O pewnych modyfikacjach/usprawnieniach algorytmu grupowania spektralnego

Sławomir T. Wierzchoń
http://www.ipipan.waw.pl/~stw/sem080517

Instytut Podstaw Informatyki PAN

Seminarium instytutowe
Warszawa, 8 maja 2017

Treść

Algorytm NJW (Ng, Jordan & Weiss, 2002)

WE: symetryczna macierz S taka, że $s_{ii} = 0$

1. wyznacz $\mathcal{S} = D^{-1/2} S D^{-1/2}$, $D = \mathrm{diag}(S\mathbf{e}) = \mathrm{diag}(\mathrm{d}_1, \ldots, \mathrm{d}_m)$
2. wyznacz $Y = (y_1, \ldots, y_k)$; y_j – j-ty dominujący wektor własny $\mathcal{S}$
3. rzutuj wiersze macierzy Y na sferę jednostkową, tzn. utwórz macierz $Z = [z_{ij}]$, $i = 1, \ldots, m$, $j = 1, \ldots, k$ o elementach

$$z_{ij} = y_{ij} / \|\mathbf{y}_{i\cdot}\|$$

4. wyznacz finalny podział danych reprezentowanych przez macierz Z.

Algorytm NJW (Ng, Jordan & Weiss, 2002)

WE: symetryczna macierz S taka, że $s_{ii} = 0$

1. wyznacz $\mathcal{S} = D^{-1/2} S D^{-1/2}$,
   $D = \mathrm{diag}(S\mathbf{e}) = \mathrm{diag}(\mathsf{d}_1, \ldots, \mathsf{d}_m)$
2. wyznacz $Y = (y_1, \ldots, y_k)$; y_j – j-ty dominujący wektor własny $\mathcal{S}$
3. rzutuj wiersze macierzy Y na sferę jednostkową, tzn. utwórz macierz $Z = [z_{ij}]$, $i = 1, \ldots, m$, $j = 1, \ldots, k$ o elementach

$$z_{ij} = y_{i,j} / \|\mathbf{y}_{i\cdot}\|$$

4. wyznacz finalny podział danych reprezentowanych przez macierz Z.

Algorytm NJW (Ng, Jordan & Weiss, 2002)

WE: symetryczna macierz S taka, że $s_{ii} = 0$

1. wyznacz $\mathcal{S} = D^{-1/2} S D^{-1/2}$,
   $D = \mathrm{diag}(S\mathbf{e}) = \mathrm{diag}(\mathrm{d}_1, \ldots, \mathrm{d}_m)$
2. wyznacz $Y = (y_1, \ldots, y_k)$; y_j – j-ty dominujący wektor własny $\mathcal{S}$
3. rzutuj wiersze macierzy Y na sferę jednostkową, tzn. utwórz macierz $Z = [z_{ij}]$, $i = 1, \ldots, m$, $j = 1, \ldots, k$ o elementach

$$z_{ij} = y_{i,j} / \|\mathbf{y}_{i\cdot}\|$$

4. wyznacz finalny podział danych reprezentowanych przez macierz Z.

Algorytm NJW (Ng, Jordan & Weiss, 2002)

WE: symetryczna macierz S taka, że $s_{ii} = 0$

1. wyznacz $\mathcal{S} = D^{-1/2} S D^{-1/2}$,
   $D = \mathrm{diag}(S\mathbf{e}) = \mathrm{diag}(\mathrm{d}_1, \ldots, \mathrm{d}_m)$
2. wyznacz $Y = (y_1, \ldots, y_k)$; y_j – j-ty dominujący wektor własny $\mathcal{S}$
3. rzutuj wiersze macierzy Y na sferę jednostkową, tzn. utwórz macierz $Z = [z_{ij}]$, $i = 1, \ldots, m$, $j = 1, \ldots, k$ o elementach

$$z_{ij} = y_{i,j} / \|\mathbf{y}_{i\cdot}\|$$

4. wyznacz finalny podział danych reprezentowanych przez macierz Z.

Algorytm NJW (Ng, Jordan & Weiss, 2002)

WE: symetryczna macierz S taka, że $s_{ii} = 0$

1. wyznacz $\mathcal{S} = D^{-1/2} S D^{-1/2}$,
   $D = \mathrm{diag}(S\mathbf{e}) = \mathrm{diag}(\mathrm{d}_1, \ldots, \mathrm{d}_m)$
2. wyznacz $Y = (y_1, \ldots, y_k)$; y_j – j-ty dominujący wektor własny $\mathcal{S}$
3. rzutuj wiersze macierzy Y na sferę jednostkową, tzn. utwórz macierz $Z = [z_{ij}]$, $i = 1, \ldots, m$, $j = 1, \ldots, k$ o elementach

$$z_{ij} = y_{i,j} / \|\mathbf{y}_{i\cdot}\|$$

4. wyznacz finalny podział danych reprezentowanych przez macierz Z.

Algorytm NJW – dlaczego działa?

Dane są 3 rozłączne grupy C_1, C_2, C_3

$$S = \begin{bmatrix} S^{(1)} & 0 & 0 \\ 0 & S^{(2)} & 0 \\ 0 & 0 & S^{(3)} \end{bmatrix} \quad \Rightarrow \quad \mathcal{L} = \begin{bmatrix} \mathcal{L}^{(1)} & 0 & 0 \\ 0 & \mathcal{L}^{(2)} & 0 \\ 0 & 0 & \mathcal{L}^{(3)} \end{bmatrix}$$

Dominujące wektory własne

$$Y = \begin{bmatrix} y^{(1)} & 0 & 0 \\ 0 & y^{(2)} & 0 \\ 0 & 0 & y^{(3)} \end{bmatrix}, y^{(i)} = \sqrt{D^{(i)}}\mathbf{e}_{C_i} \Rightarrow Z = \begin{bmatrix} \mathbf{e}_{C_1} & 0 & 0 \\ 0 & \mathbf{e}_{C_2} & 0 \\ 0 & 0 & \mathbf{e}_{C_3} \end{bmatrix}$$

Algorytm NJW – dlaczego działa?

Dane są 3 rozłączne grupy C_1, C_2, C_3

$$S = \begin{bmatrix} S^{(1)} & 0 & 0 \\ 0 & S^{(2)} & 0 \\ 0 & 0 & S^{(3)} \end{bmatrix} \quad \Rightarrow \quad \mathcal{L} = \begin{bmatrix} \mathcal{L}^{(1)} & 0 & 0 \\ 0 & \mathcal{L}^{(2)} & 0 \\ 0 & 0 & \mathcal{L}^{(3)} \end{bmatrix}$$

Dominujące wektory własne

$$Y = \begin{bmatrix} y^{(1)} & 0 & 0 \\ 0 & y^{(2)} & 0 \\ 0 & 0 & y^{(3)} \end{bmatrix}, y^{(i)} = \sqrt{D^{(i)}}\mathbf{e}_{C_i} \Rightarrow Z = \begin{bmatrix} \mathbf{e}_{C_1} & 0 & 0 \\ 0 & \mathbf{e}_{C_2} & 0 \\ 0 & 0 & \mathbf{e}_{C_3} \end{bmatrix}$$

Algorytm NJW - wyznaczanie finalnego podziału danych

Zastosuj:

1. algorytm k-średnich,
2. algorytm rotacji spektralnej (Huang, Nie, & Huang, 2013): $\min_{Q,H} \|YQ - H\|_F$ p.o. $Q^{\mathrm{T}}Q = \mathbb{I}$, $H \in \{0,1\}^{m\times k}$,
3. QR faktoryzację (Zha, He, Ding, Gu, & Simon, 2001), (Freder, & van Barel, 2010), (Damle, Minden, & Ying)

```
[Q,R,P]=qr(Y');  % Q'Q = I, P - macierz permut.
u=inv(R(1:K,1:K))*R*P';
[tmp,r1]=max(abs(u'),[],2);
```

Dla $Z \in \mathbb{R}^{2000\times 3}$: wariant (1) trwa 0.03737 ± 0.00958, a randomizowany wariat (3): $10^{-4}(4.8034 \pm 5.0504) \approx 0.001067$ (średnie z 50 powtórzeń).

Algorytm NJW - wyznaczanie finalnego podziału danych

Zastosuj:

1. algorytm k-średnich,
2. algorytm rotacji spektralnej (Huang, Nie, & Huang, 2013): $\min_{Q,H} \|YQ - H\|_F$ p.o. $Q^{\mathrm{T}}Q = \mathbb{I}$, $H \in \{0,1\}^{m\times k}$,
3. QR faktoryzację (Zha, He, Ding, Gu, & Simon, 2001), (Freder, & van Barel, 2010), (Damle, Minden, & Ying)

```
[Q,R,P]=qr(Y');  %Q'Q = I, P – macierz permut.
u=inv(R(1:K,1:K))*R*P';
[tmp,r1]=max(abs(u'),[],2);
```

Dla $Z \in \mathbb{R}^{2000\times 3}$: wariant (1) trwa 0.03737 ± 0.00958, a randomizowany wariat (3): $10^{-4}(4.8034 \pm 5.0504) \approx 0.001067$ (średnie z 50 powtórzeń).

Algorytm NJW - wyznaczanie finalnego podziału danych

Zastosuj:

1. algorytm k-średnich,
2. algorytm rotacji spektralnej (Huang, Nie, & Huang, 2013): $\min_{Q,H} \|YQ - H\|_F$ p.o. $Q^{\mathrm{T}}Q = \mathbb{I}$, $H \in \{0,1\}^{m\times k}$,
3. QR faktoryzację (Zha, He, Ding, Gu, & Simon, 2001), (Freder, & van Barel, 2010), (Damle, Minden, & Ying)

```
[Q,R,P]=qr(Y');  %Q'Q = I, P – macierz permut.
u=inv(R(1:K,1:K))*R*P';
[tmp,r1]=max(abs(u'),[],2);
```

Dla $Z \in \mathbb{R}^{2000\times 3}$: wariant (1) trwa 0.03737 ± 0.00958, a randomizowany wariat (3): $10^{-4}(4.8034 \pm 5.0504) \approx 0.001067$ (średnie z 50 powtórzeń).

Algorytm NJW - wyznaczanie finalnego podziału danych

Zastosuj:

1. algorytm k-średnich,
2. algorytm rotacji spektralnej (Huang, Nie, & Huang, 2013): $\min_{Q,H} \|YQ - H\|_F$ p.o. $Q^{\mathrm{T}}Q = \mathbb{I}$, $H \in \{0,1\}^{m\times k}$,
3. QR faktoryzację (Zha, He, Ding, Gu, & Simon, 2001), (Freder, & van Barel, 2010), (Damle, Minden, & Ying)
   [Q,R,P]=qr(Y^{T}); %$Q^{\mathrm{T}}Q = \mathbb{I}$, P – macierz permut.
   u=inv(R(1:K,1:K))*R*P';
   [tmp,r1]=max(abs(u^{T}),[],2);

Dla $Z \in \mathbb{R}^{2000\times 3}$: wariant (1) trwa 0.03737 ± 0.00958, a randomizowany wariat (3): $10^{-4}(4.8034 \pm 5.0504) \approx 0.001067$ (średnie z 50 powtórzeń).

