math.e

Hrvatski matematički elektronički časopis

Dekompozicija matrice na singularne vrijednosti i primjene

klasifikacija singularne vrijednosti

Andrej Novak
Sveučilište u Zagrebu Prirodoslovno-matematički fakultet Matematički odsjek

Danijel Pavlović
AVL-AST d.o.o. Zagreb

Zagreb, prosinac 2013.

1 Uvod

Moderna tehnologija i stalan porast broja stanovnika na Zemlji razlog su velike produkcije podataka Svaki dan se proizvede ogromna količina novih podataka koje je potrebno spremati, obraditi, a često i slati. Zbog toga je bilo potrebno osmisliti načine za efikasno spremanje podataka kako bi oni zauzimali što je manje moguće prostora, a da glavna informacija bude očuvana.
Lijep primjer toga je kompresija digitalne slike. Uzmimo za primjer sliku dimenzije 3000×2000 piksela. Kako bismo takvu sliku spremili u memoriju potrebno nam je $3000 \cdot 2000 \cdot 24 = 14400000$ bitova memorije (za pamćenje jednog piksela koristimo 24 bita) što iznosi 18 megabajta.

Slika 1: a) Izvorna slika, b) Komprimirana slika

U ovom članku bavit ćemo se matričnom dekompozicijom koja nam omogućava da iz nekog skupa podataka, prikazanog pomoću matrice, izdvojimo "najvažniji" dio, a ostatak izostavimo. Koristeći tu tehniku možemo komprimirati slike, razviti algoritme koji klasificiraju rukom pisane znamenke i još mnogo toga. Zamislite sada da ste zaposleni u pošti i da je vaša zadaća sortirati pristigla pisma prema poštanskom broju kako bi ona stigla na pravu adresu. Možda biste radije bili prometni policajac na autocesti koji je zadužen za praćenje brzine automobila koje ulaze u tunel. Vaš posao je snimiti tablice automobila koji ne poštuju ograničenje brzine, a zatim u bazi podataka registracija vozila pronaći tu tablicu te na priloženu adresu vlasnika poslati kaznu za prebrzu vožnju. Biste li svaki dan isti posao radili ručno ili biste radije pokušali taj postupak automatizirati? Srećom, ljudi su se već susreli s ovim problemom i smislili mnogo načina kako taj posao ubrzati i uštediti dragocjeno vrijeme.
Svi navedeni problemi mogu se riješiti upotrebnom jednog moćnog alata numeričke matematike koji ćemo vam pokušati približiti u ostatku ovog članka.

Dekompozicija matrice na singularne vrijednosti (SVD) je važna dekompozicija matrice bilo s

2 Dekompozicija matrice na singularne vrijednosti

teorijske ili praktične strane. U sljedećem teoremu dana je definicija SVD dekompozicije te tvrdnja o njezinoj egzistenciji za svaku matricu. Navodimo samo tvrdnju teorema, dokaz se može naći u standardnim udžbenicima iz numeričke analize.

Teorem 1. *(SVD) Neka su m i n ($m \geq n$) prirodni brojevi te A proizvoljna $m \times n$ realna matrica. Tada postoji dekompozicija $A = U\Sigma V^{\tau}$, gdje je U ortonormalna $m \times n$ matrica i V ortogonalna $n \times n$ matrica, a $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$, sa $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$.*

Definicija 2. *Stupce matrice $U = [u_1, \ldots, u_n]$ nazivamo lijevi singularni vektori, a stupce matrice $V = [v_1, \ldots, v_n]$ nazivamo desni singularni vektori. Brojeve σ_i nazivamo singularne vrijednosti.*

Uz uvedene oznake lako se vidi da vrijedi $A = U\Sigma V^{\tau} = \sum_{i=1}^{n} \sigma_i u_i v_i^{\tau}$.

Napomena 1. U slučaju kada je $m < n$, SVD definiramo za matricu A^{τ}.

3 Svojstva dekompozicija matrice na singularne vrijedn

Nakon što smo iskazali teorem koji garantira postojanje dekompozicije, u ovom odjeljku dajemo pregled nekih svojstâva dekompozicije, također bez dokaza. Ali prije iskaza teorema uvodimo još jedan pojam. Do sada smo inverz matrice definirali isključivo za kvadratne matrice, a sada želimo taj pojam proširiti na pravokutne matrice i matrice koje nisu punog ranga. Postoje tri ekvivalentna načina na koji se to može načiniti, u nastavku navodimo samo Penroseovu definiciju.

