7 SVD dekompozicija

Teorem 7.1 (Singularna dekompozicija)**.** *Neka je $A \in \mathbb{R}^{m\times n}$. Tada postoje unitarne matrice $U \in \mathbb{R}^{m\times m}$ i $V \in \mathbb{R}^{n\times n}$, te matrica $\Sigma \in \mathbb{R}^{m\times n}$ koja na mjestima (i,i) $(1 \leq i \leq \min\{m,n\})$ ima padajući niz nenegativnih realnih vrijednosti, a na ostalim mjestima nule, takve da je*

$$A = U\Sigma V^*.$$

Vrijednosti $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min\{m,n\}} \geq 0$ na dijagonali matrice Σ zovemo singularnim vrijednostima.

Stupce matrice U ponekad nazivamo lijevim, a stupce matrice V desnim singularnim vektorima.

Primijetite da, za razliku od spektralne dekompozicije, ne postoji uvjet na matricu A da bi imala singularnu dekompoziciju: može biti pravokutna, singularna, a također dozvoljava singularnu dekompoziciju i kada je kompleksna. Singularne vrijednosti uvijek su nenegativni realni brojevi.

Dokaz teorema nećemo raditi. Provodi se slično kao dokaz egzistencije spektralne dekompozicije. Također slično kao kod spektralne dekompozicije: singularne vrijednosti za matricu su jedinstvene, ali sama faktorizacija ne mora biti (pogledajte za primjer nul-matricu).

Propozicija 7.1. *Dana je matrica $A \in \mathbb{R}^{m\times n}$. Tada su netrivijalne singularne vrijednosti matrice A upravo drugi korijeni netrivijalnih svojstvenih vrijednosti matrica AA^* i A^*A.*

Dokaz. Neka je $A = U\Sigma V^*$ singularna dekompozicija matrice A. Tada vrijedi

$$AA^* = (U\Sigma V^*)(U\Sigma V^*)^* = U\Sigma\Sigma^* U^*.$$

Matrica $S := \Sigma\Sigma^*$ dijagonalna je matrica reda n. Na prvih $\min\{m,n\}$ mjesta na dijagonali nalaze se kvadrati singularnih vrijednosti matrice A, a na preostalim mjestima (ako ih ima) su nule. S druge strane, gornja jednakost pokazuje da je USU^* upravo spektralna dekompozicija matrice AA^*, odnosno da su dijagonalni elementi matrice S upravo svojstvene vrijednosti matrice AA^*. Slično dokazujemo tvrdnju za matricu A^*A. $\square$

Propozicija 7.2. *Neka je $A = U\Sigma V^*$ singularna dekompozicija matrice A. Neka je točno r njenih singularnih vrijednosti različito od nula. Neka su v_i i u_i oznake za i-te stupce matrica V i U. Tada vrijedi:*

a) $r(A) = r$;

b) $\operatorname{Ker} A = [\{v_{r+1}, \ldots, v_n\}]$;

c) $\operatorname{Im} A = [\{u_1, \ldots, u_r\}]$;

d) $A = U_r\Sigma_r V_r$, *gdje je Σ_r gornja lijeva podmatrica matrice Σ reda r, a U_r i V_r matrice sastavljene od prvih r stupaca matrica U i V;*

e) $\|A\|_F^2 = \sigma_1^2 + \ldots + \sigma_r^2$;

f) $\|A\|_2 = \sigma_1$.

Dokaz. Matrica Σ_r očito ima rang r, dok množenjem te matrice regularnim matricama ne mijenja njezin rang. Zato je $r(A) = r$.

Pomnožimo jednakost $A = U\Sigma V^*$ zdesna matricom V. Dobivamo

$$AV = U\Sigma.$$

Odavde čitamo da za svaki indeks $1 \leq i \leq r$ vrijedi $Av_i = \sigma_i u_i \neq 0$, dok za sve indekse $i > r$ vrijedi $Av_i = 0$. Sada tvrdnje b) i c) slijede iz činjenice da je $\{v_1, v_2, \ldots, v_n\}$ baza za $\mathbb{R}^n$, a $\{u_1, u_2, \ldots, u_m\}$ baza za $\mathbb{R}^m$.

Neka su $\bar{U}_r$ i $\bar{V}_r$ matrice sastavljene od zadnjih $m-r$ i $n-r$ stupaca matrica U i V redom. Tada vrijedi

$$A = U\Sigma V^* = \begin{bmatrix} U_r & \bar{U}_r \end{bmatrix} \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^* \\ \bar{V}_r^* \end{bmatrix} = \begin{bmatrix} U_r\Sigma_r & 0 \end{bmatrix} \begin{bmatrix} V_r^* \\ \bar{V}_r^* \end{bmatrix} = U_r\Sigma_r V_r^*,$$

čime je tvrdnja d) dokazana.

