_i \sigma_i^2 v_i) = \sum \sigma_i^2 |\alpha_i|^2 \\ &\geq \sum |\alpha_i|^2 \sigma_n^2 = \sigma_n^2, \qquad \text{(ez, a másik irányban volt a főkomp-nél)} \end{aligned}$$

- az optimum σ_n^2, $x = v_n$-nel el is érhető
- $\sigma_{n-1} > \sigma_n$ esetén az optimális x ± 1-szeres erejéig egyértelmű.

Inhomogén egyenletrendszerek megoldása legkisebb négyzetes hibával

- $Ax = b$ lineáris egyenletrendszer. A $m \times n$-es valós, $b \in \mathbb{R}^n$
- nem mindig van **pontos megoldás:** olyan $x \in \mathbb{R}^n$, amelyre $Ax - b = 0$
- **Közelítő megoldás:** olyan x, amelyre $|Ax - b|^2 \to \min$
- Ilyen x-re Ax éppen b-nek $A\mathbb{R}^n$-re való $\perp$ vetülete

  Tudjuk: altér legközelebbi pontja a vetület
- **Fogjuk látni:** létezik A^+ $n \times m$-es, amelyre
  AA^+ éppen az $A\mathbb{R}^n$-re való $\perp$ vetítés
- (az egyik) **legkisebb négyzetes hibájú magoldás**: $x = A^+b$.

  (mert $Ax = AA^+b$=a vetület)
- az összes mo.: a $\ker A + A^+b$ eltolt altér

A Moore–Penrose-féle pszeudoinverz

- **Def.:** A $m \times n$-es valós mátrixnak egy pszeudoinverze egy olyan A^+ $n \times m$-es valós mátrix, amelyre
  - $AA^+A = A$,
  - $A^+AA^+ = A^+$,
  - AA^+ szimmetrikus,
  - A^+A szimmetrikus.

  (Megj.: komplex A-ra szimm. helyett önadj.)
- **Pszeudoinverz az SVD-ből:** Ha A red. SVD-je $A = U'\Sigma U^T$, akkor $A' = U\Sigma^{-1}U'^T$ (egy) pszeudoinverze A-nak.

  Biz.: $U^TU = U'^TU' = I_r$-ből $AA' = U'\Sigma U^TU\Sigma^{-1}U'^T = U'U'^T$ és $A'A = U\Sigma^{-1}U'^TU'\Sigma U^T = UU^T$.

  Az $AA' = U'U'^T$ és $A'A = UU^T$ a segítségével:
  - $AA'A = U'U'^TA = U'U'^TU'\Sigma U^T = U'\Sigma U^T = A$.
  - $A'AA' = A'U'U'^T = U\Sigma^{-1}U'^TU'U'^T = U\Sigma^{-1}U'^T = A'$
  - $(AA')^T = (U'U'^T)^T = U'U'^T = AA'$.
  - $(A'A)^T = (UU^T)^T = UU^T = A'A$.

Pszeudoinverz és vetítés

Lemma: Tfh. A^+ egy pszeudoinverze A-nak. Ekkor AA^+ az $\mathbb{R}^m$ tér merőleges vetítése A képterére, A^+A pedig az $\mathbb{R}^n$ tér merőleges vetítése A^T képterére.

Biz.:

- AA^+ szimm. és $(AA^+)^2 = AA^+$, azaz egy merőleges vetítés.
  AA^+ képtere nyilván $\subseteq$ A képtere. $A = (AA^+)A$ képtere nyilván $\subseteq$ AA^+ képtere.
  Tehát AA^+ merőleges vetítés A képterére.
- A^+A is merőleges vetítés (hasonló biz.).
  A^+A szimm., így $A^+A = A^T {A^+}^T$, ezért
  A^+A képtere $\subseteq$ A^T képtere.
  $A = AA^+A$ miatt A^+A rangja $\geq$ A rangja, ami egyben A^T rangja. Ezért a két képtér megegyezik.
  Tehát A^+A merőleges vetítés A^T képterére.