Definicija 3. *Neka je $A \in C^{m \times n}$, njezin generalizirani inverz je jedinstvena matrica A^+, takva da vrijedi*

$$\begin{aligned} AA^+A &= A, & A^+AA^+ &= A^+ \\ (AA^+)^* &= AA^+, & (A^+A)^* &= A^+A. \end{aligned}$$

Generalizirani inverz ponekad nazivamo i Moore-Penroseov pseudoinverz.
Neka je $A = U\Sigma V^{\tau}$. Tada Moore-Penrose pseudoinverz za singularnu matricu A možemo izračunat kao $A^+ \equiv V\Sigma^+U^{\tau}$, gdje je $\Sigma^+ = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}^+ = \begin{bmatrix} \Sigma_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}$, a Σ_1 dijagonalna matrica sa ne-nul singularnim vrijednostima na dijagonali.

Teorem 4. *(Svojstva SVD-a) Neka je $m \geq n$ te neka je $A = U\Sigma V^\tau$ dekompozicija na singularne vrijednosti matrice A.*

1. *Neka je A simetrična matrica sa svojstvenim vektorima q_i, tj. $A = Q\Lambda Q^\tau$, pri čemu je $Q = [q_1, \ldots, q_n]$ te $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$. Tada je $A = U\Sigma V^\tau$ za $u_i = q_i$, $\sigma_i = |\lambda_i|$ i $v_i = sign(\lambda_i) q_i$.*

2. *Svojstvene vrijednosti simetrične matrice $A^\tau A$ su σ_i^2. Desni singularni vektori v_i su pripadni ortonormirani svojstveni vektori.*

3. *Svojstvene vrijednosti simetrične matrice AA^τ su σ_i^2 te $m-n$ nula. Lijevi singularni vektori u_i su pripadni ortonormirani svojstveni vektori za svojstvene vrijednosti σ_i^2.*

4. *Ako je A punog ranga, a b proizvoljan vektor, onda je $x = A^+ b = V\Sigma^{-1}U^\tau b$ rješenje problema $\min_x \|Ax - b\|_2$.*

5. *Vrijedi $\|A\|_2 = \sigma_1$. Ukoliko postoji A^{-1} tada je $\|A^{-1}\|_2 = 1/\sigma_n$.*

6. *Neka je $A_k = U\Sigma_k V^\tau = \sum_{i=1}^n \sigma_i u_i v_i^\tau$, gdje je $\Sigma_k = diag(\sigma_1, \ldots \sigma_k, 0, \ldots, 0)$. Tada matrica A_k ima rang k te je ona najbliža matrici A među svim matricama ranga k:*
$$\|A - A_k\|_2 = \min_{rang(B)=k} \|A - b\|_2$$
Također je $\|A - A_k\|_2 = \sigma_{k+1}$.

7. *Pretpostavimo da za singularne vrijednosti matrice A vrijedi $\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_n = 0$, tada je $rang(A) = r$. Nadalje, nul-potprostor od A je razapet stupcima $v_{r+1}, \ldots, v_n$ matrice V, a slika operatora A je razapeta stupcima $u_1, \ldots, u_r$ od U.*

Napomena 2. Umnožak $\|A\|_2 \|A^{-1}\|_2$ iz pete točke nazivamo uvjetovanost matrice A.
U ostatku članka, komentirat ćemo primjenu Teorema 2.

4 Neke primjene dekompozicije matrice na singul vrijednosti

4.1 Kompresija slike

Neka je dana $m \times n$ matrica A s elementima $a_{i,j} \in [0,1]$. Zamislimo da A predstavlja $m \times n$ *grayscale* sliku gdje broj na (i,j) mjestu u matrici predstavlja svjetlinu piksela na mjestu (i,j) na slici. Ako uzmemo da vrijednost nula predstavlja crno, a jedan bijelo, zanimljivo je pitanje što će se dogoditi kad napravimo $A = U\Sigma V^\tau$, a onda u matrici Σ posljednjih nekoliko ne-nul elemenata stavimo na nulu te promotrimo dobivenu sliku A_k.

