

Analiza rijetkih komponenata

Marko Filipović

1. travnja 2010.

Sadržaj

- 1 Uvod
- 2 Analiza rijetkih komponenata
 - Grupiranje podataka (clustering). Osnovne metode
 - Faktorizacija nenegativnih matrica (NMF)
 - Rijetka rekonstrukcija
- 3 Primjeri

Pododređeno razdvajanje signala

Definicija problema i terminologija

- $X \in \mathbb{R}^{m \times T}$, $m \ll T$, matrica podataka; pretpostavljamo:

$$X = AS + E$$

gdje je $A \in \mathbb{R}^{m \times n}$, $m < n \ll T$, S rijetka, a E aditivni šum, greška mjerenja ili greška modela

- Cilj: naći A i S
- “Definicija”: vektor $x \in \mathbb{R}^n$ je *rijedak* ako je “većina” elemenata od x jednaka nuli
- S rijetka: svaki *stupac* od S je rijedak
- $n > m$: problem je *pododređen* (“underdetermined”)

Pododređeno razdvajanje signala

- Terminologija:
 - retci matrice S : *komponente* (ili izvorni signali)
 - A : *matrica miješanja* (ili bazna matrica)
 - stupci od A : *bazni vektori*
 - retci matrice X : *miješani signali*
 - $t, 1 \leq t \leq T$: *(vremenski) trenutak*
 - $X \rightarrow X \approx AS$, S rijetka : *razdvajanje signala (analiza rijetkih komponenata)*
- Rješavanje u dvije faze:
 - 1 Aproximacija matrice miješanja grupiranjem podataka (clustering-om) [1]
 - 2 Rijetka rekonstrukcija - rješavamo pododređeni linearni sustav s ograničenjem rijetkosti
- Alternativa: faktorizacija nenegativnih matrica (Nonnegative Matrix Factorization); može se koristiti i za simultanu aproksimaciju A i S

Pododređeno razdvajanje signala

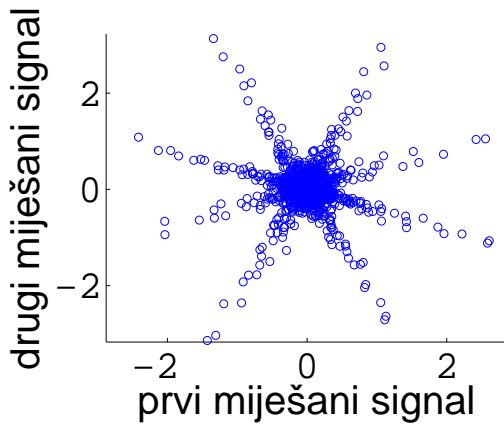
Aproksimacija matrice miješanja

- Radi jednostavnosti uzмимо $E = 0$
- Označimo $X = [x_1 \ \dots \ x_T]$, $A = [a_1 \ \dots \ a_n]$,
 $S = [s_1 \ \dots \ s_n]^T$. Za $1 \leq t \leq T$ je

$$x_t = s_{1,t} a_1 + \dots + s_{n,t} a_n$$

- Rijetkost: za većinu i , $1 \leq i \leq n$, je $s_{i,t} = 0$ (mali broj komponentata s_i je “aktivan” u trenutku t)
- $k =$ (prosječan) broj aktivnih komponentata:
 - $k = 1$ (“single-dominant case”) - grupiranje podataka
 - $k > 1$ (“multiple-dominant case”) - clustering ne pomaže; standardni pristup - alternirajuća optimizacija po A i S , uz ograničenje rijetkosti na S

Grupiranje podataka (clustering)



Klasični clustering problem

- Klasični clustering problem:

$$x_1, \dots, x_n \in \mathbb{R}^d, k \in \mathbb{N}, d(x, y) = \|x - y\|_2;$$

$$E = \sum_{i=1}^n \min_{j=1, \dots, k} d(x_i, C_j) \rightarrow \min_{C_1, \dots, C_k \in \mathbb{R}^d}$$

- Osnovne činjenice: NP-težak (čak i za $k = 2$ ili $d = 2$); za fiksne k i d može se egzaktno riješiti u vremenu $O(n^{dk+1} \log n)$ [4]?
- Klasični clustering problem je optimizacijski problem: minimiziramo nekonveksnu funkciju (problem lokalnih minimuma)
- Najbolje čemu se možemo nadati je lokalni minimum (za koji ne znamo koliko dobro aproksimira globalni)

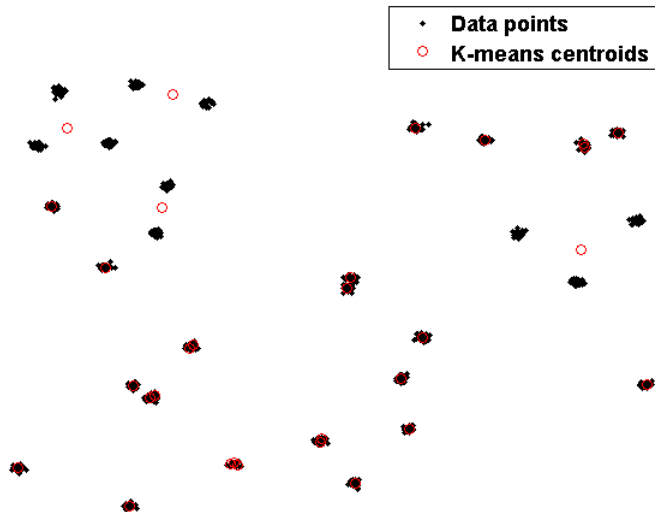
Sadržaj

- 1 Uvod
- 2 Analiza rijetkih komponenata
 - Grupiranje podataka (clustering). Osnovne metode
 - Faktorizacija nenegativnih matrica (NMF)
 - Rijetka rekonstrukcija
- 3 Primjeri

k-sredine (k-means)