Frobeniusova norma matrice je unitarno invarijantna: množenjem matrice slijeva ili zdesna unitarnom matricom ne mijenja normu te matrice. Zato vrijedi

$$\|A\|_F^2 = \|U\Sigma V^*\|_F^2 = \|\Sigma\|_F^2 = \sigma_1^2 + \ldots + \sigma_r^2,$$

čime je dokazana i tvrdnja e). Tvrdnja f) navedena je na predavanju. Za vježbu je možete dokazati sami, koristeći općenitu definiciju operatorske norme

$$\|A\|_p := \max_{\|x\|_p=1} \|Ax\|_p, \quad p \in [1, \infty\rangle,$$

i tako što dokažete da je i ta norma unitarno invarijatna. □

Teorem 7.2 (Ekhard, Young, Mirsky)**.** *Neka je $A = U\Sigma V^*$ singularna dekompozicija matrice $A \in \mathbb{R}^{m\times n}$ ranga r. Neka je $1 \leq k < r$. Tada je*

$$\min_{r(X)=k} \|C - X\|_2 = \sigma_{k+1}.$$

Minimum se ostvaruje za matricu $X = U_k\Sigma_k V_k$, gdje je Σ_k gornja lijeva podmatrica matrice Σ reda k, a U_k i V_k matrice sastavljene od prvih k stupaca matrica U i V.

Zadatak 7.1. *Zadana je matrica A*

$$A = \begin{bmatrix} 0.6 & 0.48 & 0.64 & 0 \\ -0.8 & 0.36 & 0.48 & 0 \\ 0 & -0.48 & 0.36 & 0.8 \\ 0 & 0.64 & -0.48 & 0.6 \end{bmatrix} \begin{bmatrix} 17 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.6 & 0.64 & 0 & -0.48 \\ 0 & 0.36 & -0.8 & 0.48 \\ 0 & 0.48 & 0.6 & 0.64 \\ 0.8 & -0.48 & 0 & 0.36 \end{bmatrix}$$

svojom singularnom dekompozicijom $A = U\Sigma V^*$. Odredite:

a) rang matrice A;

b) Frobeniusovu i 2-normu matrice A;

c) uvjetovanost matrice A u normi $\|\cdot\|_2$;

d) matricu A_3 kao najbolju aproksimaciju matrice A ranga 3 u 2−normi, te udaljenost tih dviju matrica u istoj normi;

e) jezgru matrice A_3^T.

Dokaz. Ovaj zadatak direktna je primjena gornjih rezultata.

Matrica A je punog ranga (tj. $r(A) = 4$). To vidimo ili zato što je dana kao umnožak regularnih matrica, ili zato što su joj sve singularne vrijednosti pozitivne.

Prema gornjim formulama imamo

$$\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2} = \sqrt{17^2 + 8^2 + 4^2 + 1^2} = \sqrt{370},$$
$$\|A\|_2 = \sigma_1 = 17.$$

Za uvjetovanost vrijedi $\kappa(A) = \|A\|_2 \cdot \|A^{-1}\|_2$. Normu matrice A smo odredili, dok za matricu A^{-1} nije teško dokazati da su joj singularne vrijednosti upravo recipročne vrijednosti singularnih vrijednosti matrice A. Zato je $\|A^{-1}\|_2 = \max\limits_{k=1,2,3,4} \sigma_k^{-1} = 1$, pa je

$$\kappa(A) = \|A\|_2 \cdot \|A^{-1}\|_2 = 17 \cdot 1 = 17.$$

Matrica A_3 prema gornjem teoremu iznosi

$$A_3 = \begin{bmatrix} 0.6 & 0.48 & 0.64 \\ -0.8 & 0.36 & 0.48 \\ 0 & -0.48 & 0.36 \\ 0 & 0.64 & -0.48 \end{bmatrix} \begin{bmatrix} 17 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 0.6 & 0.64 & 0 & -0.48 \\ 0 & 0.36 & -0.8 & 0.48 \\ 0 & 0.48 & 0.6 & 0.64 \end{bmatrix}.$$

Primijetite da ju možemo pisati i na ovaj način:

$$A_3 = \begin{bmatrix} 0.6 & 0.48 & 0.64 & 0 \\ -0.8 & 0.36 & 0.48 & 0 \\ 0 & -0.48 & 0.36 & 0.8 \\ 0 & 0.64 & -0.48 & 0.6 \end{bmatrix} \begin{bmatrix} 17 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.6 & 0.64 & 0 & -0.48 \\ 0 & 0.36 & -0.8 & 0.48 \\ 0 & 0.48 & 0.6 & 0.64 \\ 0.8 & -0.48 & 0 & 0.36 \end{bmatrix}.$$

Ponovno prema teoremu, dana udaljenost iznosi $\|A - A_3\|_2 = \sigma_4 = 1$.