Slika 2: Testna slika A dimenzije 376×350 u JPG formatu zauzima 32 KB

Preciznije, neka je $A = U\Sigma V^T$, SVD matrice A. Prema šestoj tvrdnji prethodnog teorema, vrijedi da je $A_k = \sum_{i=0}^{k} \sigma_i u_i v_i^T$ najbolja aproksimacija matrice A matricom ranga k, u smislu $\|A - A_k\|_2 = \sigma_{k+1}$. Da bismo spremili podatke $u_1, \dots u_k$ te $\sigma_1 v_1, \dots, \sigma_k v_k$ iz kojih možemo dobiti matricu A_k potrebno je pamtiti samo $(m+n) \cdot k$ brojeva budući da su u_i i v_j vektori iz $\mathbb{R}^m$ odnosno iz $\mathbb{R}^n$. S druge strane, za spremanje matrice A potrebno je pamtiti svih $m \cdot n$ brojeva. Stoga, koristimo A_k kao kompresiju slike A.

Slika 3: Kompresija A_k s a) $k = 50$, b) $k = 25$, c) $k = 15$. Omjer kompresije a) 0.275, b) 0.137, c) 0.082, a zauzimaju respektivno 25 KB, 22 KB te 20 KB.

Prethodne simulacije su rađene u programskom paketu Matlab. S obzirom na jednostavnost programske izvedbe u nastavku prilažemo i pojednostavljeni kôd kojim se čitatelj i sam može uvjeriti u korisnost i elegantnost ove primjene.

```
A = imread('test1.png');
A = A(:,:,1);
A = double(A);
[U,S,V] = svd(A);
k = 25;
Ak = U(:,1:k)*S(1:k,1:k)*V(:,1:k)';
image(Ak) colormap(gray(256));
```

4.2 Osjetljivost sustava linearnih jednadžbi na pogreške u podacima

Neka je dan sustav linearnih jednadžbi u matričnom zapisu $Ax = b$, gdje je A regularna matrica. Rješenje sustava je tada $x = A^{-1}b$. Valja se prisjetiti iz linearne algebre da sustavi u kojima je A singularna nemaju rješenje ili ono nije jedinstveno. Prilikom traženja rješenja linearnog sustava jednadžbi u uvjetima konačne aritmetike možemo očekivati poteškoće ako je matrica "blizu" singularne. Upravo Teorem 2 nam kaže da je najbliža singularna matrica od A udaljena za σ_n stoga je razumno očekivati da će točnost s kojim možemo odrediti rješenje sustava imati neke veze sa σ_k Iako numerički nikako ne želimo računati inverz matrice A, ovakav zapis ima svojih prednosti. Zamislimo primjerice da radimo simulaciju gdje je matrica A fiksna, a jedino što mijenjamo je desna strana sustava. Rješavanje sustava (kad jednom znamo inverz!) se svodi na operaciju množenja matrice vektorom.

Pretpostavimo da smo podatke dobili mjerenjem - što sugerira da oni nisu nužno točni, već sadrže neku grešku. Pretpostavimo da smo umjesto egzaktne vrijednosti b izmjerili vrijednosti $b + \delta b$ na desnoj strani sustava. Rješenje sustava se tada promijenilo iz x u $x + \delta x$. Zanima nas koliko je velika pogreška u rješenju (δx) u odnosu na pogrešku u mjerenju (δb). Rješavamo, dakle

$$x + \delta x = A^{-1}(b + \delta b).$$

Budući da je $x = A^{-1}b$ slijedi $\delta x = A^{-1}\delta b$.
Uzimanjem norme slijedi

$$\|\delta x\| = \|A^{-1}\delta b\| \leq \|A^{-1}\|\|\delta b\|.$$

Ako je $\|A^{-1}\|$ velika lako uviđamo da male greške u podacima b mogu dovesti do velikih grešaka u rješenju.
Dijeljenjem prethodne nejednakosti s $\|x\|$ i uvažavanjem $\frac{\|b\|}{\|A\|} \leq \|x\|$ na desnoj strani nejednakosti slijedi ocjena

$$\frac{\|\delta x\|}{\|x\|} \leq \|A\|\|A^{-1}\| \frac{\|\delta b\|}{\|b\|}.$$

U napomeni 2 već smo spomenuli uvjetovanost matrice $k(A) = \sigma_{max}/\sigma_{min}$.
Time smo pokazali da je relativna pogreška u x manja ili jednaka umnošku uvjetovanosti matrice sustava i relativne pogreške u podacima b.

\newpage

4.3 Problem najmanjih kvadrata za singularnu matricu A

U ovoj sekciji ne samo da ćemo dokazati tvrdnju 4 iz Teorema 4, već i njeno poopćenje za matrice koje nisu punog ranga. Mnogi problemi iz života mogu se formulirati kao problem najmanih kvadrat što ćemo vidjeti i u sljedećem poglavlju koje se bavi klasifikacijom rukom pisanih znamenki.