- k-means [5] - najjednostavniji i najpoznatiji clustering algoritam:
 - inicijalizacija (najčešće slučajna) centara C_1, \dots, C_k
 - ponavljaj do konvergencije:
 - 1 pridruži s_i j -tom centru C_j t.d. je
$$d(x_i, C_j) = \min_{l=1, \dots, k} d(x_i, C_l)$$
 - 2 $l(C_j) =$ točke pridružene centru C_j ; $C_j \leftarrow \sum_{i \in l(C_j)} x_i / |l(C_j)|$
- d je ovdje euklidska udaljenost; inače treba prilagoditi korak (2)
- Konvergira k lokalnom minimumu ciljne funkcije E u konačno mnogo iteracija
- Složenost: $O(nkd)$ po iteraciji
- Osnovna mana: osjetljivost na inicijalizaciju zbog lokalnih minimuma

k-means



k-means

- Ponekad je korisno koristiti kosinusnu mjeru udaljenosti:
 $d(x, y) = 1 - \frac{|\langle x, y \rangle|}{\|x\| \|y\|}$; treba prilagoditi k-means algoritam
- Za $x_{i_1}, \dots, x_{i_j} \in I(C_j)$ tražimo $\arg \min_{C_j} \sum_p d(x_{i_p}, C_j)$, odnosno
 $\arg \max_{C_j} \sum_p \frac{|\langle x_{i_p}, C_j \rangle|}{\|x_{i_p}\| \|C_j\|}$
- Označimo $m_j = \frac{1}{n_j} \sum_{x \in I(C_j)} x$, gdje je $n_j = |I(C_j)|$. Centroide računamo s $C_j = \frac{m_j}{\|m_j\|}$. Uočimo: za svaki z , $\|z\|_2 = 1$, iz Cauchyjeve nejednakosti slijedi

$$\sum_{x \in I(C_j)} x^T z \leq \sum_{x \in I(C_j)} x^T C_j \quad (1)$$

- Treba pokazati da za ovakav odabir centroida algoritam konvergira

k-means

- S $C_j^{(t)}$ označimo vrijednost C_j u t-toj iteraciji, a s $I_j^{(t+1)}$ indekse pripadnih točaka,
 $I_j^{(t+1)} = \{x : x^\tau C_j^{(t)} > x^\tau C_l^{(t)}, l \neq j\}$. Vrijedi:

$$\begin{aligned} \sum_{j=1}^k \sum_{x \in I_j^{(t)}} x^\tau C_j^{(t)} &= \sum_{j=1}^k \left(\sum_{l=1}^k \sum_{x \in I_j^{(t)} \cap I_l^{(t+1)}} x^\tau C_j^{(t)} \right) \\ &\leq \sum_{j=1}^k \left(\sum_{l=1}^k \sum_{x \in I_j^{(t)} \cap I_l^{(t+1)}} x^\tau C_l^{(t)} \right) \\ &= \sum_{l=1}^k \sum_{j=1}^k \sum_{x \in I_j^{(t)} \cap I_l^{(t+1)}} x^\tau C_l^{(t)} \\ &= \sum_{l=1}^k \sum_{x \in I_l^{(t+1)}} x^\tau C_l^{(t)} \\ &\stackrel{(1)}{\leq} \sum_{l=1}^k \sum_{x \in I_l^{(t+1)}} x^\tau C_l^{(t+1)} \end{aligned}$$

- Dakle, algoritam konvergira (općenito k lokalnom minimumu) [6]

k-means varijacije

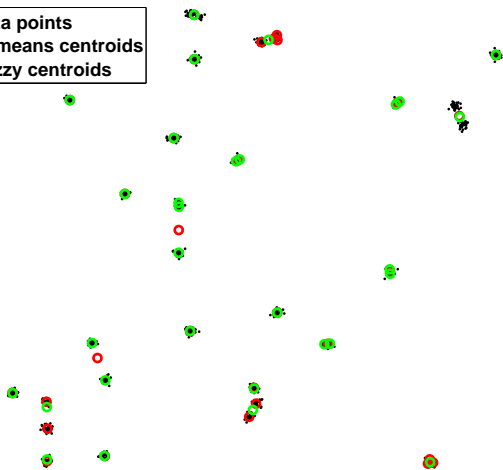
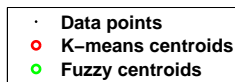
Algoritmi temeljeni na centru

- Fuzzy k-means (c-means):

$$E(U, C) = \sum_{i=1}^n \sum_{j=1}^k u_{ij}^q d(x_i, C_j) \rightarrow \min_{U, C}, \text{ gdje je:}$$