Algorytm NJW - wyznaczanie finalnego podziału danych

Zastosuj:

1. algorytm k-średnich,
2. algorytm rotacji spektralnej (Huang, Nie, & Huang, 2013): $\min_{Q,H} \|YQ - H\|_F$ p.o. $Q^{\mathrm{T}}Q = \mathbb{I}$, $H \in \{0,1\}^{m\times k}$,
3. QR faktoryzację (Zha, He, Ding, Gu, & Simon, 2001), (Freder, & van Barel, 2010), (Damle, Minden, & Ying)
   [Q,R,P]=qr(Y^{T}); %$Q^{\mathrm{T}}Q = \mathbb{I}$, P – macierz permut.
   u=inv(R(1:K,1:K))*R*P';
   [tmp,r1]=max(abs(u^{T}),[],2);

Dla $Z \in \mathbb{R}^{2000\times 3}$: wariant (1) trwa 0.03737 ± 0.00958, a randomizowany wariat (3): $10^{-4}(4.8034 \pm 5.0504) \approx 0.001067$ (średnie z 50 powtórzeń).

Algorytm NJW - wyznaczanie finalnego podziału danych

Zastosuj:

1. algorytm k-średnich,
2. algorytm rotacji spektralnej (Huang, Nie, & Huang, 2013): $\min_{Q,H} \|YQ - H\|_F$ p.o. $Q^{\mathrm{T}}Q = \mathbb{I}$, $H \in \{0,1\}^{m\times k}$,
3. QR faktoryzację (Zha, He, Ding, Gu, & Simon, 2001), (Freder, & van Barel, 2010), (Damle, Minden, & Ying)
   $[\mathrm{Q,R,P}]=\mathrm{qr}(\mathrm{Y^T})$; %$Q^{\mathrm{T}}Q = \mathbb{I}$, P – macierz permut.
   $\mathrm{u=inv(R(1{:}K,1{:}K))*R*P'}$;
   $[\mathrm{tmp,r1}]=\max(\mathrm{abs(u^T)},[],2)$;

Dla $Z \in \mathbb{R}^{2000\times 3}$: wariant (1) trwa 0.03737 ± 0.00958, a randomizowany wariat (3): $10^{-4}(4.8034 \pm 5.0504) \approx 0.001067$ (średnie z 50 powtórzeń).

Algorytm NJW a błądzenie losowe

Jeżeli (λ, ϕ) jest parą własną $\mathcal{S} = D^{-1/2} S D^{-1/2}$ to:

$$D^{-1/2} S D^{-1/2} \phi = \lambda \phi \,|\, * D^{-1/2}$$

$$\underbrace{D^{-1} S}_{P} \underbrace{(D^{-1/2} \phi)}_{\psi} = \lambda \underbrace{(D^{-1/2} \phi)}_{\psi}$$

tzn. $(\lambda, D^{-1/2}\phi)$ jest parą własną macierzy $P = D^{-1} S$.
W kroku (2) algorytmu NJW można liczyć dominujące wektory macierzy P (Meilă, & Shi, 2000).

Algorytm NJW a błądzenie losowe

Jeżeli (λ, ϕ) jest parą własną $\mathcal{S} = D^{-1/2} S D^{-1/2}$ to:

$$D^{-1/2} S D^{-1/2} \phi = \lambda \phi | * D^{-1/2}$$

$$\underbrace{D^{-1} S}_{P} \underbrace{(D^{-1/2} \phi)}_{\psi} = \lambda \underbrace{(D^{-1/2} \phi)}_{\psi}$$

tzn. $(\lambda, D^{-1/2}\phi)$ jest parą własną macierzy $P = D^{-1} S$.
W kroku (2) algorytmu NJW można liczyć dominujące wektory macierzy P (Meilă, & Shi, 2000).

Algorytm NJW a błądzenie losowe

Jeżeli (λ, ϕ) jest parą własną $\mathcal{S} = D^{-1/2} S D^{-1/2}$ to:

$$D^{-1/2} S D^{-1/2} \phi = \lambda \phi | * D^{-1/2}$$

$$\underbrace{D^{-1} S}_{P} \underbrace{(D^{-1/2} \phi)}_{\psi} = \lambda \underbrace{(D^{-1/2} \phi)}_{\psi}$$

tzn. $(\lambda, D^{-1/2}\phi)$ jest parą własną macierzy $P = D^{-1}S$.
W kroku (2) algorytmu NJW można liczyć dominujące wektory macierzy P (Meilă, & Shi, 2000).

Algorytm NJW a błądzenie losowe

Jeżeli (λ, ϕ) jest parą własną $\mathcal{S} = D^{-1/2} S D^{-1/2}$ to:

$$D^{-1/2} S D^{-1/2} \phi = \lambda \phi | * D^{-1/2}$$

$$\underbrace{D^{-1} S}_{P} \underbrace{(D^{-1/2} \phi)}_{\psi} = \lambda \underbrace{(D^{-1/2} \phi)}_{\psi}$$

tzn. $(\lambda, D^{-1/2}\phi)$ jest parą własną macierzy $P = D^{-1} S$.
W kroku (2) algorytmu NJW można liczyć dominujące wektory macierzy P (Meilă, & Shi, 2000).

Wyzwania podejścia spektralnego

- Złożoność pamięciowa: $O(m^2)$,
- Złożoność czasowa $O(m^3)$
- Jak klasyfikować nowe obserwacje?

Wyzwania podejścia spektralnego

- Złożoność pamięciowa: $O(m^2)$,
- Złożoność czasowa $O(m^3)$
- Jak klasyfikować nowe obserwacje?

Wyzwania podejścia spektralnego

- Złożoność pamięciowa: $O(m^2)$,
- Złożoność czasowa $O(m^3)$
- Jak klasyfikować nowe obserwacje?

Wyzwania podejścia spektralnego

- Złożoność pamięciowa: $O(m^2)$,
- Złożoność czasowa $O(m^3)$
- Jak klasyfikować nowe obserwacje?

(Alzate & Suykens, 2010):

- It is not clear what we are optimizing when doing spectral clustering.
- Due to the lack of a clear optimization problem, the parameters selection is not straightforward.
- Clustering of new points should rely on approximation techniques.

PCA & KPCA (Schölkopf, Smola, & Müller, 2005)

PCA

1. $X_c = (x_1 - \mu, \ldots, x_m - \mu)$, $\mu = \frac{1}{m}\sum_{i=1}^{m} x_i$
2. $\Sigma = \mathtt{cov}(X_c) = \frac{1}{m} X_c X_c^{\mathrm{T}}$
3. $[U, \Lambda] = \mathtt{eig}(\Sigma)$, $U_k = (u_1, \ldots, u_k)$,
4. $\mathtt{proj}_j(y) = u_j^{\mathrm{T}} y$

KPCA

1. $\mathrm{K}(x_i, x_j) = \phi(x_i)^{\mathrm{T}}\phi(x_j) = s_{ij}$, $\phi\colon X \to \mathcal{F}$
2. $\mathrm{K}_c = M_c \mathrm{K} M_c$, $M_c = \mathbb{I} - \frac{1}{m}\mathbf{e}\mathbf{e}^{\mathrm{T}}$
3. $[U, \Lambda] = \mathtt{eig}(\mathrm{K}_c)$, $U_k = (u_1, \ldots, u_k)$
4. $\mathtt{proj}_j(y) = \sum_{i=1}^{m} u_{ij} \mathrm{K}_c(y, x_i)$

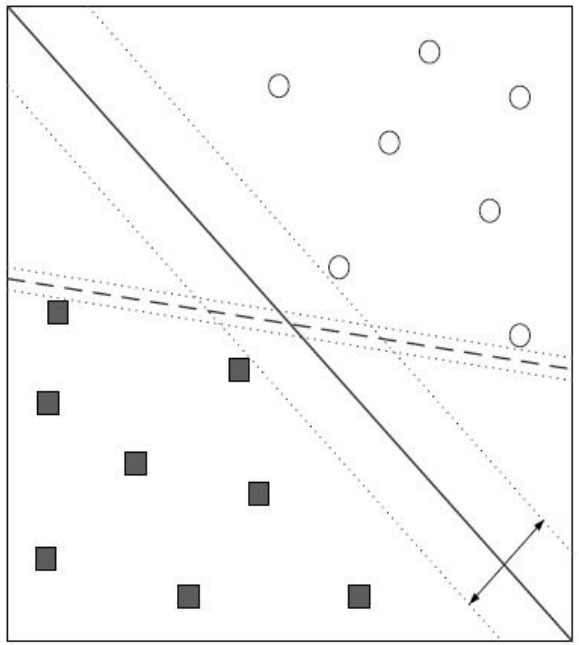
SVM a LS-SVM (1)

$\{(x_i, y_)\}_{i=1}^m$ – zbiór trenujący, $y_i \in \{-1, +1\}$. Jeżeli dane są liniowo separowalne, to szukamy hiperpłaszczyzny $w^{\mathrm{T}}x + b$ takiej, że

$$w^{\mathrm{T}}x_i + b \begin{cases} > 0 & \text{gdy } x_i \in C_+ \\ < 0 & \text{gdy } x_i \in C_- \end{cases}$$

Sformułowanie:

$$\min_w \frac{1}{2} w^{\mathrm{T}} w \quad \text{p.o. } y_i(w^{\mathrm{T}}x_i + b) \geqslant 1$$



SVM a LS-SVM (2)

Gdy dane nie są liniowo separowalne, wprowadza się zmienne dopełniające ξ_i, co prowadzi do zadania:

$$\min_{w,b,e_i} \frac{1}{2} w^{\mathrm{T}} w + c \sum_{i=1}^{m} \xi_i, \quad \text{p.w. } y_i(w^{\mathrm{T}} x_i + b) \geqslant 1 - \xi_i, \xi_i \geqslant 0$$

Jego rozwiązanie nie jest trywialne!
Suykens & Vandewalle (1999) zaproponowali alternatywne sformułowanie

$$\min_{w,b,e_i} \frac{1}{2} w^{\mathrm{T}} w + \frac{1}{2} \sum_{i=1}^{m} e_i^2, \quad \text{p.w. } y_i(w^{\mathrm{T}} x_i + b) = 1 - e_i$$

Ważony jądrowy algorytm PCA

$$\min J(\mathbf{u}_\ell, \boldsymbol{\rho}_\ell, b_\ell) = \frac{1}{2m}\sum_{\ell=1}^{n_k} \gamma_\ell \boldsymbol{\rho}_\ell^{\mathrm{T}} W \boldsymbol{\rho}_\ell - \frac{1}{2}\sum_{\ell=1}^{n_k} \mathbf{u}_\ell^{\mathrm{T}}\mathbf{u}_\ell$$

$$\text{s.t. } \boldsymbol{\rho}_\ell = \mathbf{F}\mathbf{u}_\ell + b_\ell \mathbf{e}, \quad \ell = 1, \ldots, n_k \geqslant k-1$$

gdzie $\boldsymbol{F} = (\phi(x_1); \ldots; \phi(x_m))$, W – macierz wag, np. $W = D^{-1}$.
Lagrangian:

$$\mathfrak{L}(\mathbf{u}_\ell, \boldsymbol{\rho}_\ell, b_\ell) = \frac{1}{2m}\sum_{\ell=1}^{n_k} \gamma_\ell \boldsymbol{\rho}_\ell^{\mathrm{T}} W \boldsymbol{\rho}_\ell - \frac{1}{2}\sum_{\ell=1}^{n_k} \mathbf{u}_\ell^{\mathrm{T}}\mathbf{u}_\ell - \sum_{\ell=1}^{n_k} \boldsymbol{\alpha}_\ell^{\mathrm{T}}(\boldsymbol{\rho}_\ell - \mathbf{F}\mathbf{u}_\ell - b_\ell \mathbf{e})$$

gdzie $\boldsymbol{\alpha}_\ell$ – mnożniki Lagrange'a.