Već smo vidjeli da je rješenje problema $\min_x ||Ax - b||_2$ za matricu A punog ranga dano sa $x = V\Sigma^{-1}U^T b$. Promotrimo sada slučaj kada matrica A nije punog ranga, već ima rang r koji može biti i manji od n. Kako bismo izveli formulu za rješenje takvog problema, prvo valja saznati nešto više o naravi problema.

Propozicija 5. *Neka je A matrica dimenzije $m \times n$ gdje je ($m \geq n$) i rang od A je r, gdje je $r \leq n$. Također, neka je b proizvoljan vektor dimenzije $m \times 1$. Tada postoji skup dimenzije barem $n - r$ vektora x koji minimiziraju $||Ax - b||_2$.*

Proof. Neka je z vektor iz jezgre od A, tj. neka vrijedi $Az = 0$. Tada ako x minimizira $||Ax - b||_2$, onda i vektor $x + z$ minimizira navedeni izraz jer vrijedi:

$$||A(x + z) - b||_2 = ||Ax + Az - b||_2 = ||Ax - b||_2.$$

Time smo, zbog činjenice da je dimenzija jezgre matrice A upravo jednaka $n - r$, pokazali da je dimenzija skupa rješenja barem $n - r$. ∎

Iz Propozicije 5 vidimo da izbor vektora x nije jedinstven kada je $r < n$.
Slijedi rezultat koji daje ocjenu na normu vektora rješenja x.

Propozicija 6. *Neka je σ_{min}, najmanja ne-nul singularna vrijednost matrice A. Ako x minimizira $||Ax - b||_2$, onda je $||x||_2 \geq |u_n^T b|/\sigma_{min}$ gdje je u_n zadnji stupac od U u rastavu $A = U\Sigma V^T$*

Proof. Kako je $x = A^+ b = V\Sigma^{-1}U^T b$, slijedi da je
$||x||_2 = ||\Sigma^{-1}U^T b||_2 \geq |(\Sigma^{-1}U^T b)_n| = |u_n^T b|/\sigma_{min}$ ∎

Nakon što smo se uvjerili da norma rješenja ovisi o najmanjoj ne-nul singularnoj vrijednosti matrice A, u sljedećoj propoziciji ćemo dati karakterizaciju vektora x koji minimizira $||Ax - b||_2$ za matricu A koja nije punog ranga.

Propozicija 7. *Neka je $A = U\Sigma V^\tau$ matrica ranga $r < n$. Tada je možemo zapisati kao*

$$A = [\, U_1, \quad U_2 \,] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} [\, V_1, \quad V_2 \,]^\tau = U_1 \Sigma_1 V_1^\tau$$

gdje je Σ_1 $r \times r$ regularna matrica, a U_1 i V_1 imaju r stupaca. Neka je σ_{min} najmanja singularna vrijednost od A različita od nule. Tada vrijedi:

(1)

Sva rješenja x mogu se zapisati kao $x = V_1 \Sigma_1^{-1} U_1^\tau b + V_2 z$, gdje je z proizvoljan vektor.

(2)

Rješenje x ima minimalnu normu $||x||_2$ kada je $z = 0$, u kojem slučaju $x = V_1 \Sigma_1^{-1} U_1^\tau b$ i $||x||_2 \leq \frac{||b||_2}{\sigma_{min}}$.

(3)

Promjenom vektora b u vektor $b + \delta b$, rješenje x se može promijeniti za najviše $\frac{||\delta b||_2}{\sigma_{min}}$.

Drugim riječima, norma i kondicija rješenja x ovise o najmanjoj ne-nul singularnoj vrijednosti matrice A.

Proof. Izaberimo $\tilde{U}$ tako da je $[U, \tilde{U}]$ ortogonalna matrica. Tada

$$\begin{aligned} ||Ax - b||_2^2 = ||[U, \tilde{U}]^\tau (Ax - b)||_2^2 &= || \begin{bmatrix} U^\tau \\ \tilde{U}^\tau \end{bmatrix} (U_1 \Sigma_1 V_1^\tau x - b)||_2^2 \\ &= \left|\left| \begin{bmatrix} \Sigma_1 V_1^\tau x - U_1^\tau b \; U_2^\tau b \; \tilde{U}^\tau b \end{bmatrix} \right|\right|_2^2 \\ &= ||\Sigma_1 V_1^\tau x - U_1^\tau b||_2^2 + ||U_2^\tau b||_2^2 + ||\tilde{U}^\tau b||_2^2 \end{aligned}$$

(1)

Budući da zadnja dva pribrojnika u gornjem izrazu ne ovise o x, $||Ax - b||_2$ je minimalna kada je prvi pribrojnik minimalan. Možemo postići da je on jednak nuli ako stavimo $\Sigma_1 V_1^\tau x = U_1^\tau b$, odnosn $x = V_1 \Sigma_1^{-1} U_1^\tau b + V_2 z$ jer je $V_1^\tau V_2 z = 0$ za svaki z.