- $U \in \mathbb{R}^{n \times k}$, $\sum_{i=1}^n u_{ij} > 0$, $\sum_{j=1}^k u_{ij} = 1$; u_{ij} određuje nivo pripadnosti točke x_i clusteru C_j ("soft assignment")
- $q > 1$; parametar q kontrolira nivo pripadnosti ("fuzziness") točaka grupama
- Za $q = 1$ algoritam je sličan k-meansu

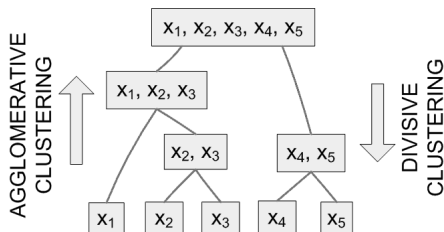
Fuzzy k-means



Hijerarhijski clustering

- Particijski clustering - particija skupa podataka (npr. kmeans), hijerarhijski - niz ugnježđenih particija

- Osnovna podjela: *aglomerativni* (češći) i divizivni



- Zaustavljanje: zadani broj clustera, prevelika udaljenost među clusterima...
- Prednosti nad particijskim metodama: ne zahtijeva broj clustera kao ulaz, radi s proizvoljnom metrikom

Hijerarhijski clustering

- Lance-Williams formula:

$$D(C, C_i \cup C_j) = \alpha_i D(C, C_i) + \alpha_j D(C, C) + \\ + \beta D(C_i, C_j) + \gamma |D(C, C_i) - D(C, C_j)|$$

- Neki posebni slučajevi:
 - $D(C_1, C_2) = \min_{x \in C_1, y \in C_2} d(x, y)$ ("single-link")
 - $D(C_1, C_2) = \max_{x \in C_1, y \in C_2} d(x, y)$ ("complete-link")
 - $D(C_1, C_2) = \frac{1}{|C_1||C_2|} \sum_{x \in C_1, y \in C_2} d(x, y)$ ("group-average")
- Neki hijerarhijski algoritmi zahtijevaju $O(n^2)$ memorijskog prostora (matrica udaljenosti), što je nepraktično za velike n

Spektralni clustering [7]

- Uzmimo $k = 2$; $\{C_1, C_2\}$ particija skupa podataka, y indikator particije ($y_i = \begin{cases} 1 & , x_i \in C_1 \\ -1 & , x_i \in C_2 \end{cases}$), $w_{ij} = \exp(-d^2(x_i, x_j)/\sigma^2)$, $W = (w_{ij})$, $D = \text{diag}(d_1, \dots, d_n)$, $d_i = \sum_j w_{ij}$; želimo

$$\frac{1}{4} \sum_{i \in C_1, j \in C_2} w_{ij} (y_i - y_j)^2 = \frac{1}{4} y^\tau (D - W) y \rightarrow \min_{y \in \{-1, 1\}^n}$$

- $L \doteq D - W$; uz dodatnu pretpostavku $|C_1| = |C_2|$, imamo dodatni uvjet na y : $\sum_i y_i = 0$, odn. uz oznaku $e = [1, \dots, 1]^\tau$, $y \cdot e = 0$; dakle imamo diskretni problem $y^\tau L y \rightarrow \min_{y \in \{-1, 1\}^n, y \cdot e = 0}$ koji aproksimiramo neprekidnim $y^\tau L y \rightarrow \min_{\|y\|_2=1, y \cdot e = 0}$ čije je rješenje drugi najmanji svojstveni vektor od L

Spektralni clustering

- Drugi pristup [8]: definiramo $\text{cut}(C_1, C_2) = \sum_{i \in C_1, j \in C_2} w_{ij}$, $V = \{s_1, \dots, s_n\}$; minimiziramo funkciju

$$\text{Ncut}(C_1, C_2) = \frac{\text{cut}(C_1, C_2)}{\text{cut}(C_1, V)} + \frac{\text{cut}(C_1, C_2)}{\text{cut}(C_2, V)}$$

- Uz prethodne oznake, slično kao prije, rješenje možemo aproksimirati rješavanjem

$$\min_{y^T D e = 0} \frac{y^T L y}{y^T D y}$$

Rješenje je drugi najmanji svojstveni vektor generaliziranog problema svojstvenih vrijednosti $Ly = \lambda Dy$

- Proširenje na $k > 2$: ponavljamo algoritam na dobivenim clusterima
- Problem: L je $n \times n$ matrica (\rightarrow općenito $O(n^2)$ memorijskog prostora)

Bayesovske tehnike

Grupiranje bazirano na statističkom modelu podataka

- Podaci su slučajni uzorak iz neke distribucije s gustoćom f
- Najčešće - mješavine Gausovskih distribucija ("mixtures of Gaussians"):

$$f(x) = \sum_{j=1}^k \pi_j \Phi(x|\mu_j, \Sigma_j) \quad (2)$$

gdje je

$$\Phi(x|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} \det \Sigma} \exp\left(-\frac{1}{2}(x - \mu)^\tau \Sigma^{-1}(x - \mu)\right)$$

gustoća višedimenzionalne normalne distribucije

Bayesovske tehnike

- Motivacija: z binarna slučajna varijabla, jedan element je $= 1$, svi ostali $= 0$; neka je gustoća od z $P(z_k = 1) = \pi_k$, $0 \leq \pi_k \leq 1$, $\sum_k \pi_k = 1$; dakle, $P(z) = \prod_k \pi_k^{z_k}$; z je *latentna* varijabla
- Za distribuciju podataka x pretpostavljamo $P(x|z_k = 1) = N(x|\mu_k, \Sigma_k)$; slijedi