Ważony jądrowy algorytm PCA

Warunki optymalności

1. $\frac{\partial \mathfrak{L}}{\partial \mathbf{u}_\ell} = 0 \Rightarrow \mathbf{u}_\ell = \mathbf{F}^{\mathrm{T}} \boldsymbol{\alpha}_\ell$
2. $\frac{\partial \mathfrak{L}}{\partial \boldsymbol{\rho}_\ell} = 0 \Rightarrow \boldsymbol{\alpha}_\ell = \frac{\gamma}{m} D^{-1} \boldsymbol{\rho}_\ell$
3. $\frac{\partial \mathfrak{L}}{\partial b_\ell} = 0 \Rightarrow \mathbf{e}^{\mathrm{T}} \boldsymbol{\alpha}_\ell$
4. $\frac{\partial \mathfrak{L}}{\partial \boldsymbol{\alpha}_\ell} = 0 \Rightarrow \boldsymbol{\rho}_\ell - \mathbf{F}\mathbf{u}_\ell - b_\ell \mathbf{e} = 0$

Problem dualny

$$WMS\boldsymbol{\alpha}_\ell = \lambda_\ell \boldsymbol{\alpha}_\ell, \quad \ell = 1, \ldots, n_k,$$
$$\lambda_\ell = m/\gamma_\ell,$$
$$S(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{F}\mathbf{F}^{\mathrm{T}}(\mathbf{x}_i, \mathbf{x}_j) = \mathrm{K}(\mathbf{x}_i, \mathbf{x}_j),$$
oraz

$$M = \mathbb{I} - \frac{1}{\sum_{i,j} w_{ij}} \mathbf{e}\mathbf{e}^{\mathrm{T}} W$$

Gdy $W = \mathbb{I} \Rightarrow$ WKPCA=KPCA.

$$\mathbf{b} = -\frac{1}{\sum_{i,j} w_{ij}} \mathbf{e}^{\mathrm{T}} WSA_{n_k}, \; A_{n_k} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}$$

Ważony jądrowy algorytm PCA

Warunki optymalności

1. $\frac{\partial \mathfrak{L}}{\partial \mathbf{u}_\ell} = 0 \Rightarrow \mathbf{u}_\ell = \mathbf{F}^{\mathrm{T}} \boldsymbol{\alpha}_\ell$
2. $\frac{\partial \mathfrak{L}}{\partial \boldsymbol{\rho}_\ell} = 0 \Rightarrow \boldsymbol{\alpha}_\ell = \frac{\gamma}{m} D^{-1} \boldsymbol{\rho}_\ell$
3. $\frac{\partial \mathfrak{L}}{\partial b_\ell} = 0 \Rightarrow \mathbf{e}^{\mathrm{T}} \boldsymbol{\alpha}_\ell$
4. $\frac{\partial \mathfrak{L}}{\partial \boldsymbol{\alpha}_\ell} = 0 \Rightarrow \boldsymbol{\rho}_\ell - \mathbf{F}\mathbf{u}_\ell - b_\ell \mathbf{e} = 0$

Dalsze własności

1. (1) & (4): $Z = SA_{nk} + \boldsymbol{b}\mathbf{e}$, a dla nowych obserwacji $\{\mathbf{y}_1, \ldots, \mathbf{y}_{m_{new}}\}$: $Z^{new} = S^{new} A_{nk} + \boldsymbol{b}\mathbf{e}$, gdzie $S^{new}(\mathbf{y}_i, \mathbf{x}_j) = \mathrm{K}(\mathbf{y}_i, \mathbf{x}_j)$
2. (2): $\mathtt{sign}(\boldsymbol{\rho}_j) = \mathtt{sign}(\boldsymbol{\alpha}_j)$
3. (3): $\mathbf{e}\boldsymbol{\alpha}_1 = 0 \Rightarrow n_k = k - 1$

Komentarz

Problem dualny do problemu

$$\min J(\mathbf{u}_\ell, \boldsymbol{\rho}_\ell, b_\ell) = \frac{1}{2m}\sum_{\ell=1}^{n_k} \gamma_\ell \boldsymbol{\rho}_\ell^{\mathrm{T}} W \boldsymbol{\rho}_\ell - \frac{1}{2}\sum_{\ell=1}^{n_k} \mathbf{u}_\ell^{\mathrm{T}}\mathbf{u}_\ell$$
$$\text{s.t. } \boldsymbol{\rho}_\ell = \mathbf{F}\mathbf{u}_\ell, \quad \ell = 1, \ldots, n_k \geqslant k-1$$

ma postać

$$WS\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha}$$

Gdy $W = D^{-1}$ otrzymujemy algorytm MS, a przyjmując $\boldsymbol{\phi}_\ell = \boldsymbol{\alpha}_\ell$ – algorytm NJW.

Grupowanie spektralne z WKPCA - klasyfikacja obserwacji

Error correcting output codes (Dietterich, Bakiri, 1995):

- Binaryzuj wiersze a^i macierzy A_{n_k}:
  $c_i = \mathrm{sign}(a^i) \in \{-1, 1\}^{k-1}$, $i = 1, \ldots, m$
- Utwórz książkę kodową z k najbardziej popularnych słów.
- *przydziel obserwację x_i do grupy* $g_i = \arg\min_j d_H(sign(\alpha_i), c_j)$
- dla nowej obserwacji y_ℓ wyznacz wektor $z_\ell = S(y_\ell, \cdot)A_{n_k} + b$
- przydziel obserwację y_ℓ do grupy $g_\ell = \arg\min_j d_H(\mathrm{sign}(z_\ell), c_j)$

Grupowanie spektralne z WKPCA - klasyfikacja obserwacji

Error correcting output codes (Dietterich, Bakiri, 1995):

- Binaryzuj wiersze a^i macierzy A_{n_k}:
  $c_i = \mathrm{sign}(a^i) \in \{-1, 1\}^{k-1}$, $i = 1, \ldots, m$
- Utwórz książkę kodową z k najbardziej popularnych słów.
- *przydziel obserwację* x_i *do grupy* $g_i = \arg\min_j d_H(sign(\alpha_i), c_j)$
- dla nowej obserwacji y_ℓ wyznacz wektor $z_\ell = S(y_\ell, \cdot)A_{n_k} + b$
- przydziel obserwację y_ℓ do grupy $g_\ell = \arg\min_j d_H(\mathrm{sign}(z_\ell), c_j)$

Grupowanie spektralne z WKPCA - klasyfikacja obserwacji

Error correcting output codes (Dietterich, Bakiri, 1995):

- Binaryzuj wiersze a^i macierzy A_{n_k}:
  $c_i = \mathrm{sign}(a^i) \in \{-1, 1\}^{k-1}$, $i = 1, \ldots, m$
- Utwórz książkę kodową z k najbardziej popularnych słów.
- *przydziel obserwację x_i do grupy* $g_i = \arg\min_j d_H(sign(\alpha_i), c_j)$
- dla nowej obserwacji y_ℓ wyznacz wektor $z_\ell = S(y_\ell, \cdot)A_{n_k} + b$
- przydziel obserwację y_ℓ do grupy $g_\ell = \arg\min_j d_H(\mathrm{sign}(z_\ell), c_j)$

Grupowanie spektralne z WKPCA - klasyfikacja obserwacji

Error correcting output codes (Dietterich, Bakiri, 1995):

- Binaryzuj wiersze a^i macierzy A_{n_k}:
  $c_i = \mathrm{sign}(a^i) \in \{-1, 1\}^{k-1}$, $i = 1, \ldots, m$
- Utwórz książkę kodową z k najbardziej popularnych słów.
- *przydziel obserwację* x_i *do grupy* $g_i = \arg\min_j d_H(sign(\alpha_i), c_j)$
- dla nowej obserwacji y_ℓ wyznacz wektor $z_\ell = S(y_\ell, \cdot)A_{n_k} + b$
- przydziel obserwację y_ℓ do grupy $g_\ell = \arg\min_j d_H(\mathrm{sign}(z_\ell), c_j)$

Grupowanie spektralne z WKPCA - klasyfikacja obserwacji

Error correcting output codes (Dietterich, Bakiri, 1995):

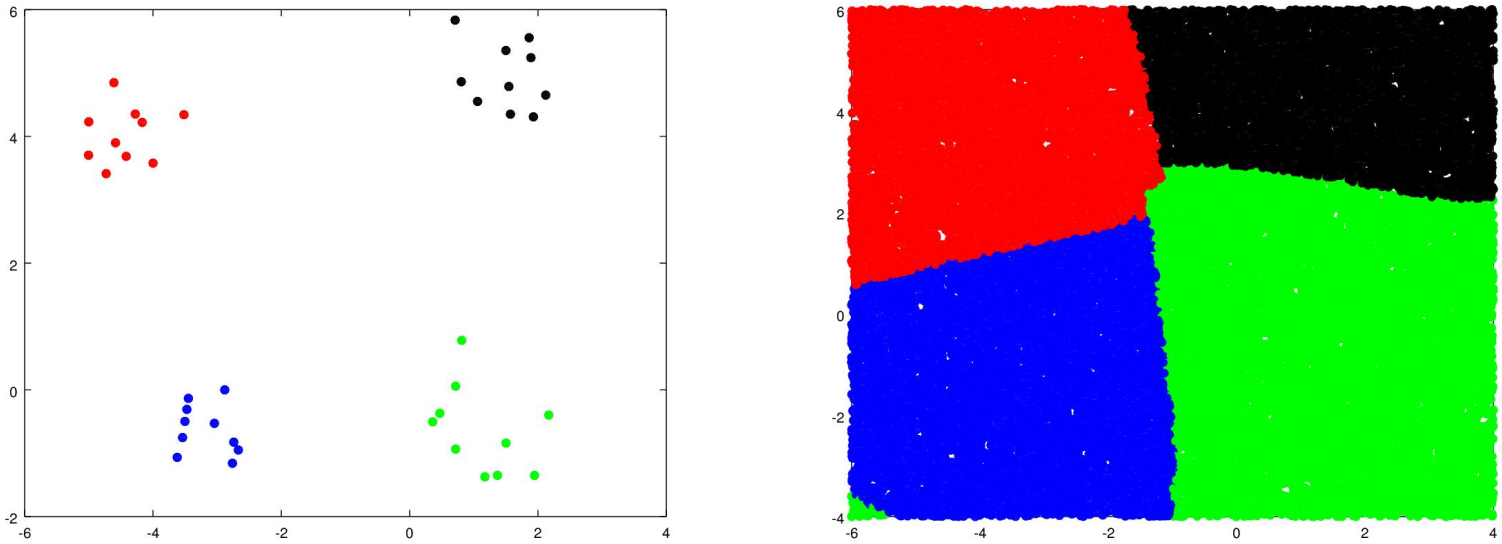
- Binaryzuj wiersze a^i macierzy A_{n_k}:
  $c_i = \mathrm{sign}(a^i) \in \{-1, 1\}^{k-1}$, $i = 1, \ldots, m$
- Utwórz książkę kodową z k najbardziej popularnych słów.
- *przydziel obserwację* x_i *do grupy* $g_i = \arg\min_j d_H(sign(\alpha_i), c_j)$
- dla nowej obserwacji y_ℓ wyznacz wektor $z_\ell = S(y_\ell, \cdot)A_{n_k} + b$
- przydziel obserwację y_ℓ do grupy $g_\ell = \arg\min_j d_H(\mathrm{sign}(z_\ell), c_j)$

Grupowanie spektralne z WKPCA - klasyfikacja obserwacji

Error correcting output codes (Dietterich, Bakiri, 1995):