Konačno, za matricu A_3^T transponiranjem zadnjeg matričnog umnoška dobivamo da joj je upravo

$$A_3^T = \begin{bmatrix} 0.6 & 0 & 0 & 0.8 \\ 0.64 & 0.36 & 0.48 & -0.48 \\ 0 & -0.8 & 0.6 & 0 \\ -0.48 & 0.48 & 0.64 & 0.36 \end{bmatrix} \begin{bmatrix} 17 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.6 & -0.8 & 0 & 0 \\ 0.48 & 0.36 & -0.48 & 0.64 \\ 0.64 & 0.48 & 0.36 & -0.48 \\ 0 & 0 & 0.8 & 0.6 \end{bmatrix}$$

singularna dekompozicija. Zato direktnom primjenom gornjih rezultata dobivamo da joj je jezgra razapeta zadnjim stupcem matrice V, što je transponirani zadnji redak najdesnije matrice u gornjem umnošku: $\mathrm{Ker}\, A_3^T = [\{(0, 0, 0.8.0.6)^T\}]$. □

Zadatak 7.2 (DZ). *Dokažite da su općenito singularne vrijednosti matrice A^{-1} upravo recipročne vrijednosti singularnih vrijednosti matrice A.*

Definicija 7.1. *Neka je dana matrice* $A \in \mathbb{R}^{m\times n}$*. Za matricu* $A^\dagger$ *kažemo da je Moore-Penroseov inverz matrice* A *ako vrijedi*

- $AA^\dagger A = A$;
- $A^\dagger AA^\dagger = A^\dagger$;
- $(AA^\dagger)^* = AA^\dagger$;
- $(A^\dagger A)^* = A^\dagger A$.

Očito je za regularne kvadratne matrice Moore-Penroseov inverz upravo pravi inverz matrice A. Pokazuje se općenito da je Moore-Penroseov inverz matrice jedinstven, te ima eksplicitnu formulu: ako je $A = U\Sigma V^*$ singularna dekompozicija matrica matrice A, tada je $A^\dagger = V\tilde{\Sigma}U^*$, gdje smo s $\tilde{\Sigma}$ označili matricu dobivenu od matrice Σ transponiranjem i zamjenom nenul elemenata njihovim recipročnim vrijednostima.

Zadatak 7.3 (DZ)**.** *Dokažite da je* $A^\dagger = V\tilde{\Sigma}U^*$ *Moore-Penroseov inverz matrice* A*.*

Primjer 7.1. *Moore-Penroseov inverz matrice* A_3 *iz prošlog zadatka je*

$$A_3^\dagger = \begin{bmatrix} 0.6 & 0 & 0 & 0.8 \\ 0.64 & 0.36 & 0.48 & -0.48 \\ 0 & -0.8 & 0.6 & 0 \\ -0.48 & 0.48 & 0.64 & 0.36 \end{bmatrix} \begin{bmatrix} \frac{1}{17} & 0 & 0 & 0 \\ 0 & \frac{1}{8} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.6 & -0.8 & 0 & 0 \\ 0.48 & 0.36 & -0.48 & 0.64 \\ 0.64 & 0.48 & 0.36 & -0.48 \\ 0 & 0 & 0.8 & 0.6 \end{bmatrix}.$$

7.1 Primjene SVD-a

Kompresija slika

- Slika na ekranu je predstavljena kao mreža piksela određene boje.
- Svaka boja je predstavljena kao mješavina crvenih, zelenih i plavih piksela različitog zasićenja, u rasponu od 0-255 (0 - boja nije zastupljena, 255 - boja je u potpunosti zastupljena).
- Prema tome, sliku možemo predstaviti kao matricu. Svaki piksel je predstavljen brojem, a stupac i redak matrice određuju njegovu poziciju na originalnoj slici.
- U slučaju crno–bijele slike, zasićenost svake tri boje u jednom pikselu je jednaka, pa moramo pamtiti samo jednu vrijednost.
- U slučaju slike u boji, za svaki piksel moramo pamtiti tri vrijednosti: zasićenost crvene, zelene i plave boje.