$$P(x) = \sum_z P(z) P(x|z) = \sum_k \pi_k N(x|\mu_k, \Sigma_k)$$

- Na ovaj način moguće je modelirati podatke raznim distribucijama
- Parametre μ_k, Σ_k procjenjujemo metodom maksimalne vjerodostojnosti: maksimiziramo logaritam funkcije vjerodostojnosti

$$\ln P(X|\pi, \mu, \Sigma) = \sum_{i=1}^N \ln \left\{ \sum_k \pi_k N(x_i|\mu_k, \Sigma_k) \right\} \quad (3)$$

gdje je $X = \{x_1, \dots, x_n\}$ skup podataka (slučajni uzorak)

Bayesovske tehnike

Expectation-maximization (EM) algoritam [9]

- Maksimizacija funkcije u (3) je netrivialna
- EM je općeniti algoritam za procjenu parametara vjerojatnosnih modela
- EM za mješavine Gausovskih distribucija:
 - 1 Inicijalizacija parametara μ_k , Σ_k i π_k
 - 2 "E-korak":
$$\gamma(z_{ik}) = P(z_k = 1 | x_i) = \frac{P(z_k=1)P(x_i|z_k=1)}{\sum_j P(z_j=1)P(x_i|z_j=1)} = \frac{\pi_k N(x_i | \mu_k, \Sigma_k)}{\sum_j \pi_j N(x_i | \mu_j, \Sigma_j)}$$
 - 3 "M-korak": update parametara, $\mu_k^{new} \leftarrow \frac{1}{N_k} \sum_i \gamma(z_{ik}) x_i$,
 $\Sigma_k \leftarrow \frac{1}{N_k} \sum_i \gamma(z_{ik}) (x_i - \mu_k^{new})(x_i - \mu_k^{new})^T$, $\pi_k \leftarrow \frac{N_k}{n}$, gdje je $N_k = \sum_i \gamma(z_{ik})$
 - 4 Ponavljanje (2) i (3) do konvergencije
- M-korak: gledamo stacionarne točke funkcije vjerodostojnosti (3)

Bayesovske tehnike

Općeniti EM algoritam

- X podaci, Z latentne varijable, Θ parametri, $P(X, Z|\Theta)$ zajednička distribucija određena parametrima Θ ; cilj je maksimizirati funkciju vjerodostojnosti $P(X|\Theta)$ po Θ
- EM algoritam:
 - 1 Inicijalizacija parametara, Θ^{old}
 - 2 E-korak: $P(Z|X, \Theta^{old})$
 - 3 M-korak: $\Theta^{new} = \arg \max_{\Theta} \sum_Z P(Z|X, \Theta^{old}) \ln P(X, Z|\Theta)$
 - 4 Ponavljanje (2) i (3) do konvergencije
- Fuzzy k-means je vrlo specijalan slučaj EM-a

Problemi

Preprocesiranje skupa podataka

- Glavni problem: veličina skupa podataka; korisni algoritmi prije clusteringa reduciraju broj (možda i dimenzionalnost) podataka
- Neke metode redukcije skupa podataka:
 - “random subsampling”
 - particioniranje [16]: ideja je particionirati skup podataka na p particija veličine $\frac{n}{p}$; zatim se radi grupiranje u $\frac{n}{pq}$ grupa, za neki $q > 1$, na svakoj particiji posebno; zatim se dobivenih $\frac{n}{q}$ grupa spoji prema nekom od kriterija, kao u hijerarhijskom grupiranju

Redukcija skupa podataka

Random subsampling

- Ideja: slučajno (uniformno) se generira podskup $D = \{x_1, \dots, x_s\}$ skupa podataka X
- Pitanje: koliki je minimalan s da s velikom vjerojatnošću za svaku grupu $C_j \subseteq X$ vrijedi $|C_j \cap D| \geq f \cdot |C_j|$, $0 \leq f \leq 1$ unaprijed zadan
- Pretpostavke: X_j slučajna varijabla, C neki cluster,
$$X_j = \begin{cases} 1 & x_j \in C \\ 0 & x_j \notin C \end{cases} \Rightarrow P(X_j = 1) = \frac{|C|}{n}, X_j \text{ nezavisne}$$
Bernoullijeve
- $X = \sum_j X_j$ je broj točaka u $C \cap D$; to je binomna slučajna varijabla; aproksimiramo ju Poissonovom (k njoj i teži za $n \rightarrow \infty$)

Redukcija skupa podataka

Random subsampling

- Chernoff-ljeva ograda [2]: za X Poissonovu, $0 < \varepsilon \leq 1$, uz oznaku $\mu = E(X)$, vrijedi

$$P(X < (1 - \varepsilon)\mu) < \exp\left(\frac{-\mu\varepsilon^2}{2}\right)$$

- Za zadani δ dobijemo minimalni s za koji je $P(X < f \cdot |C|) < \delta$ [16] ($\mu = \frac{s|C|}{n}$)