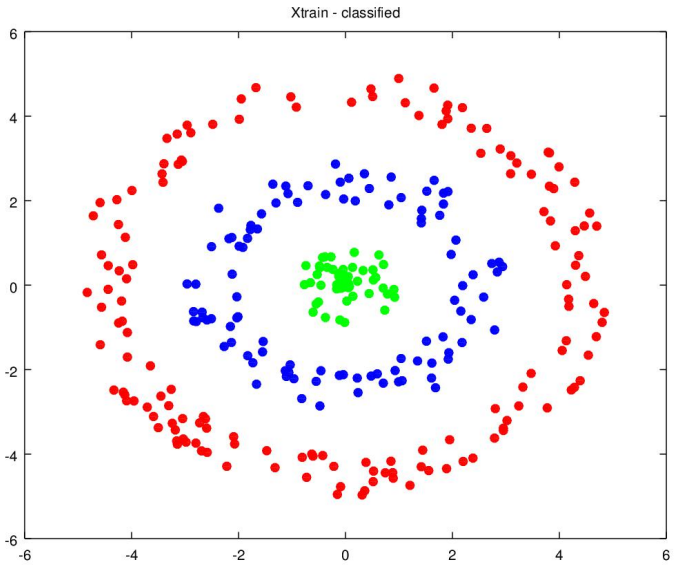
- Binaryzuj wiersze a^i macierzy A_{n_k}:
  $c_i = \mathrm{sign}(a^i) \in \{-1, 1\}^{k-1}$, $i = 1, \ldots, m$
- Utwórz książkę kodową z k najbardziej popularnych słów.
- *przydziel obserwację* x_i *do grupy* $g_i = \arg\min_j d_H(sign(\alpha_i), c_j)$
- dla nowej obserwacji y_ℓ wyznacz wektor $z_\ell = S(y_\ell, \cdot)A_{n_k} + b$
- przydziel obserwację y_ℓ do grupy $g_\ell = \arg\min_j d_H(\mathrm{sign}(z_\ell), c_j)$

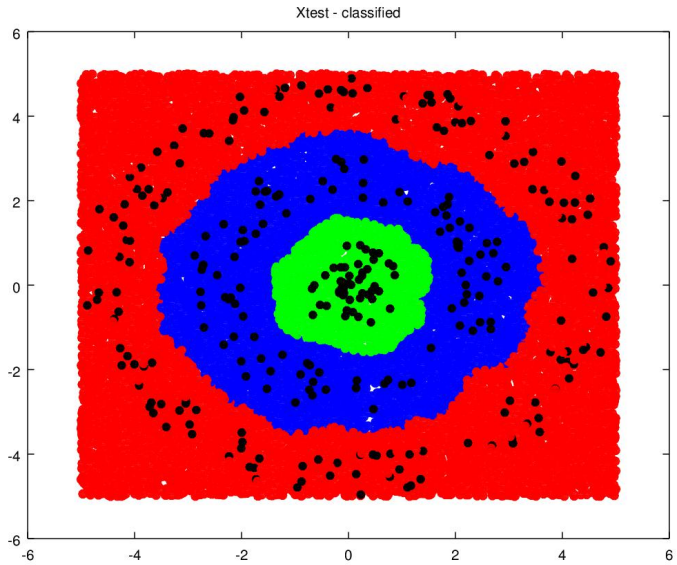
Podział przestrzeni obserwacji (1)



Rysunek: Dane trenujące i ich rozszerzenie

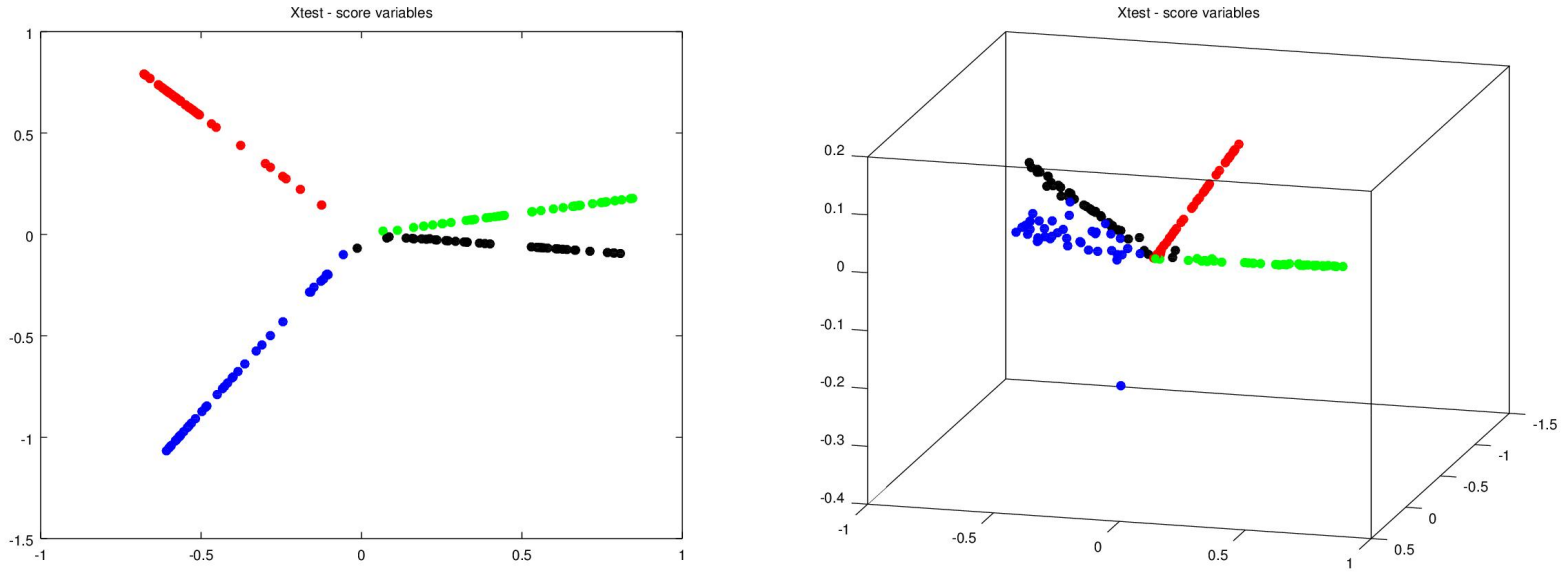
Podział przestrzeni obserwacji (2)





Rysunek: Dane trenujące i ich rozszerzenie

Projekcje danych testowych na dominujące wektory własne



Rysunek: Grupy liniowo separowalne oraz grupy częściowo nakładające się

Grupowanie ewolucyjne

W przypadku danych ewoluujących w czasie, wyniki grupowania powinny odzwierciedlać aktualny stan a jednocześnie nie powinny dramatyczne różnić się od poprzednich wyników. Podejścia:

- (Chakrabarti, Kumar, & Tomkins, 2006)
- (Chi, & *et al.*, 2009)
- (Langone, van Barel, & Suykens, 2016): $\max_{Z^{(t)}} \mathtt{tr}\{(Z^{(t)})^{\mathtt{T}}[\eta\mathcal{S}^{(t)} + (1-\eta)\mathcal{S}^{(t-1)}]Z^{(t)}\}$, p.o. $(Z^{(t)})^{\mathtt{T}}Z^{(t)} = \mathbb{I}^{(t)}$

Nowe podejścia do pokonywania złożoności

1. **RanNLA** – Randomized Numerical Linear Algebra (Drineas, & Mahoney, 2016),(Mahoney, 2016): Randomizacja jako narzędzie wspomagające konstrukcję algorytmów o małym czasie wykonywania i lepszej stabilności. Jednym z głównych chwytów jest *sketching*: zastąpienie dużej macierzy A mniejszą, np. $\widehat{A} = A\Omega$, która reprezentuje istotne własności A.
2. **Sparsity** (wybór podzbioru obserwacji z X): (Alzate, & Suykens, 2011), (Federix, & van Barel, 2010), oraz
3. **Compressed sensing**, np. (Tremblay, Puy, Gribonval, & Vandergheynst, 2016)

Nowe podejścia do pokonywania złożoności

1. **RanNLA** – Randomized Numerical Linear Algebra (Drineas, & Mahoney, 2016),(Mahoney, 2016): Randomizacja jako narzędzie wspomagające konstrukcję algorytmów o małym czasie wykonywania i lepszej stabilności. Jednym z głównych chwytów jest *sketching*: zastąpienie dużej macierzy A mniejszą, np. $\widehat{A} = A\Omega$, która reprezentuje istotne własności A.
2. **Sparsity** (wybór podzbioru obserwacji z X): (Alzate, & Suykens, 2011), (Federix, & van Barel, 2010), oraz
3. **Compressed sensing**, np. (Tremblay, Puy, Gribonval, & Vandergheynst, 2016)

Nowe podejścia do pokonywania złożoności

1. **RanNLA** – Randomized Numerical Linear Algebra (Drineas, & Mahoney, 2016),(Mahoney, 2016): Randomizacja jako narzędzie wspomagające konstrukcję algorytmów o małym czasie wykonywania i lepszej stabilności. Jednym z głównych chwytów jest *sketching*: zastąpienie dużej macierzy A mniejszą, np. $\widehat{A} = A\Omega$, która reprezentuje istotne własności A.
2. **Sparsity** (wybór podzbioru obserwacji z X): (Alzate, & Suykens, 2011), (Federix, & van Barel, 2010), oraz
3. **Compressed sensing**, np. (Tremblay, Puy, Gribonval, & Vandergheynst, 2016)

Nowe podejścia do pokonywania złożoności

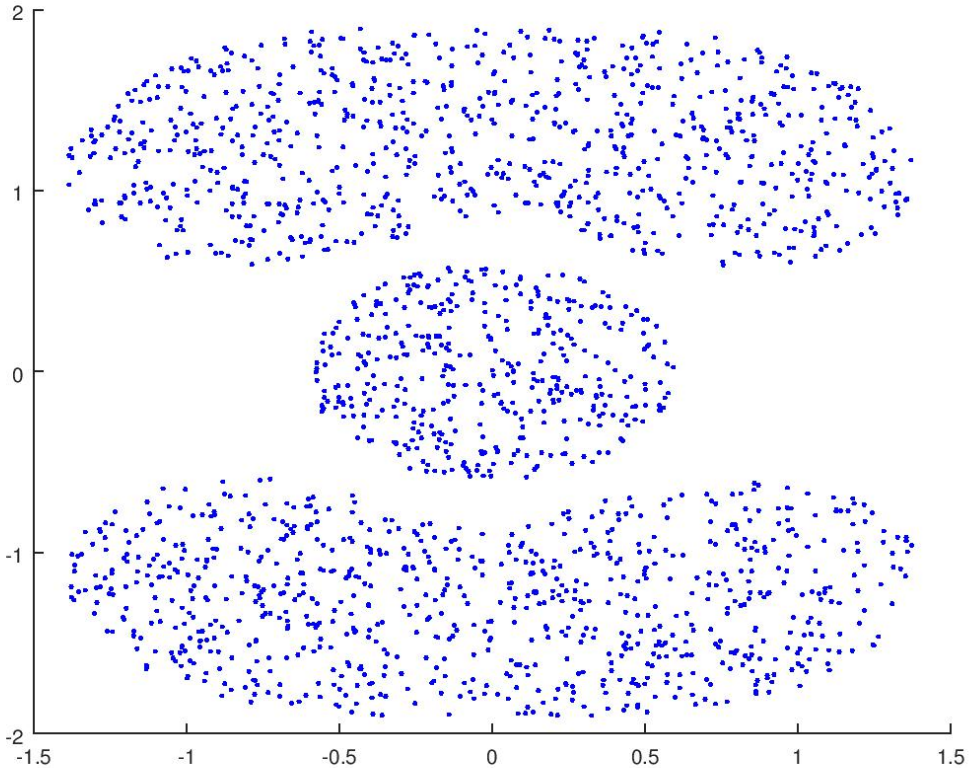
1. **RanNLA** – Randomized Numerical Linear Algebra (Drineas, & Mahoney, 2016),(Mahoney, 2016): Randomizacja jako narzędzie wspomagające konstrukcję algorytmów o małym czasie wykonywania i lepszej stabilności. Jednym z głównych chwytów jest *sketching*: zastąpienie dużej macierzy A mniejszą, np. $\widehat{A} = A\Omega$, która reprezentuje istotne własności A.
2. **Sparsity** (wybór podzbioru obserwacji z X): (Alzate, & Suykens, 2011), (Federix, & van Barel, 2010), oraz
3. **Compressed sensing**, np. (Tremblay, Puy, Gribonval, & Vandergheynst, 2016)