(2)

Kako su V_1 i V_2 ortogonalne, $||x||_2^2 = ||V_1 \Sigma_1^{-1} U_1^\tau b||_2^2 + ||V_2 z||_2^2$ a ta je norma minimalna za $z = 0$. Ocjena $||x||_2 \leq \frac{||b||_2}{\sigma_{min}}$ slijedi iz nejednakosti $||A \cdot B|| \leq ||A|| \cdot ||B||$ i definicije matrične 2-norme.

(3)

Promjenom vektora b za δb mijenja se x za najviše $||V_1 \Sigma_1^{-1} U_1^\tau \delta b||_2 \leq ||\Sigma_1^{-1}||_2 ||\delta b||_2 = \frac{||\delta b||_2}{\sigma_{min}}$. ∎

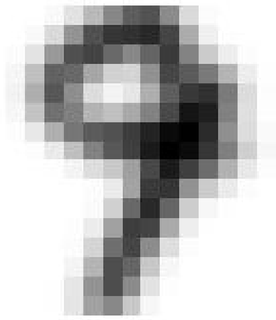
Zaključujemo da je, u slučaju kada je matrica A punog ranga, rješenje problema najmanjih kvadrata vektor $x = A^+ b$, a kada matrica A nije punog ranga, onda među svim vektorima x koji minimiziraju $||Ax - b||_2$, vektor $x = A^+ b$ ima najmanju normu.

4.4 Klasifikacija rukom pisanih znamenki

Možda najzanimljivija primjena SVD-a koju ćemo razmatrati u ovom članku je klasifikacija rukom pisanih znamenki. Potreba za algoritmom koji u kratkom vremenu može prepoznati rukom pisane znamenke pojavila se u Američkoj pošti. Budući da pošta dnevno prima velike količine pisama koja je potrebno razvrstati prema poštanskom broju, trebalo je pronaći način da se postupak razvrstavanja automatizira. Razvijeni su brojni algoritmi koji provode klasifikaciju, a u nastavku ovog članka predstavljamo jedan koji se bazira na SVD-u.

4.4.1 Prikaz rukom pisanih znamenki u računalu

Prvi problem s kojim se susrećemo u razvoju klasifikacijskog algoritma je način prikaza podataka u računalu. Logičan način unošenja rukom pisanih znamenki u računalo je skeniranje. Na taj način svaku znamenku možemo poistovijetiti sa slikom rezolucije $n \times n$ piksela (u daljnjem tekstu koristi ćemo $n = 16$ jer se pokazalo da je ta rezolucija dovoljna za razlikovanje znamenki). Kako nam boja olovke kojom je znamenka napisana nije bitna, sliku možemo pretvoriti u *grayscale* format. Na taj način smo postigli da je svaki piksel na slici predstavljen nekom numeričkom vrijednošću koja određuje kojom će nijansom sive boje piksel biti iscrtan na zaslonu računala. Prema tome, svaku znamenku smo poistovjetili s matricom realnih brojeva iz segmenta $[0,1]$ dimenzije 16×16.



Slika 4: Primjer skenirane rukom pisane znamenke 9 u rezoluciji 16×16 piksela

4.4.2 Opis algoritma

Algoritam bismo mogli ugrubo podijeliti u dvije faze.

- Prva faza se sastoji od ručnog klasificiranja dovoljno velikog skupa znamenki, tj. stvaranja tzv. "skupa za vježbu". Skup za vježbu se može nadalje podijeliti na deset podskupova, gdje svaki podskup sadrži slike koje predstavljaju jednu od znamenaka $0,1,2,3,\cdots,9$.
- Druga faza se sastoji od odabira neke neklasificirane rukom pisane znamenke (znamenke iz tzv. "skupa za testiranje") i određivanja podskupa skupa za vježbu kojem je ta nepoznata znamenka "najbliža".