Sadržaj

- 1 Uvod
- 2 **Analiza rijetkih komponenata**
 - Grupiranje podataka (clustering). Osnovne metode
 - **Faktorizacija nenegativnih matrica (NMF)**
 - Rijetka rekonstrukcija
- 3 Primjeri

Konveksni NMF

- Želimo $X = AS$, S rijetka (i možda nenegativnost A i S); alternirajuća minimizacija; glavni problem: lokalni minimumi, suboptimalno rješenje
- $A = XW$, odnosno $a_i = \sum_j w_{ji}x_j$; $W \geq 0$; dakle, $X = XWS$; stupce od A modeliramo kao centroide [17]
- Prednost nad k-meansom: k-means staje kad nema promjene pripadnosti točkaka clusterima jer bi promjena *jedne* točke povećala grešku; ali, promjena *više* točkaka bi ju možda smanjila! K-means je vrlo "lokalan" algoritam
- Dodatna prednost: može biti i $k > 1$, iako ovdje ne razmatramo taj slučaj

Konveksni NMF

- Minimiziramo

$$J = \|X - XWS\|_F^2 \quad (4)$$

po W i S , uz uvjet $W \geq 0$

Definicija

Neka je $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Funkcija $g : \mathbb{R}^{2n} \rightarrow \mathbb{R}$ je pomoćna funkcija za f ako vrijedi

$$g(x, y) \geq f(x), g(x, x) = f(x)$$

za sve $x, y \in \mathbb{R}^n$.

- Ako definiramo $x^{(t+1)} = \arg \min_x g(x, x^{(t)})$, vidi se da vrijedi $f(x^{(t)}) = g(x^{(t)}, x^{(t)}) \geq g(x^{(t+1)}, x^{(t)}) \geq f(x^{(t+1)})$, dakle niz $f(x^{(t)})$ je opadajući

Konveksni NMF

Teorem

Uz oznake $B = X^\tau X S^\tau$, $A = X^\tau X$, $C = S S^\tau$, funkcija

$$\begin{aligned} Z(W, \tilde{W}) = & \operatorname{Tr}(A) - \sum_{i,k} 2B_{ik}^+ \tilde{W}_{ik} \left(1 + \log \frac{W_{ik}}{\tilde{W}_{ik}}\right) \\ & + \sum_{ik} B_{ik}^- \frac{W_{ik}^2 + \tilde{W}_{ik}^2}{\tilde{W}_{ik}} + \sum_{ik} \frac{(A^+ \tilde{W} C)_{ik} W_{ik}^2}{\tilde{W}_{ik}} \\ & - \sum_{ijkl} A_{ij}^- \tilde{W}_{jk} C_{kl} \tilde{W}_{il} \left(1 + \log \frac{W_{jk} W_{il}}{\tilde{W}_{jk} \tilde{W}_{il}}\right) \end{aligned} \quad (5)$$

gdje je $W \geq 0$, je pomoćna funkcija za $J = J(W)$ iz (4).

Konveksni NMF

- $Z(W, \tilde{W})$ je konveksna funkcija od W ; minimum se postiže u

$$W_{ik} = \tilde{W}_{ik} \sqrt{\frac{[(X^\tau X)^+ S^\tau]_{ik} + [(X^\tau X)^- \tilde{W} S S^\tau]_{ik}}{[(X^\tau X)^- S^\tau]_{ik} + [(X^\tau X)^+ \tilde{W} S S^\tau]_{ik}}} \quad (6)$$

- Treba provjeriti da li fiksna točka zadovoljava nužne uvjete optimalnosti (KKT uvjeti) za početni problem
- Ako da, možemo očekivati konvergenciju k lokalnom minimumu
- Pokazuje se da je manja osjetljivost na početne uvjete u odnosu na k-means (barem empirijski)

Konveksni NMF

- Uvjeti optimalnosti su

$$\begin{aligned}W &\geq 0 \\ \lambda = (\lambda_{ij}) &\geq 0 \\ (\nabla_W J(W))_{ij} - \lambda_{ij} &= (-2X^\tau X S^\tau + 2X^\tau X W S S^\tau)_{ij} - \lambda_{ij} = 0 \\ \lambda_{ij} W_{ij} &= 0\end{aligned}$$

- Zadnja dva uvjeta povlače $(-X^\tau X S^\tau + X^\tau X W S S^\tau)_{ij} W_{ij} = 0$
- Fiksna točka iteracija (6) zadovoljava $(-X^\tau X S^\tau + X^\tau X W S S^\tau)_{ij} W_{ij}^2 = 0$, što je ekvivalentno gornjem uvjetu
- Uvjet $W \geq 0$ je implicitno zadovoljen ako je $W^{(0)} \geq 0$ (početna vrijednost)
- Isti postupak i za S

Višeslojni lokalni NMF

- Ciljna funkcija: $D(X||AS) = \frac{1}{2} \|X - AS\|_F^2 + \alpha J(S)$;
minimiziramo po A i S ; $J(S)$ je regularizacijski član koji forsira rijetkost od S
- Definiramo lokalne ciljne funkcije:

$$D^{(m)}(X^{(m)}||a_m \underline{s}_m) = \frac{1}{2} \left\| X^{(m)} - a_m \underline{s}_m \right\|_F^2 + \alpha^{(m)} J(\underline{s}_m)$$

gdje je a_m m -ti stupac od A , \underline{s}_m m -ti redak od S , i

$$X^{(m)} = X - \sum_{j \neq m} a_j \underline{s}_j$$

- $J(\underline{s}_m) = \sum_t s_{mt}$
- $D^{(m)}(X^{(m)}||a_m \underline{s}_m)$ je konveksna funkcija po a_m i \underline{s}_m posebno