Proste recepty (1)

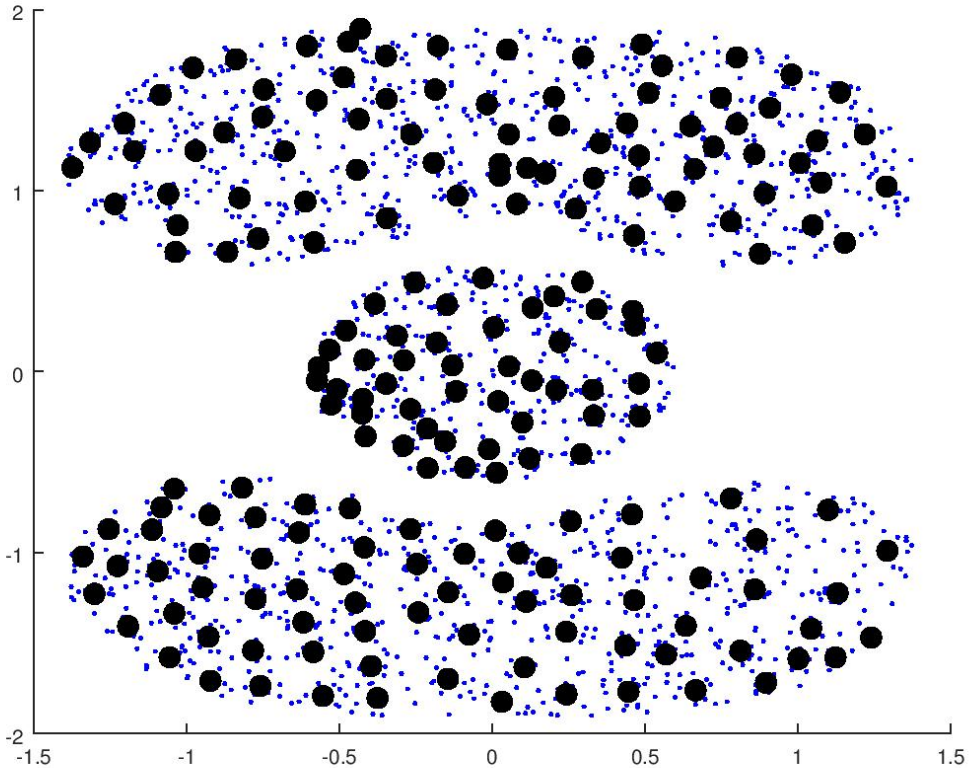
Wybierz $p \ll m$ reprezentantów i wykonaj grupowanie spektralne na macierzy $S_{red} \in \mathbb{R}^{p \times p}$. Przydziel pozostałe punkty do skupień wyznaczonych przez najbliższych reprezentantów. Por.

- Shinnou & Sasaki, 2008
- Yan, Huang & Jordan, 2009

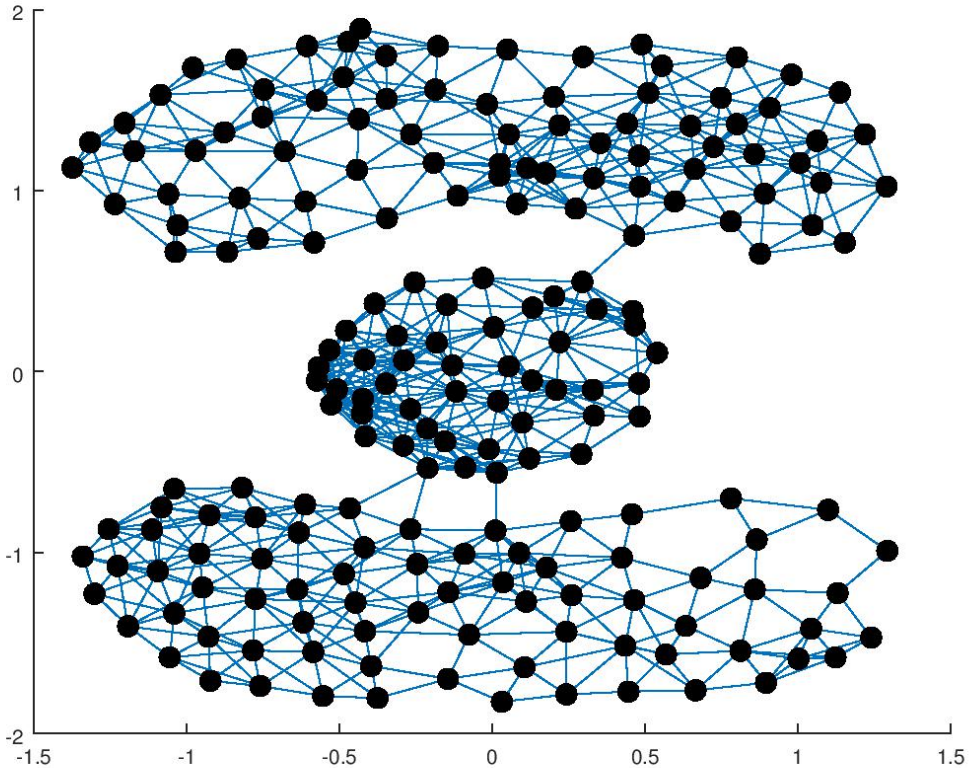
Proste recepty – ilustracja (dane wejściowe (`cassini`))



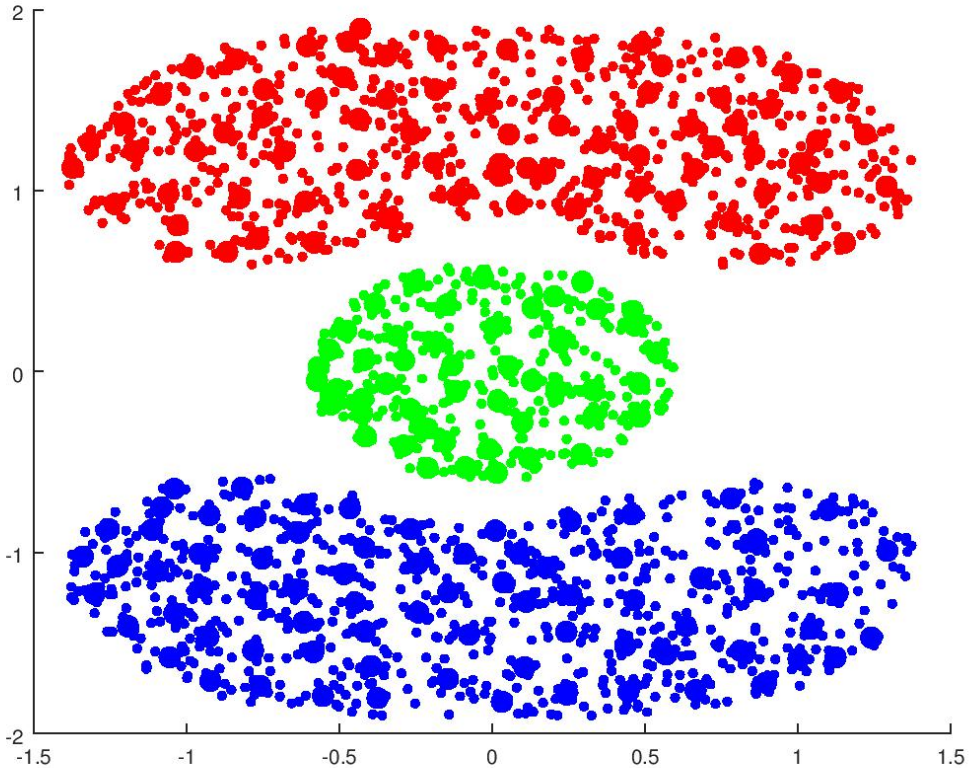
Proste recepty – ilustracja (wybór reprezentantów)



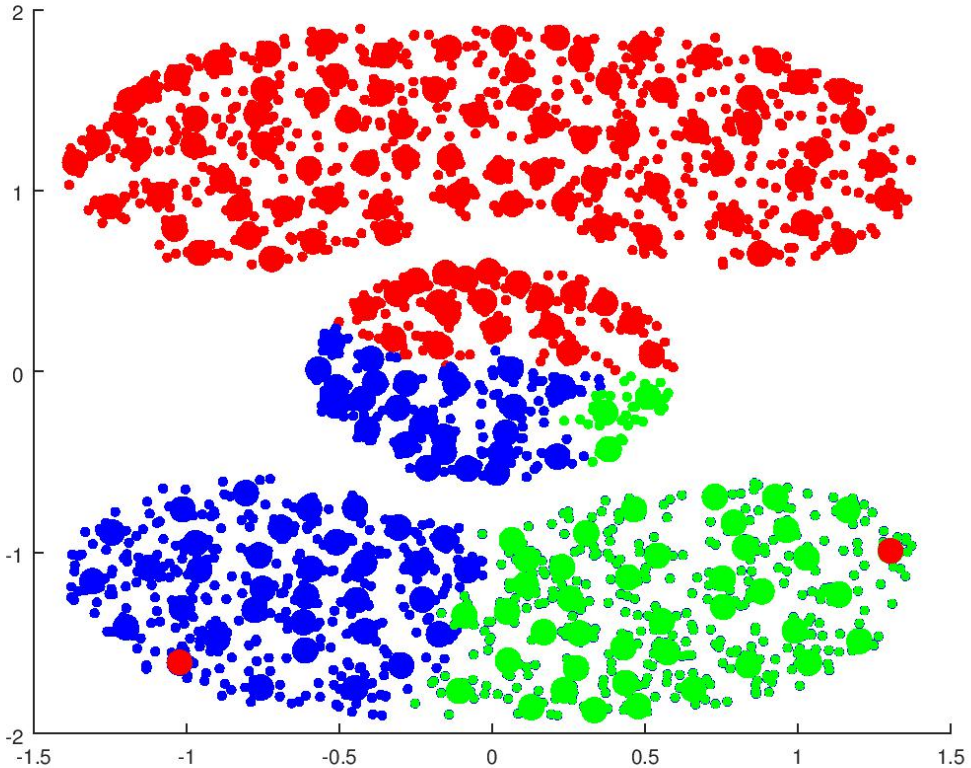
Proste recepty – ilustracja (graf podobieństwa)



Proste recepty – ilustracja (podział końcowy)



Proste recepty – ilustracja (podział z k-means)



Proste recepty (2a) (Chen & Cai, 2011)

- Dane: $\tilde{X} = (x_1, \ldots, x_m) \in \mathbb{R}^{n \times m}$.
- NMF reprezentacja danych: $\tilde{X} \approx WH$, gdzie $W = (w_1, \ldots, w_p) \in \mathbb{R}^{n \times p}$ to zbiór pojęć/składowych, $p < m$, $H = (h1, \ldots, h_m) \in \mathbb{R}^{p \times m}$. Wtedy:

$$x_j \approx \sum_{i=1}^{p} w_i h_{ij} = W h_j$$

- Rzadka reprezentacja: macierz H jest rzadka, np.:

$$h_{ij} = \begin{cases} \frac{K(w_i, x_j)}{\sum_{\ell=1}^{r} K w_\ell, x_j)} & w_i \in N_W(j, r) \\ 0 & wp.p. \end{cases}$$

$N_W(j, r)$ – zbiór $r < p$ pojęć najbliższych x_j, $K(w_i, x_j)$ – stopień podobieństwa w_i do x_j.