Prva faza algoritma nam nije interesantna jer se provodi ručno, odnosno bez upotrebe računala, a drugu fazu ćemo detaljnije proučiti.

Kao što smo već rekli, svaku znamenku u računalu zapisujemo kao matricu dimenzije 16×16. Svaki podskup $T_i, i = 0,\cdots,9$ skupa za vježbu sastoji se od k_i matrica koje predstavljaju istu znamenku. Kako bismo jedan cijeli podskup T_i prikazali kao jednu matricu, a ne njih k_i načinit ćemo sljedeće:

(1)

Svaku 16×16 matricu iz podskupa T_i pretvorimo u vektor dimenzije 256 tako da prvih 16 elemenata vektora budu elementi prvog stupca matrice, sljedećih 16 elemenata vektora elementi drugog stupca matrice itd. Postupak je lakše prikazati grafički:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,16} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,16} \\ \vdots & \vdots & \cdots & \vdots \\ a_{16,1} & a_{16,2} & \cdots & a_{16,16} \end{bmatrix} \rightarrow \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{16,1} \\ a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{16,16} \end{bmatrix}$$

(2)

Dobivenih k_i vektora presložimo tako da tvore stupce matrice $A^{(i)}$ dimenzije $256 \times k_i$

Ovim postupkom generirali smo deset matrica $A^{(i)}$ čiji stupci razapinju potprostor od $\mathbb{R}^{256}$ koji sadrži podskup T_i skupa za vježbu. Jednostavnije rečeno, svaka rukom pisana znamenka

$i, i = 0, \cdots, 9$ iz skupa T_i odgovara nekom stupcu matrice $A^{(i)}$. Razumno je očekivati da se svaka znamenka iz skupa za testiranje može prikazati kao linearna kombinacija stupaca matrice $A^{(i)}$, odnosno da leži u slici te matrice. Valja napomenuti da potprostor razapet stupcima matrice $A^{(i)}$, za fiksan i, ne bi smio biti velike dimenzije, jer bi inače potprostori koji pripadaju različitim znamenkama i imali presjek velike dimenije, što znači da bi se neka nepoznata znamenka mogla prikazati pomoću stupaca dviju različitih matrica $A^{(i_1)}$ i $A^{(i_2)}$ što bi otežalo klasifikaciju.

Sada koristeći SVD dekompoziciju iz:

$$A^{(i)} = U^{(i)}\Sigma^{(i)}V^{(i)\tau} = \sum_{j=1}^{256} \sigma_j^{(i)} u_j^{(i)} v_j^{(i)\tau} \tag{1}$$

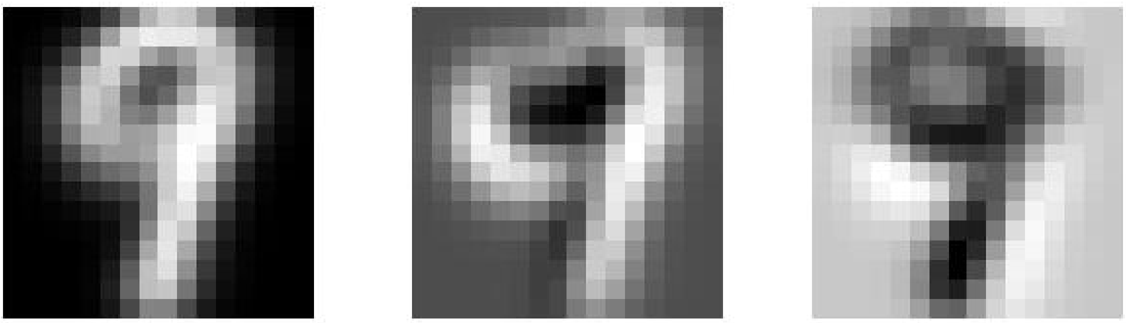
slijedi da se svaki stupac $a_l^{(i)}, l = 1, \cdots, k_i$ matrice $A^{(i)}$ može zapisati kao

$$a_l^{(i)} = A^{(i)} e_l = \sum_{j=1}^{256} \left(\sigma_j^{(i)} v_{j,l}^{(i)}\right) u_j^{(i)} \tag{2}$$

Iz (2) možemo zaključiti da lijevi singularni vektori $u_j^{(i)}$ tvore ortogonalnu bazu za potprostor koji odgovara nekoj znamenci i, a da su koeficijenti u zapisu l-te znamenke iz skupa T_i u toj bazi upravo $\sigma_j^{(i)} v_{j,l}^{(i)}$.