Višeslojni lokalni NMF

- Alternirajuća minimizacija:

$$\begin{aligned}\underline{s}_m &\leftarrow [a_m^\tau X^{(m)} - \alpha e]_+, \quad m = 1, \dots, M \\ A &\leftarrow [XS^\tau (SS^\tau + \lambda I)^{-1}]_+ \\ a_m &\leftarrow \frac{a_m}{\|a_m\|_2}, \quad m = 1, \dots, M\end{aligned}$$

gdje je e vektor jedinica, $[\xi]_+ = \max(\xi, \varepsilon)$, ε mala konstanta (npr. $\varepsilon = 10^{-16}$), M broj redaka od S , i λ regularizacijska konstanta

- Empirijski se pokazuje da se efikasnost ovog algoritma može značajno poboljšati uzastopnom primjenom (u više “slojeva”): $X \approx A^{(1)}S^{(1)}$, $S^{(1)} \approx A^{(2)}S^{(2)}$, ..., $X \approx A^{(1)}A^{(2)} \dots A^{(L)}S^{(L)}$
- Naime, to je jedna od heuristika za izbjegavanje lokalnih minimuma; zajedno s “multi-start” inicijalizacijom empirijski se pokazuje vrlo korisnom [11, 12, 10]

Sadržaj

1 Uvod

2 Analiza rijetkih komponenata

- Grupiranje podataka (clustering). Osnovne metode
- Faktorizacija nenegativnih matrica (NMF)
- Rijetka rekonstrukcija

3 Primjeri

Konveksna aproksimacija

- Clusteringom dobijemo matricu A ; za svaki $1 \leq i \leq n$ želimo naći rijetko rješenje pododređenog linearnog sustava $As_i = x_i$
- Idealna, ali nepraktična mjera rijetkosti je $\|x\|_0 = |\{i : x_i \neq 0\}|$; najčešće se zamjenjuje neprekidnom, konveksnom aproksimacijom $\|x\|_1 = \sum_i |x_i|$ (1-norma)
- Dakle, rješavamo

$$\min_s \|s\|_1 \text{ uz uvjet } \|As - x_i\|_2 \leq \varepsilon \quad (7)$$

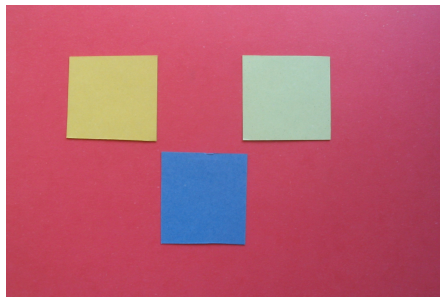
- To je problem konveksne optimizacije, može se riješiti standardnim metodama [3]; može se pokazati [13, 14]: ako je najrjeđe (u smislu mjere $\|\cdot\|_0$) rješenje (odnosno približno rješenje) sustava $As = x_i$ dovoljno rijetko, uz neke prihvatljive pretpostavke na matricu A (npr. koherencija od A , $\mu(A) = \max_{i,j} \frac{|\langle a_i, a_j \rangle|}{\|a_i\|_2 \|a_j\|_2}$, dovoljno mala), to rješenje možemo aproksimirati rješavanjem (7)

Pohlepne metode

- “Matching pursuit” (MP) algoritam:
 - 1 Inicijalizacija: rješenje $s = 0$, rezidual $r = x_i$, normalizacija $\|a_i\|_2 = 1, \forall i$
 - 2 Ponavljaj do konvergencije: $\hat{i} = \arg \max_i \langle r, a_i \rangle$;
 $s_{\hat{i}} \leftarrow s_{\hat{i}} + \langle r, a_{\hat{i}} \rangle a_{\hat{i}}$; $r \leftarrow r - \langle r, a_{\hat{i}} \rangle a_{\hat{i}}$
- Moguće su razne varijacije
- Pokazuje se da konveksna aproksimacija ima bolja svojstva od pohlepnih algoritama (stabilnost, teorijske garancije); prednost pohlepnih algoritama je brzina [15]

Segmentacija višespektralne slike

- Segmentacija slike: particija slike na homogene dijelove
- RGB slika = 3D tenzor ($\in \mathbb{R}^{m \times n \times 3}$); pikseli = vektori u \mathbb{R}^3