Proste recepty (2a) (Chen & Cai, 2011)

- Dane: $\tilde{X} = (x_1, \ldots, x_m) \in \mathbb{R}^{n \times m}$.
- NMF reprezentacja danych: $\tilde{X} \approx WH$, gdzie $W = (w_1, \ldots, w_p) \in \mathbb{R}^{n \times p}$ to zbiór pojęć/składowych, $p < m$, $H = (h1, \ldots, h_m) \in \mathbb{R}^{p \times m}$. Wtedy:

$$x_j \approx \sum_{i=1}^{p} w_i h_{ij} = W h_j$$

- Rzadka reprezentacja: macierz H jest rzadka, np.:

$$h_{ij} = \begin{cases} \frac{K(w_i, x_j)}{\sum_{\ell=1}^{r} K w_\ell, x_j)} & w_i \in N_W(j, r) \\ 0 & wp.p. \end{cases}$$

$N_W(j, r)$ – zbiór $r < p$ pojęć najbliższych x_j, $K(w_i, x_j)$ – stopień podobieństwa w_i do x_j.

Proste recepty (2a) (Chen & Cai, 2011)

- Dane: $\tilde{X} = (x_1, \ldots, x_m) \in \mathbb{R}^{n \times m}$.
- NMF reprezentacja danych: $\tilde{X} \approx WH$, gdzie $W = (w_1, \ldots, w_p) \in \mathbb{R}^{n \times p}$ to zbiór pojęć/składowych, $p < m$, $H = (h1, \ldots, h_m) \in \mathbb{R}^{p \times m}$. Wtedy:

$$x_j \approx \sum_{i=1}^{p} w_i h_{ij} = W h_j$$

- Rzadka reprezentacja: macierz H jest rzadka, np.:

$$h_{ij} = \begin{cases} \frac{K(w_i, x_j)}{\sum_{\ell=1}^{r} K w_\ell, x_j)} & w_i \in N_W(j, r) \\ 0 & wp.p. \end{cases}$$

$N_W(j, r)$ – zbiór $r < p$ pojęć najbliższych x_j, $K(w_i, x_j)$ – stopień podobieństwa w_i do x_j.

Proste recepty (2a) (Chen & Cai, 2011)

- Dane: $\tilde{X} = (x_1, \ldots, x_m) \in \mathbb{R}^{n \times m}$.
- NMF reprezentacja danych: $\tilde{X} \approx WH$, gdzie $W = (w_1, \ldots, w_p) \in \mathbb{R}^{n \times p}$ to zbiór pojęć/składowych, $p < m$, $H = (h1, \ldots, h_m) \in \mathbb{R}^{p \times m}$. Wtedy:

$$x_j \approx \sum_{i=1}^{p} w_i h_{ij} = W h_j$$

- Rzadka reprezentacja: macierz H jest rzadka, np.:

$$h_{ij} = \begin{cases} \frac{K(w_i, x_j)}{\sum_{\ell=1}^{r} K w_\ell, x_j)} & w_i \in N_W(j, r) \\ 0 & wp.p. \end{cases}$$

$N_W(j, r)$ – zbiór $r < p$ pojęć najbliższych x_j, $K(w_i, x_j)$ – stopień podobieństwa w_i do x_j.

Proste recepty (2b) Algorytm LSC (Chen & Cai, 2011)

- Rzadka macierz podobieństwa $S = Z^T Z = H^T C^{-1} H$, gdzie $Z = C^{-1/2} H$ i $C = \mathrm{diag}(c_1, \ldots, c_p)$, $c_i = \sum_j h_{ij}$.
- Skoro $\sum_i h_{ij} = 1$ to $\sum_j s_{ij} = 1$, czyli $D = \mathbb{I} \Rightarrow$ $L = \mathcal{L} = \mathbb{I} - Z^T Z$.
- Jeżeli $Z = A \Sigma B^T$, gdzie $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$, to: $ZZ^T = A \Sigma^2 A^T$, $Z^T Z = B \Sigma^2 B^T$, czyli kolumny A (odp. B) to wektory własne macierzy ZZ^T (odp. $Z^T Z = S$). Macierz $ZZ^T \in \mathbb{R}^{p \times p}$, tzn. czas wyznaczenia A $O(p^3)$. $B^T = \Sigma^{-1} A^T \hat{Z}$.
- Problemy: (a) Jak wybierać kolumny W? Losowo, albo stosując k-means. (b) błąd $\|\tilde{X} - WH\|_F$ może być duży.

Proste recepty (2b) Algorytm LSC (Chen & Cai, 2011)

- Rzadka macierz podobieństwa $S = Z^{\mathrm{T}} Z = H^{\mathrm{T}} C^{-1} H$, gdzie $Z = C^{-1/2} H$ i $C = \mathrm{diag}(c_1, \ldots, c_p)$, $c_i = \sum_j h_{ij}$.
- Skoro $\sum_i h_{ij} = 1$ to $\sum_j s_{ij} = 1$, czyli $D = \mathbb{I} \Rightarrow$ $L = \mathcal{L} = \mathbb{I} - Z^{\mathrm{T}} Z$.
- Jeżeli $Z = A \Sigma B^{\mathrm{T}}$, gdzie $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$, to: $Z Z^{\mathrm{T}} = A \Sigma^2 A^{\mathrm{T}}$, $Z^{\mathrm{T}} Z = B \Sigma^2 B^{\mathrm{T}}$, czyli kolumny A (odp. B) to wektory własne macierzy $Z Z^{\mathrm{T}}$ (odp. $Z^{\mathrm{T}} Z = S$). Macierz $Z Z^{\mathrm{T}} \in \mathbb{R}^{p \times p}$, tzn. czas wyznaczenia A $O(p^3)$. $B^{\mathrm{T}} = \Sigma^{-1} A^{\mathrm{T}} \hat{Z}$.
- Problemy: (a) Jak wybierać kolumny W? Losowo, albo stosując k-means. (b) błąd $\|\tilde{X} - WH\|_F$ może być duży.

Proste recepty (2b) Algorytm LSC (Chen & Cai, 2011)

- Rzadka macierz podobieństwa $S = Z^{\mathrm{T}} Z = H^{\mathrm{T}} C^{-1} H$, gdzie $Z = C^{-1/2} H$ i $C = \mathrm{diag}(c_1, \ldots, c_p)$, $c_i = \sum_j h_{ij}$.
- Skoro $\sum_i h_{ij} = 1$ to $\sum_j s_{ij} = 1$, czyli $D = \mathbb{I} \Rightarrow$ $L = \mathcal{L} = \mathbb{I} - Z^{\mathrm{T}} Z$.
- Jeżeli $Z = A \Sigma B^{\mathrm{T}}$, gdzie $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$, to: $Z Z^{\mathrm{T}} = A \Sigma^2 A^{\mathrm{T}}$, $Z^{\mathrm{T}} Z = B \Sigma^2 B^{\mathrm{T}}$, czyli kolumny A (odp. B) to wektory własne macierzy $Z Z^{\mathrm{T}}$ (odp. $Z^{\mathrm{T}} Z = S$). Macierz $Z Z^{\mathrm{T}} \in \mathbb{R}^{p \times p}$, tzn. czas wyznaczenia A $O(p^3)$. $B^{\mathrm{T}} = \Sigma^{-1} A^{\mathrm{T}} \hat{Z}$.
- Problemy: (a) Jak wybierać kolumny W? Losowo, albo stosując k-means. (b) błąd $\|\tilde{X} - WH\|_F$ może być duży.

Proste recepty (2b) Algorytm LSC (Chen & Cai, 2011)

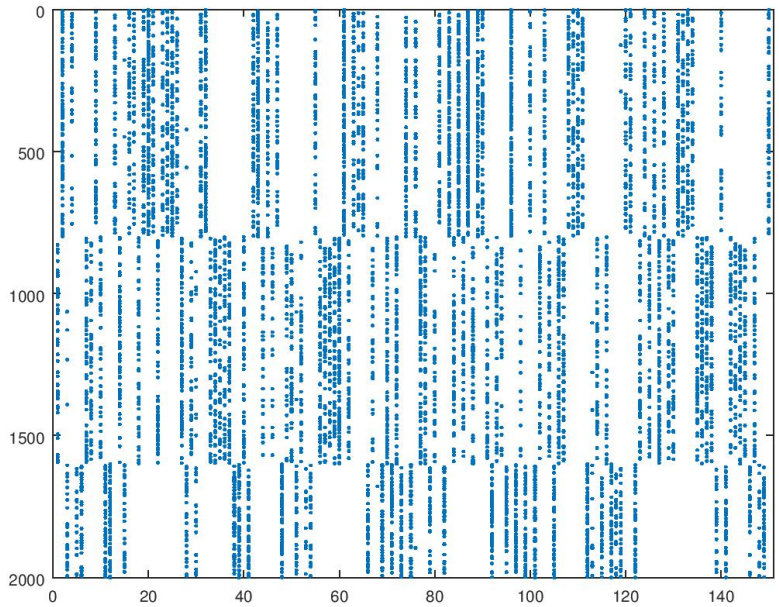
- Rzadka macierz podobieństwa $S = Z^{\mathrm{T}} Z = H^{\mathrm{T}} C^{-1} H$, gdzie $Z = C^{-1/2} H$ i $C = \mathrm{diag}(c_1, \ldots, c_p)$, $c_i = \sum_j h_{ij}$.
- Skoro $\sum_i h_{ij} = 1$ to $\sum_j s_{ij} = 1$, czyli $D = \mathbb{I} \Rightarrow$ $L = \mathcal{L} = \mathbb{I} - Z^{\mathrm{T}} Z$.
- Jeżeli $Z = A \Sigma B^{\mathrm{T}}$, gdzie $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$, to: $Z Z^{\mathrm{T}} = A \Sigma^2 A^{\mathrm{T}}$, $Z^{\mathrm{T}} Z = B \Sigma^2 B^{\mathrm{T}}$, czyli kolumny A (odp. B) to wektory własne macierzy $Z Z^{\mathrm{T}}$ (odp. $Z^{\mathrm{T}} Z = S$). Macierz $Z Z^{\mathrm{T}} \in \mathbb{R}^{p \times p}$, tzn. czas wyznaczenia A $O(p^3)$. $B^{\mathrm{T}} = \Sigma^{-1} A^{\mathrm{T}} \hat{Z}$.
- Problemy: (a) Jak wybierać kolumny W? Losowo, albo stosując k-means. (b) błąd $\|\tilde{X} - WH\|_F$ może być duży.