Pszeudoinverz - egyértelműség

- **Áll.:** A^+ és A' két pszeudoinverze A-nak $\Rightarrow A' = A^+$.
  - **Biz.:** A lemma miatt AA^+ is és AA' az ortogonális projekció A képterére, így
  - $AA' = AA^+$.
  - Hasonlóan, $A'A = A^+A$.
  - Ezekből $A' = A'AA' = A'AA^+ = A^+AA^+ = A^+$.
- **Köv.:** Ha $A = U'\Sigma U^T$ az A mátrix red. SVD-je, akkor A pszeudoinverze $A^+ = U\Sigma^{-1}U'^T$.

Pszeudoinverz - példa I.

- $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}.$
- SVD: $A = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix} \sqrt{6} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$
- $A^+ = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \frac{1}{\sqrt{6}} \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} = \frac{1}{6} A^T$

Pszeudoinverz - példa II.

- $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$
- $A^T A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$
- SVD: $A = \begin{pmatrix} 1 \\ 0 \end{pmatrix} 1 \begin{pmatrix} 0 & 1 \end{pmatrix},$
- $A^+ = \begin{pmatrix} 0 \\ 1 \end{pmatrix} 1 \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = A^T.$

Pszeudoinverz - feladatok

- $(A^+)^+ = A$.
- $(A^T)^+ = (A^+)^T$.
- A^+ rangja $=$ A rangja.
- Ha A invertálható négyzetes mátrix, akkor $A^+ = A^{-1}$.
- Ha A egy merőleges vetítés (négyzetes) mátrixa, akkor $A^+ = A$.

Pszeudoinverz - példa III.

- Legyen $A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$
- $AB = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \end{pmatrix}$, $BA = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{pmatrix}$
- A és B merőleges vetítések, így $A^+ = A$, $B^+ = B$.
- $(AB)(BA) = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{pmatrix} = \frac{1}{2}A$
- $(AB)(BA)$ nem vetítés, de $(AB)(AB)^+$ az kell legyen, így
- $(AB)^+ \neq BA = B^+A^+$.
- AB SVD-je $AB = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \end{pmatrix}$
- $(AB)^+ = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \sqrt{2} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = 2BA.$

A pszeudoinverz és a magtér

A $m \times n$-es valós

- **Áll.:** A^T képtere $= (\ker A)^\perp$
  - **Biz.:**
  - $\subseteq$: Ha $x \in \ker A$ és $y \in \mathbb{R}^m$, akkor $(A^T y)^T x = y^T A x = 0$, azaz $x \perp Ay$. Tehát $A^T \mathbb{R}^m \leq (\ker A)^\perp$.
  - dimenziók $=$:
    $\dim A^T = A^T$ rangja $= A$ rangja $= \dim A\mathbb{R}^n =$
    $n - \dim \ker A = \dim (\ker A)^\perp$.
- **Köv.:** $I - A^+ A$ merőleges vetítés A magterére
  - **Biz.:** $A^+ A$ merőleges vetítés $A^T \mathbb{R}^m$-re
  - $I - A^+ A$ merőleges vetítés $(A^T \mathbb{R}^m)^\perp$-re
  - $(A^T \mathbb{R}^m)^\perp = (\ker A)^{\perp\perp} = \ker A$

SVD - feladatok

- **számpéldák önálló gyakorlásra** zh101029, 3. és 4. feladatok; pzh101108, 3. és 4. feladatok; pzh091106 2.feladat
- (zh091030, 6. feladat) Legyen A egy $m \times n$-es valós. Legyen B a következő

$$B = \left(\begin{array}{c|c} 0 & A \\ \hline A^T & 0 \end{array}\right).$$

Milyen összefüggés van A szinguláris értékei és B sajátértékei között?