Budući da prvi singularni vektor (onaj koji pripada najvećoj singularnoj vrijednosti) određuje "dominantni smjer" matrice $A^{(i)}$, za očekivati je da će vektor $u_1^{(i)}$ prikazan kao slika izgledati kao znamenka i, a vektori $u_2^{(i)}, u_3^{(i)}, \cdots$ će predstavljati "dominantne varijacije" u podskupu skupa za vježbu T_i.

Neformalno, kažemo da su singularni vektori $u_1^{(i)}, \cdots, u_m^{(i)}$ dominantni smjerovi ako su singularne vrijednosti $\sigma_1, \cdots, \sigma_m$ velike u odnosu na $\sigma_{m+1}, \cdots, \sigma_n$, odnosno, ako je omjer $\frac{\sigma_1}{\sigma_{m+1}} < \epsilon$ za neku unaprijed odabranu malenu toleranciju ϵ. Prema Teoremu 4, to znači da je matrica $A_m^{(i)}$ aproksimacija matrice $A^{(i)}$ s pogreškom manjom od ϵ, te da vektori $u_1^{(i)}, \cdots, u_m^{(i)}$ približno razapinju isti potprostor kao svi stupci matrice $A^{(i)}$. Stoga očekujemo da ćemo znamenke iz testne skupine moći približno prikazati kao linearnu kombinaciju dominantnih singularnih vektora.



Slika 5: (a) Vektor $u_1^{(9)}$. (b) Vektor $u_2^{(9)}$. (c) Vektor $u_3^{(9)}$.

Bitno je naglasiti da se ostatak algoritma zasniva na sljedećim prirodnim pretpostavkama:

(1)

Svaka znamenka (u skupu za vježbu i skupu za testiranje) je dobro karakterizirana s nekoliko dominantnih singularnih vektora $u_1^{(i)}, \cdots, u_m^{(i)}$ za odgovarajuću znamenku i. Drugim riječima, očekujemo da singularne vrijednosti matrice $A^{(i)}$ brzo opadaju, te da je norma $||A^{(i)} - A_m^{(i)}||_2$ malena već za malene vrijednosti m.

(2)

Zapis neklasificirane znamenke kao linearne kombinacije nekoliko dominantnih vektora $u_j^{(i)}$ se razlikuje značajno za različite vrijednosti i, npr. zapis u bazi trojki se znatno razlikuje od zapisa u bazi petica.

(3)

Ako je nepoznata znamenka bolje aproksimirana u jednoj bazi singularnih vektora (npr. $u_j^{(5)}$), nego u ostalim bazama, tada je nepoznata znamenka vjerojatno znamenka 5.

Jedino što još treba napraviti je na neki način odrediti koliko je dobro nepoznata znamenka opisana u deset različitih baza, odnosno odrediti koeficijente zapisa u bazi $u^{(i)}$ za koje je greška najmanja. To ćemo učiniti tako da izračunamo vektor reziduala kao problem najmanjih kvadrata tipa:

$$\min_{\alpha_i} \left\| z - \sum_{j=1}^{m} \alpha_j u_j^{(i)} \right\|, \tag{3}$$

gdje je z nepoznata znamenka, $u_j^{(i)}$ lijevi singularni vektori koji odgovaraju znamenci i, a m broj koji određuje koliko dominantnih singularnih vektora želimo koristiti za aproksimaciju.

Izraz (3) možemo zapisati i kao $\min_\alpha \left\| z - U_m^{(i)} \alpha \right\|$, gdje je $U_m^{(i)} = [u_1^{(i)}, u_2^{(i)}, \cdots, u_m^{(i)}]$ matrica čiji su stupci m dominantnih singularnih vektora koji pripadaju znamenci i, a $\alpha^\tau = [\alpha_1, \alpha_2, \cdots, \alpha_m]$. Budući da su stupci matrice $U_m^{(i)}$ ortonormirani, rješenje problema najmanjih kvadrata je dano s $\alpha = U_m^{(i)\tau} z$, pa je norma reziduala je dana s:

$$\left\| (I - U_m^{(i)} U_m^{(i)\tau}) z \right\|_2. \tag{4}$$

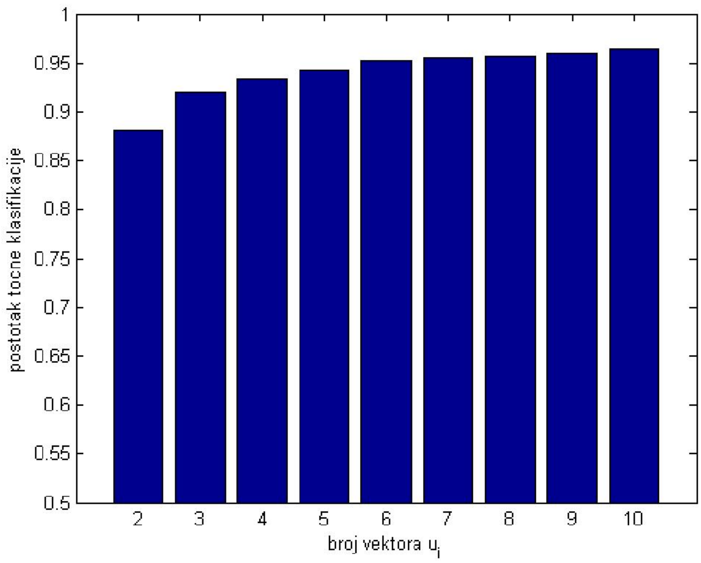
Konačno, možemo zapisati pseudokod algoritma:

- Izračunaj SVD dekompozicije matrica $A^{(i)}$ čiji stupci reprezentiraju znamenke iste vrste iz skupa za vježbu.
- Za danu znamenku iz skupa za testiranje izračunaj po formuli (4) norme reziduala u svih deset baza. Ako je rezidual u bazi $U^{(i)}$ bitno manji od ostalih, klasificiraj nepoznatu znamenku kao znamenku i.

4.4.3 Rezultati testiranja

Algoritam je implementiran u MATLAB-u i testiran na zamenkama iz baze U.S. Postal Service. Baza se sastoji od 9298 klasificiranih rukom pisanih znamenki podijeljenih u dva skupa od 4649 znamenki. Jedan skup je korišten za vježbu, a drugi za testiranje. Za primjer možemo reći da je skup za vježb sadržavao 786 slika koje odgovaraju znamenci 0, tj. matrica $A^{(0)}$ je imala 786 stupaca. Znamenke korištene prilikom izrade ovog članka mogu se pronaći u formatu pogodnom za MATLAB na adresi: http://rrr.soundsoftware.ac.uk/rrr/usps-handwritten-digits-dataset .

Algoritam je na ovim podacima pokazao zavidne rezultate. Klasifikacija jedne znamenke traje u prosjeku 0.02 sekunde, a uspješnost klasifikacije (broj točno klasificiranih znamenki) ovisi o broju m singularnih vektora korištenih pri računanju reziduala.



Slika 6: Uspješnost klasifikacije u ovisnosti o broju vektora u_i.

Slika 6 prikazuje uspješnost klasifikacije u ovisnosti o broju vektora korištenih pri računanju reziduala. Možemo vidjeti da je za dva korištena singularna vektora uspješnost 88% dok je za veći broj vektora uspješnost veća te za 10 korištenih vektora iznosi čak 96.5%.

Za kraj, iz rezultata testiranja možemo izvući dva bitna zaključka koja nam ilustriraju zašto je dekompozicija matrice na singularne vrijednosti iznimno korisna prilikom rješavanja praktičnih problema:

(1)

Algoritam prepoznavanja testne znamenke z kao one koja daje najmanji rezidual kod rješavanja problema najmanjih kvadrata $\min_x ||z - A^{(i)}x||$ ima točnost veću od 95%.

(2)

Potprostori razapeti stupcima matrice $A^{(i)}$ jako su dobro aproksimirani sa svega nekoliko dominantnih singularnih vektora. Stoga nije potrebno rješavati problem najmanjih kvadrata s potencijalno velikom matricom $A^{(i)}$, nego s puno manjom matricom čiji su stupci dominantni singularni vektori. Tako proces rješavanja postaje puno efikasniji i brži uz minimalne posljedice na točnost prepoznavanja.

Literatura

(1)

J. Demmel. Applied Numerical Linear Algebra. SIAM. MIT, 1996.

(2)

J. Demmel. Berkeley mathematics: Lecture notes, Volume 1. Center for Pure and Applied Mathematics, University of California, Berkley, California, 1993.

(3)

L. Elden. Matrix Methods in Data Mining and Pattern Recognition. SIAM. Philadelphia, 2007.

(4)

Z. Drmač. Bilješke s predavanja iz kolegija Uvod u složeno pretraživanje podataka, PMF-Matematičk odsjek, Zagreb, 2009.

ISSN 1334-6083