- Klasični “toy example” za segmentaciju slike;
rezolucija slike:
 1380×2048

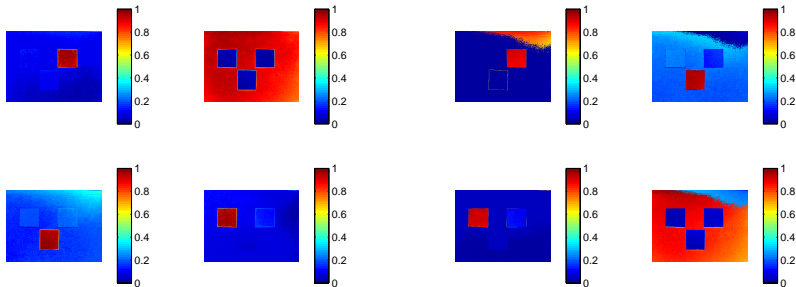
Segmentacija višespektralne slike

- RGB slika $R \in \mathbb{R}^{m \times n \times 3} \rightarrow \hat{R} \in \mathbb{R}^{3 \times mn}$, $\hat{R}(:, i) = R(i_1, i_2, :)$ za neke i_1 i i_2 (“matricizacija” tenzora)
- Interpretacija modela $X = AS + E$:
 - $X = \hat{R}$ (matrica podataka)
 - $A \in \mathbb{R}^{3 \times k}$ - matrica “spektralnih profila” materijala prisutnih na slici (broj materijala k je najčešće nepoznat, ovdje $k = 4$)
 - S - komponente (retci od S) predstavljaju prostornu distribuciju materijala
- Ideja: matricu spektralnih profila aproksimiramo grupiranjem piksela (stupaca matrice \hat{R})
- Nakon grupiranja, za svaki piksel se traži prikaz u obliku rijetke linearne kombinacije dobivenih spektralnih profila

Segmentacija višespektralne slike

- Glavni problem: već za sliku osrednje rezolucije (npr. 1000×1000) imamo 10^6 točaka; matrica udaljenosti bi uzela $O(10^{12})$ prostora
- Nužna je redukcija broja točaka
- Pokazuje se da random subsampling kao metoda preprocesiranja daje dobre rezultate; reduciramo s faktorom 16000
- Hijerarhijsko grupiranje je najpouzdanije jer ne zahtijeva inicijalizaciju; na istom skupu uvijek daje isti rezultat
- Mogući nedostatak metode je što zanemaruje prostorni položaj piksela u slici

Segmentacija višespektralne slike

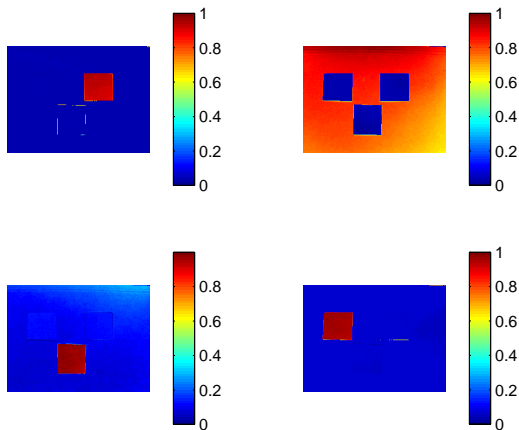


Dobivena segmentacija:
hijerarhijsko grupiranje +
minimizacija 1-norme

Tipična segmentacija dobivena
k-means algoritmom + MP
rekonstrukcija

Segmentacija višespektralne slike

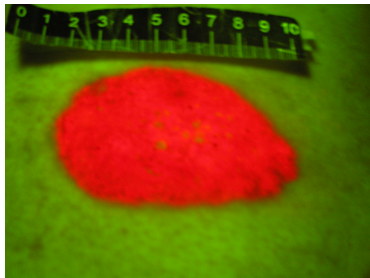
- Hijerarhijsko grupiranje + MP rekonstrukcija



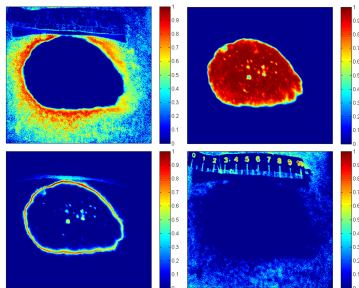
- Brzina izvršavanja: konveksna optimizacija - $\sim 1h$, MP - < 5 sekundi

Segmentacija višespektralne slike

Realni primjer



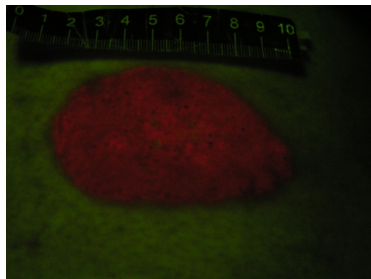
RGB slika kožnog tumora



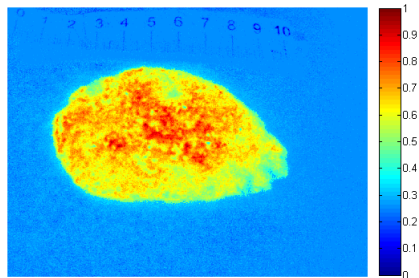
Dobivena segmentacija
višeslojnim lokalnim NMF-om
[10]