Proste recepty (2b) Algorytm LSC (Chen & Cai, 2011)

- Rzadka macierz podobieństwa $S = Z^{\mathrm{T}} Z = H^{\mathrm{T}} C^{-1} H$, gdzie $Z = C^{-1/2} H$ i $C = \mathrm{diag}(c_1, \ldots, c_p)$, $c_i = \sum_j h_{ij}$.
- Skoro $\sum_i h_{ij} = 1$ to $\sum_j s_{ij} = 1$, czyli $D = \mathbb{I} \Rightarrow$ $L = \mathcal{L} = \mathbb{I} - Z^{\mathrm{T}} Z$.
- Jeżeli $Z = A \Sigma B^{\mathrm{T}}$, gdzie $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$, to: $Z Z^{\mathrm{T}} = A \Sigma^2 A^{\mathrm{T}}$, $Z^{\mathrm{T}} Z = B \Sigma^2 B^{\mathrm{T}}$, czyli kolumny A (odp. B) to wektory własne macierzy $Z Z^{\mathrm{T}}$ (odp. $Z^{\mathrm{T}} Z = S$). Macierz $Z Z^{\mathrm{T}} \in \mathbb{R}^{p \times p}$, tzn. czas wyznaczenia A $O(p^3)$. $B^{\mathrm{T}} = \Sigma^{-1} A^{\mathrm{T}} \hat{Z}$.
- Problemy: (a) Jak wybierać kolumny W? Losowo, albo stosując k-means. (b) błąd $\|\tilde{X} - WH\|_F$ może być duży.

Ilustracja algorytmu LSC

Dla zbioru `cassini` przyjęto:
$m = 2000$, $p = 150$, $r = 4$.
Czas obliczeń: 0.728232 s
Dokładność: 100%
Algorytm NJW wymaga $> 1,4$ s.



Rysunek: Macierz H^{T}

Randomizowany SVD (Halko *et al.*, 2011)

Etapy randomizowanego SVD:

- znajdź ortogonalną macierz $Q \in \mathbb{R}^{m \times r}$ taką, że $\|(\mathbb{I} - QQ^{\mathrm{T}})A\| \leqslant \epsilon$:
  - losowa macierz gaussowska $\Omega \in \mathbb{R}^{n \times r}$, $r \leqslant \min(m, n)$
  - $Y = (AA^{\mathrm{T}})^q A\Omega$, $q \in \{1, 2\}$
  - $[Q, R] = QR(Y)$
- właściwy SVD
  - wyznacz macierz $B = Q^{\mathrm{T}}A \in \mathbb{R}^{r \times n}$
  - znajdź rozkład USV macierzy B
  - $A \approx (QU)SV$

Złożoność dla macierzy wymiaru $m \times m$: $O((q + r)rm)$.
Dokładność – por. s. 10 w (Halko *et al.*, 2011). Przykładowy kod.

Randomizowany SVD (Halko *et al.*, 2011)

Etapy randomizowanego SVD:

- znajdź ortogonalną macierz $Q \in \mathbb{R}^{m \times r}$ taką, że $\|(\mathbb{I} - QQ^{\mathrm{T}})A\| \leqslant \epsilon$:
  - losowa macierz gaussowska $\Omega \in \mathbb{R}^{m \times r}$, $r < \min(m, n)$
  - $Y = (AA^{\mathrm{T}})^q A\Omega$, $q \in \{1, 2\}$
  - $[Q, R] = QR(Y)$
- właściwy SVD
  - wyznacz macierz $B = Q^{\mathrm{T}} A \in \mathbb{R}^{r \times m}$
  - znajdź rozkład $U\Sigma V^{\mathrm{T}}$ macierzy B
  - $A \approx (QU)\Sigma V^{\mathrm{T}}$

Złożoność dla macierzy wymiaru $m \times m$: $O((q + r)rm)$.
Dokładność – por. s. 10 w (Halko *et al.*, 2011). Przykładowy kod.

Randomizowany SVD (Halko *et al.*, 2011)

Etapy randomizowanego SVD:

- znajdź ortogonalną macierz $Q \in \mathbb{R}^{m \times r}$ taką, że $\|(\mathbb{I} - QQ^{\mathrm{T}})A\| \leqslant \epsilon$:
  - losowa macierz gaussowska $\Omega \in \mathbb{R}^{m \times r}$, $r < \min(m, n)$
  - $Y = (AA^{\mathrm{T}})^q A\Omega$, $q \in \{1, 2\}$
  - $[Q, R] = QR(Y)$
- właściwy SVD
  - wyznacz macierz $B = Q^{\mathrm{T}} A \in \mathbb{R}^{r \times m}$
  - znajdź rozkład $U\Sigma V^{\mathrm{T}}$ macierzy B
  - $A \approx (QU)\Sigma V^{\mathrm{T}}$

Złożoność dla macierzy wymiaru $m \times m$: $O((q + r)rm)$.
Dokładność – por. s. 10 w (Halko *et al.*, 2011). Przykładowy kod.

Randomizowany SVD (Halko *et al.*, 2011)

Etapy randomizowanego SVD:

- znajdź ortogonalną macierz $Q \in \mathbb{R}^{m \times r}$ taką, że $\|(\mathbb{I} - QQ^{\mathrm{T}})A\| \leqslant \epsilon$:
  - losowa macierz gaussowska $\Omega \in \mathbb{R}^{m \times r}$, $r < \min(m, n)$
  - $Y = (AA^{\mathrm{T}})^q A\Omega$, $q \in \{1, 2\}$
  - $[Q, R] = QR(Y)$
- właściwy SVD
  - wyznacz macierz $B = Q^{\mathrm{T}} A \in \mathbb{R}^{r \times m}$
  - znajdź rozkład $U\Sigma V^{\mathrm{T}}$ macierzy B
  - $A \approx (QU)\Sigma V^{\mathrm{T}}$

Złożoność dla macierzy wymiaru $m \times m$: $O((q + r)rm)$.
Dokładność – por. s. 10 w (Halko *et al.*, 2011). Przykładowy kod.

Randomizowany SVD (Halko *et al.*, 2011)

Etapy randomizowanego SVD:

- znajdź ortogonalną macierz $Q \in \mathbb{R}^{m \times r}$ taką, że $\|(\mathbb{I} - QQ^{\mathrm{T}})A\| \leqslant \epsilon$:
  - losowa macierz gaussowska $\Omega \in \mathbb{R}^{m \times r}$, $r < \min(m, n)$
  - $Y = (AA^{\mathrm{T}})^q A\Omega$, $q \in \{1, 2\}$
  - $[Q, R] = QR(Y)$
- właściwy SVD
  - wyznacz macierz $B = Q^{\mathrm{T}} A \in \mathbb{R}^{r \times m}$
  - znajdź rozkład $U\Sigma V^{\mathrm{T}}$ macierzy B
  - $A \approx (QU)\Sigma V^{\mathrm{T}}$

Złożoność dla macierzy wymiaru $m \times m$: $O((q + r)rm)$.
Dokładność – por. s. 10 w (Halko *et al.*, 2011). Przykładowy kod.

Randomizowany SVD (Halko *et al.*, 2011)

Etapy randomizowanego SVD:

- znajdź ortogonalną macierz $Q \in \mathbb{R}^{m \times r}$ taką, że $\|(\mathbb{I} - QQ^{\mathrm{T}})A\| \leqslant \epsilon$:
  - losowa macierz gaussowska $\Omega \in \mathbb{R}^{m \times r}$, $r < \min(m, n)$
  - $Y = (AA^{\mathrm{T}})^q A\Omega$, $q \in \{1, 2\}$
  - $[Q, R] = QR(Y)$
- właściwy SVD
  - wyznacz macierz $B = Q^{\mathrm{T}}A \in \mathbb{R}^{r \times m}$
  - znajdź rozkład $U\Sigma V^{\mathrm{T}}$ macierzy B
  - $A \approx (QU)\Sigma V^{\mathrm{T}}$

Złożoność dla macierzy wymiaru $m \times m$: $O((q + r)rm)$.
Dokładność – por. s. 10 w (Halko *et al.*, 2011). Przykładowy kod.

Randomizowany SVD (Halko *et al.*, 2011)

Etapy randomizowanego SVD:

- znajdź ortogonalną macierz $Q \in \mathbb{R}^{m \times r}$ taką, że $\|(\mathbb{I} - QQ^{\mathrm{T}})A\| \leqslant \epsilon$:
  - losowa macierz gaussowska $\Omega \in \mathbb{R}^{m \times r}$, $r < \min(m, n)$
  - $Y = (AA^{\mathrm{T}})^q A\Omega$, $q \in \{1, 2\}$
  - $[Q, R] = QR(Y)$
- właściwy SVD
  - wyznacz macierz $B = Q^{\mathrm{T}}A \in \mathbb{R}^{r \times m}$
  - znajdź rozkład $U\Sigma V^{\mathrm{T}}$ macierzy B
  - $A \approx (QU)\Sigma V^{\mathrm{T}}$

Złożoność dla macierzy wymiaru $m \times m$: $O((q + r)rm)$.
Dokładność – por. s. 10 w (Halko *et al.*, 2011). Przykładowy kod.

Randomizowany SVD (Halko *et al.*, 2011)

Etapy randomizowanego SVD:

- znajdź ortogonalną macierz $Q \in \mathbb{R}^{m \times r}$ taką, że $\|(\mathbb{I} - QQ^{\mathrm{T}})A\| \leqslant \epsilon$:
  - losowa macierz gaussowska $\Omega \in \mathbb{R}^{m \times r}$, $r < \min(m, n)$
  - $Y = (AA^{\mathrm{T}})^q A\Omega$, $q \in \{1, 2\}$
  - $[Q, R] = QR(Y)$
- właściwy SVD
  - wyznacz macierz $B = Q^{\mathrm{T}}A \in \mathbb{R}^{r \times m}$
  - znajdź rozkład $U\Sigma V^{\mathrm{T}}$ macierzy B
  - $A \approx (QU)\Sigma V^{\mathrm{T}}$

Złożoność dla macierzy wymiaru $m \times m$: $O((q + r)rm)$.
Dokładność – por. s. 10 w (Halko *et al.*, 2011). Przykładowy kod.

Randomizowany SVD (Halko *et al.*, 2011)

Etapy randomizowanego SVD:

- znajdź ortogonalną macierz $Q \in \mathbb{R}^{m \times r}$ taką, że $\|(\mathbb{I} - QQ^{\mathrm{T}})A\| \leqslant \epsilon$:
  - losowa macierz gaussowska $\Omega \in \mathbb{R}^{m \times r}$, $r < \min(m, n)$
  - $Y = (AA^{\mathrm{T}})^q A\Omega$, $q \in \{1, 2\}$
  - $[Q, R] = QR(Y)$
- właściwy SVD
  - wyznacz macierz $B = Q^{\mathrm{T}}A \in \mathbb{R}^{r \times m}$
  - znajdź rozkład $U\Sigma V^{\mathrm{T}}$ macierzy B
  - $A \approx (QU)\Sigma V^{\mathrm{T}}$

Złożoność dla macierzy wymiaru $m \times m$: $O((q + r)rm)$.
Dokładność – por. s. 10 w (Halko *et al.*, 2011). Przykładowy kod.

Randomizowany SVD, cd. (Gittens *et al.*, 2014)

Jeżeli $\|U - \tilde{U}\|_2 \leqslant \epsilon$, to grupowania wyznaczone z U i $\tilde{U}$ są identyczne. Krok (2) w algorytmie NJW zastąpiono przez:

- Niech $\tilde{Y} \in \mathbb{R}^{m \times k}$ zawiera k dominujących wektorów osobliwych macierzy $B = (SS^{\mathrm{T}})^q S\Omega$, gdzie $\Omega \in \mathbb{R}^{m \times k}$ jest macierzą gaussowską
- Zastosuj algorytm k średnich do wierszy macierzy $\tilde{Y}$.

Złożoność: $O((m^2kq + k^2m)$.