Primjer 7.2. *Pretpostavimo da imamo crno–bijelu sliku dimenzije* 100×100 *piksela. Svaki piksel je jedan podatak u matrici reda* 100×100 *raspona 0–255. Ako želimo pohraniti tu sliku, moramo pamtiti* 100×100 *brojeva. Za usporedbu kako broj podataka koje moramo pamtiti raste:*

- *za crno–bijelu sliku veličine* 1280×1024 *treba pamtiti* 1310720 *podataka (1.25 MB),*

- *za sliku u boji iste veličine moramo pamtiti tri puta više podataka* 3932169 *(3.75 MB),*
- *za film imamo 30-60 frameova u sekundi!*

Da bi uštedjeli memoriju možemo izračunati SVD aproksimaciju nižeg ranga. Recimo da se odlučimo za rang 10 (uzmemo prvih 10 stupaca matrica U i V i prvih 10 dijagonalnih elemenata matrice Σ) broj podataka koje trebamo pamtiti je 2010. *(Za ilustraciju pogledati predavanja).*

U slučaju slika u boji, prvi korak je sliku prebaciti u tri slike koje predstavljaju zasićenost crvene, zelene i plave boje. Zatim napravimo SVD za svaku od njih. Izaberemo aproksimaciju nižeg ranga i skupimo u jednu sliku. Primjer kompresije za sliku veličine 2800×2052*:*

$$\begin{aligned} \textit{Original:}\ & 16.4\,\mathrm{MB}, \\ \textit{Rang 100:}\ & 1.4\,\mathrm{MB}, \\ \textit{Rang 75:}\ & 1.0\,\mathrm{MB}, \\ \textit{Rang 50:}\ & 0.7\,\mathrm{MB}\,. \end{aligned}$$

Više o sažimanju slika koristeći SVD možete pročitati u članku:

- Mathews, Brady. "Image compression using singular value decomposition (svd)." *Univ. Utah* (2014).

Prepoznavanje rukom pisanih znamenki

- Primjena: automatsko čitanje zip koda sa omotnica.
- Matematički model: za zadani skup ručno klasificiranih znamenki (*training set*) treba klasificirati nepoznate znamenke (*test set*).
- Znamenka: crno–bijela slika dimenzije 16×16 prikažemo kao vektor reda 256 (stupce matrice poredamo u vektor jedan iznad drugog).
- Sve slike iste znamenke (npr. znamenke 3) iz *training* skupa poredamo u matricu $A \in \mathbb{R}^{256\times n}$.
- Stupci matrice A razapinju potprostor od $\mathbb{R}^{256}$ koji ne može biti velike dimenzije, inače bi se potprostori koje razapinju različite znamenke presijecali.
- Ortogonalna baza tih potprostora se računa koristeći SVD matrice A.
- Svaki stupac od A predstavlja jednu sliku znamenke 3, prema tome lijevi singularni vektori (tzv. singularne slike) predstavljaju ortogonalnu bazu u prostoru slika broja 3.
- Prvi singularni vektor predstavlja dominantni smjer podataka u matrici.

Pretpostavke potrebne za klasifikacijski algoritam koji koristi SVD:

- Svaka znamenka je dobro predstavljena sa prvih nekoliko singularnih slika svoje vrste (koliko mnogo je nekoliko se vidi eksperimentalno)

- Rastav na prvih nekoliko singularnih slika pravi dobru razliku između različitih znamenki.
- Ako nepoznatu znamenku možemo bolje aproksimirati u bazi singularnih slika jedne određene znamenke (npr. 3) nego u bazi preostalih znamenki, tada je najvjerojatnije nepoznata znamenka jednaka 3.

Algoritam se sastoji od sljedećih dijelova:

1. Za svaku znamenku 0-9 složimo matricu koristeći *training set* $A_d \in \mathbb{R}^{256\times n_d}$, $d = 0, 1, \dots, 9$.
2. Izračunamo SVD dekompoziciju $A_d = U_d \Sigma_d V_d^T = \sum_{i=0}^{256} \sigma_i^d u_i^d (v_i^d)^T$. Za svaku znamenku d izaberemo s koliko singularnih slika k_d ćemo aproksimirati potprostor razapet tom znamenkom.
2. Za zadanu nepoznatu znamenku predstavljenu vektorom $z \in \mathbb{R}^{256}$ riješimo 10 problema najmanjih kvadrata
$$\min_{\alpha_i^d} \left\| z - \sum_{i=1}^{k_d} \alpha_i^d u_i^d \right\|. \tag{1}$$
3. Nepoznata znamenka je ona znamenka d za koju smo dobili najmanju vrijednost (1).