Segmentacija višespektralne slike

Realni primjer



Slika niskog intenziteta



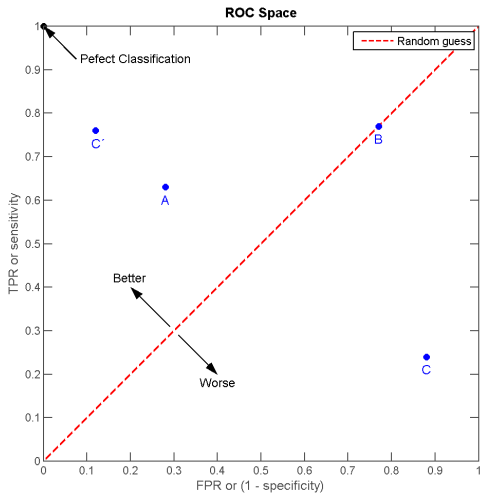
Segmentacija: convex NMF +
MP

Segmentacija višespektralne slike

Realni primjer

- Statistička analiza klasifikacije (segmentacije): za svaki piksel gledamo osjetljivost klasifikacije (tumor ili zdravo tkivo) u odnosu na prag (veće/manje vrijednosti - prisutnost/odsutnost tumora); 4 moguća ishoda klasifikacije:
 - “true positive”: piksel je ispravno klasificiran kao tumor
 - greška tipa 1 (“false positive”): piksel je pogrešno klasificiran kao tumor
 - greška tipa 2 (“false negative”): piksel pogrešno nije klasificiran kao tumor
 - “true negative”: piksel ispravno nije klasificiran kao tumor
- “True positive rate” (TPR): udio ispravno klasificiranih točaka;
- “false positive rate” (FPR): udio pogrešno klasificiranih točaka
- TPR i FPR su funkcije praga
- Grafički prikaz: “Receiver Operating Characteristic” (ROC) krivulja - TPR vs. FPR

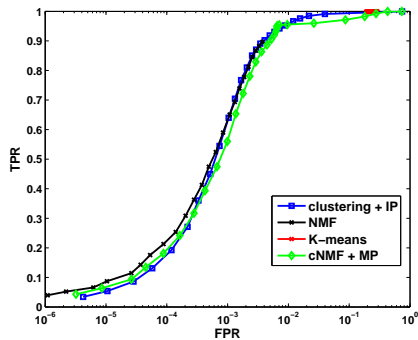
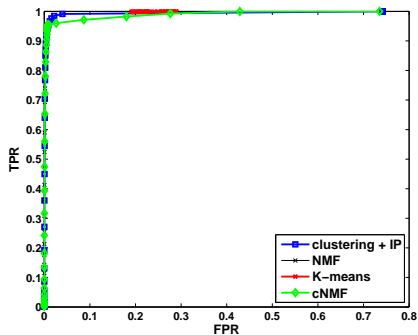
ROC krivulja



http://en.wikipedia.org/wiki/File:ROC_space-2.png

ROC krivulja za segmentaciju

ROC krivulje za sliku tumora niskog intenziteta



Zaključak

- Razdvajanje signala - klasični inverzni problem; analiza rijetkih komponenata - jedna metoda rješavanja
- Rijetkost - grupiranje podataka; druga faza - konveksna optimizacija (računski najzahtjevniji dio! - paralelizacija)
- Grupiranje podataka - NP-težak problem; minimizacija nekonveksne funkcije
- Razne heuristike za izbjegavanje lokalnih minimuma; globalna optimizacija
- Nužna je redukcija skupa podataka
- Nema najbolje metode; ovisi o danom problemu
- Možda najviše teorijski utemeljen pristup: Bayesovski

Literatura I

- [1] G. Gan, C. Ma, J. Wu.
Data clustering: Theory, algorithms and applications.
ASA-SIAM, 2007.
- [2] R. Motwani, P. Raghavan.
Randomized algorithms.
Cambridge University Press, 1995.
- [3] Boyd, Vandenberghe.
Convex optimization.
- [4] M. Inaba, N. Katoh, H. Imai.
Applications of weighted Voronoi diagrams and randomization
to variance-based k-clustering.
10th ACM Symposium on Computational Geometry, 1994.

Literatura II

- [5] J. MacQueen.
Some methods for classification and analysis of multivariate observations.
1967.
- [6] I. S. Dhillon, D. S. Modha.
Concept decompositions for large sparse text data using clustering.
Machine learning, 2001.
- [7] Y. Weiss.
Segmentation using eigenvectors: A unifying view.
Computer Vision, 1999.
- [8] Shi, Malik.
Normalized cuts and image segmentation.
IEEE Trans. PAMI, 2000.

Literatura III

- [9] A. P. Dempster, N. M. Laird, D. B. Rubin.
Maximum likelihood from incomplete data via the EM algorithm.
1977.
- [10] I. Kopriva, A. Cichocki.
Blind decomposition of low-dimensional multi-spectral image by sparse component analysis.
Journal of Chemometrics, 2009.
- [11] A. Cichocki, R. Zdunek, S. Amari.
Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization.
LNCS 4666, pp. 169-176, 2007.
- [12] A. Cichocki, R. Zdunek.
Multilayer nonnegative matrix factorization.
Electronics Letters 42(16), 2006.

Literatura IV

- [13] Gribonval, Nielsen.
Sparse representations in unions of bases.
IEEE Trans. Inf. Theory, 49(12), 2003.
- [14] Candes, Tao.
Decoding by linear programming.
IEEE Trans. Inf. Theory 51, 2005.
- [15] J. A. Tropp, S. J. Wright.
Computational methods for sparse solution of linear inverse problems.
CalTech ACM Tech. Rep. 2009-01
- [16] S. Guha, R. Rastogi, K. Shim.
CURE: An efficient clustering algorithm for large databases.
Information Systems, 26(1), 2001.

- [17] C. Ding, T. Li, M. I. Jordan.
Convex and semi-nonnegative matrix factorizations.
IEEE Trans. PAMI, 2010.