Više o prepoznavanju rukom pisanih znamenki možete pročitati u knjizi:

- Eldén, Lars. Matrix methods in data mining and pattern recognition. *Society for Industrial and Applied Mathematics*, 2007.

Pretraživanje podataka

- Problem: imamo bazu od d dokumenata (knjige, članci) i za zadani upit želimo izlistati relevantne dokumente.
- Svaki dokument je predstavljen ključnim riječima – terminima. Svaki od termina ima određenu težinu (najčešće koliko često se pojavljuje u dokumentu).
- Za bazu dokumenata imamo t termina kojima su ti dokumenti opisani (težina termina u određenom dokumentu može biti 0).
- Baza dokumenata je predstavljena matricom A reda $t \times d$. Na mjestu a_{ij} se nalazi težina i-tog termina u j-tom dokumentu.
- Upit je predstavljen vektorom $q \in \mathbb{R}^t$. Za svaki termin je dodijeljena težina (može biti i 0).
- Relevantnost dokumenta j za upit q se mjeri kao kosinus kuta između upita i dokumenta:

$$\cos\theta_j = \frac{a_j^T q}{\|a_j\|_2 \|q\|_2}, \quad a_j = Ae_j. \tag{2}$$

- Problem: moguće je da termin koji pretražujemo nije direktna ključna riječ nekog dokumenta iako je on relevantan za upit (npr. ključna riječ *baking* i dokument *Pastry: A Book of Best French Recipes*).
- Rješenje: Bazu, tj. matricu A zamijenimo sa aproksimacijom nižeg ranga:
  - Neka je $A_k = U_k \Sigma_k V_k^T$ aproksimacija matrice A matricom ranga k.
  - Relevantnost sada možemo računati kao
$$\cos\theta_j = \frac{(A_k e_j)^T q}{\|A_k e_j\|_2 \|q\|_2} = \frac{e_j^T V_k \Sigma_k (U_k^T q)}{\|\Sigma_k V_k^T e_j\|_2 \|q\|_2}.$$
  - Elementi vektora $U_k^T q$ su koordinate u bazi projekcije $U_k U_k^T q$ vektora upita na prostor razapet stupcima matrice A_k. Prema tome, može se koristiti i alternativna formula
$$\cos\theta_j' = \frac{e_j^T V_k \Sigma_k (U_k^T q)}{\|\Sigma_k V_k^T e_j\|_2 \|U_k^T q\|_2}. \tag{3}$$

Više o pretraživanju podataka možete pročitati u članku:

- Berry, Michael W., Zlatko Drmac, and Elizabeth R. Jessup. "Matrices, vector spaces, and information retrieval." *SIAM review* 41.2 (1999): 335-362.

Netflix problem

- Problem: predložiti korisniku film na osnovu njegovih ocjena za već odgledane filmove (tj. procijeniti ocjenu za filmove koje korisnik nije pogledao).
- Netflix je nudio nagradu za algoritam koji će poboljšati njihov postojeći algoritam za 10%.
- Algoritam koji koristi SVD (bilo ih je više) je pobijedio.
- Ideja: retci matrice $A \in \mathbb{R}^{C\times M}$ predstavljaju korisnike, stupci matrice predstavljaju filmove. Na mjestu a_{ij} se nalazi ocjena korisnika i za film j.
- Nađemo SVD matrice $A = U\Sigma V^T = (V\sqrt{\Sigma})(\sqrt{\Sigma}V^T)$.
- Aproksimiramo matricu A matricom nižeg ranga k i dobijemo dvije matrice $\tilde{U}_k = U_k\sqrt{\Sigma_k} \in \mathbb{R}^{C\times k}$ i $\tilde{V}_k = \sqrt{\Sigma}_k V_k \in \mathbb{R}^{M\times k}$.
- Aproksimacija ocjene za film j korisnika i je $e_i^T \tilde{U}_k \tilde{V}_k^T e_j$.
- U realnosti nemamo ocjene za svaki film od svakog korisnika, pa se koriste druge metode za aproksimaciju vrijednosti koje nedostaju.
- Aproksimacija nižeg ranga se može interpretirati kao svrstavanje filmova u k kategorija. Točnije, element na mjestu (i,p) matrice U_k predstavlja koliko je korisnik i zainteresiran za filmove iz kategorije p. S druge strane, element na mjestu (p,j) matrice V_k^T predstavlja s kojim faktorom film j pripada kategoriji p.

Više o ovom problemu i njegovom rješavanju možete pročitati u članku jednih od natjecatelja:

- Koren, Yehuda, Robert Bell, and Chris Volinsky. "Matrix factorization techniques for recommender systems." *Computer* 42.8 (2009): 30-37.