Randomizowany SVD, cd. (Gittens *et al.*, 2014)

Jeżeli $\|U - \tilde{U}\|_2 \leqslant \epsilon$, to grupowania wyznaczone z U i $\tilde{U}$ są identyczne. Krok (2) w algorytmie NJW zastąpiono przez:

- Niech $\tilde{Y} \in \mathbb{R}^{m \times k}$ zawiera k dominujących wektorów osobliwych macierzy $B = (\mathcal{S}\mathcal{S}^{\mathrm{T}})^q \mathcal{S}\Omega$, gdzie $\Omega \in \mathbb{R}^{m \times k}$ jest macierzą gaussowską
- Zastosuj algorytm k średnich do wierszy macierzy $\tilde{Y}$.

Złożoność: $O((m^2 kq + k^2 m)$.

Randomizowany SVD, cd. (Gittens *et al.*, 2014)

Jeżeli $\|U - \tilde{U}\|_2 \leqslant \epsilon$, to grupowania wyznaczone z U i $\tilde{U}$ są identyczne. Krok (2) w algorytmie NJW zastąpiono przez:

- Niech $\tilde{Y} \in \mathbb{R}^{m \times k}$ zawiera k dominujących wektorów osobliwych macierzy $B = (\mathcal{S}\mathcal{S}^{\mathrm{T}})^q \mathcal{S}\Omega$, gdzie $\Omega \in \mathbb{R}^{m \times k}$ jest macierzą gaussowską
- Zastosuj algorytm k średnich do wierszy macierzy $\tilde{Y}$.

Złożoność: $O((m^2kq + k^2m)$.

Niepełny rozkład Choleskiego (1)

1. Rozkład Choleskiego-Banachiewicza: Jeżeli $K \succ 0$, to $K = GG^{\mathrm{T}}$, gdzie $G \in \mathbb{R}^{m \times m}$ - macierz trójkątna dolna
2. Niepełny rozkład Choleskiego: Dla $K \succ 0$ szukamy trójkątnej macierzy $C \in \mathbb{R}^{m \times r}$ takiej, że $\|K - CC^{\mathrm{T}}\|_F \leqslant \eta$, przy czym $r \ll m$ (Bach & Jordan, 2002).

   Zalety:
   (a) złożoność czasowa $O(mr^2)$,
   (b) wystarcza znajomość diagonali K (gdy K jest gaussowskie, $\mathrm{diag}(K) = \mathbf{e}$),
   (c) jeżeli wartości własne K szybko zanikają, dekompozycja skutkuje niską wartością η oraz $r \ll m$ (Alzate, & Suykens, 2011) (Frederix, & van Barel, 2013).

Niepełny rozkład Choleskiego (1)

1. Rozkład Choleskiego-Banachiewicza: Jeżeli $K \succ 0$, to $K = GG^{\mathrm{T}}$, gdzie $G \in \mathbb{R}^{m \times m}$ - macierz trójkątna dolna
2. Niepełny rozkład Choleskiego: Dla $K \succ 0$ szukamy trójkątnej macierzy $C \in \mathbb{R}^{m \times r}$ takiej, że $\|K - CC^{\mathrm{T}}\|_F \leqslant \eta$, przy czym $r \ll m$ (Bach & Jordan, 2002).
   Zalety:
   (a) złożoność czasowa $O(mr^2)$,
   (b) wystarcza znajomość diagonali K (gdy K jest gaussowskie, $\mathrm{diag}(K) = \mathbf{e}$),
   (c) jeżeli wartości własne K szybko zanikają, dekompozycja skutkuje niską wartością η oraz $r \ll m$ (Alzate, & Suykens, 2011) (Frederix, & van Barel, 2013).

Niepełny rozkład Choleskiego (1)

1. Rozkład Choleskiego-Banachiewicza: Jeżeli $K \succ 0$, to $K = GG^{\mathrm{T}}$, gdzie $G \in \mathbb{R}^{m \times m}$ - macierz trójkątna dolna
2. Niepełny rozkład Choleskiego: Dla $K \succ 0$ szukamy trójkątnej macierzy $C \in \mathbb{R}^{m \times r}$ takiej, że $\|K - CC^{\mathrm{T}}\|_F \leqslant \eta$, przy czym $r \ll m$ (Bach & Jordan, 2002).
   Zalety:
   (a) złożoność czasowa $O(mr^2)$,
   (b) wystarcza znajomość diagonali K (gdy K jest gaussowskie, $\mathrm{diag}(K) = \mathbf{e}$),
   (c) jeżeli wartości własne K szybko zanikają, dekompozycja skutkuje niską wartością η oraz $r \ll m$ (Alzate, & Suykens, 2011) (Frederix, & van Barel, 2013).

Niepełny rozkład Choleskiego (3)

Aby wyznaczyć dominujące wektory własne macierzy $\mathcal{S} = D^{-1/2} S D^{-1/2}$, $S \succ 0$,:

1. $S = CC^{\mathrm{T}}$, % C = chol_gauss(X,sigma,tol)
2. $D \approx diag(CC^{\mathrm{T}})\mathbf{e} = C(C^{\mathrm{T}}\mathbf{e}) = \mathrm{C} * \mathrm{sum}(\mathrm{C}', 2)$
3. $D^{-1/2}C = QR$, % $[Q, R] = qr(D^{-1/2}C, 0)$
   $Q \in \mathbb{R}^{m \times r}, R \in \mathbb{R}^{r \times r}$
4. $S = QRR^{\mathrm{T}}Q^{\mathrm{T}}$
5. $RR^{\mathrm{T}} = U\Lambda U'$, % $[U, \Lambda] = \mathrm{eig}(RR^{\mathrm{T}})$, $RR^{\mathrm{T}} \in \mathbb{R}^{r \times r}$
6. $S = D^{-1/2} S D^{-1/2} = (QU)\Lambda(QU)^{\mathrm{T}}$,
   tzn. kolumny macierzy QU są wektorami własnymi $\mathcal{S}$.

Niepełny rozkład Choleskiego (3)

Aby wyznaczyć dominujące wektory własne macierzy $\mathcal{S} = D^{-1/2} S D^{-1/2}$, $S \succ 0$,:

1. $S = CC^{\mathrm{T}}$, $\quad \%\ \mathrm{C} = \mathrm{chol\_gauss(X,sigma,tol)}$
2. $D \approx diag(CC^{\mathrm{T}})\mathbf{e} = C(C^{\mathrm{T}}\mathbf{e}) = \mathrm{C} * \mathrm{sum}(\mathrm{C}', 2)$
3. $D^{-1/2} C = QR$, $\quad \%\ [Q, R] = qr(D^{-1/2} C, 0)$
   $Q \in \mathbb{R}^{m \times r}, R \in \mathbb{R}^{r \times r}$
4. $\mathcal{S} = QRR^{\mathrm{T}} Q^{\mathrm{T}}$
5. $RR^{\mathrm{T}} = U \Lambda U'$, $\quad \%\ [U, \Lambda] = \mathrm{eig}(RR^{\mathrm{T}})$, $RR^{\mathrm{T}} \in \mathbb{R}^{r \times r}$
6. $S = D^{-1/2} S D^{-1/2} = (QU)\Lambda(QU)^{\mathrm{T}}$,
   tzn. kolumny macierzy QU są wektorami własnymi $\mathcal{S}$.

Niepełny rozkład Choleskiego (3)

Aby wyznaczyć dominujące wektory własne macierzy $\mathcal{S} = D^{-1/2} S D^{-1/2}$, $S \succ 0$,:

1. $S = CC^{\mathrm{T}}$, $\quad \%\ \mathrm{C} = \mathrm{chol\_gauss(X,sigma,tol)}$
2. $D \approx diag(CC^{\mathrm{T}})\mathbf{e} = C(C^{\mathrm{T}}\mathbf{e}) = \mathrm{C} * \mathrm{sum}(\mathrm{C}', 2)$
3. $D^{-1/2} C = QR$, $\quad \%\ [Q, R] = qr(D^{-1/2} C, 0)$
   $Q \in \mathbb{R}^{m \times r}, R \in \mathbb{R}^{r \times r}$
4. $\mathcal{S} = QRR^{\mathrm{T}} Q^{\mathrm{T}}$
5. $RR^{\mathrm{T}} = U \Lambda U'$, $\quad \%\ [U, \Lambda] = \mathrm{eig}(RR^{\mathrm{T}})$, $RR^{\mathrm{T}} \in \mathbb{R}^{r \times r}$
6. $S = D^{-1/2} S D^{-1/2} = (QU) \Lambda (QU)^{\mathrm{T}}$,
   tzn. kolumny macierzy QU są wektorami własnymi $\mathcal{S}$.

Niepełny rozkład Choleskiego (3)

Aby wyznaczyć dominujące wektory własne macierzy $\mathcal{S} = D^{-1/2} S D^{-1/2}$, $S \succ 0$,:

1. $S = CC^{\mathrm{T}}$, $\quad \%\ \mathrm{C} = \mathrm{chol\_gauss(X,sigma,tol)}$
2. $D \approx diag(CC^{\mathrm{T}})\mathbf{e} = C(C^{\mathrm{T}}\mathbf{e}) = \mathrm{C} * \mathrm{sum}(\mathrm{C}', 2)$
3. $D^{-1/2}C = QR$, $\quad \%\ [Q, R] = qr(D^{-1/2}C, 0)$
   $Q \in \mathbb{R}^{m \times r}, R \in \mathbb{R}^{r \times r}$
4. $\mathcal{S} = QRR^{\mathrm{T}}Q^{\mathrm{T}}$
5. $RR^{\mathrm{T}} = U \Lambda U'$, $\quad \%\ [U, \Lambda] = \mathrm{eig}(RR^{\mathrm{T}})$, $RR^{\mathrm{T}} \in \mathbb{R}^{r \times r}$
6. $S = D^{-1/2} S D^{-1/2} = (QU)\Lambda(QU)^{\mathrm{T}}$,
   tzn. kolumny macierzy QU są wektorami własnymi $\mathcal{S}$.

Niepełny rozkład Choleskiego (3)

Aby wyznaczyć dominujące wektory własne macierzy $\mathcal{S} = D^{-1/2} S D^{-1/2}$, $S \succ 0$,:

1. $S = CC^{\mathrm{T}}$, $\quad \%\ \mathrm{C} = \mathrm{chol\_gauss(X,sigma,tol)}$
2. $D \approx diag(CC^{\mathrm{T}})\mathbf{e} = C(C^{\mathrm{T}}\mathbf{e}) = \mathrm{C} * \mathrm{sum}(\mathrm{C}', 2)$
3. $D^{-1/2}C = QR$, $\quad \%\ [Q, R] = qr(D^{-1/2}C, 0)$
   $Q \in \mathbb{R}^{m \times r}, R \in \mathbb{R}^{r \times r}$
4. $\mathcal{S} = QRR^{\mathrm{T}}Q^{\mathrm{T}}$
5. $RR^{\mathrm{T}} = U \Lambda U'$, $\quad \%\ [U, \Lambda] = \mathrm{eig}(RR^{\mathrm{T}})$, $RR^{\mathrm{T}} \in \mathbb{R}^{r \times r}$
6. $\mathcal{S} = D^{-1/2} S D^{-1/2} = (QU)\Lambda(QU)^{\mathrm{T}}$,
   tzn. kolumny macierzy QU są wektorami własnymi $\mathcal{S}$.

Niepełny rozkład Choleskiego (3)

Aby wyznaczyć dominujące wektory własne macierzy $\mathcal{S} = D^{-1/2} S D^{-1/2}$, $S \succ 0$,:

1. $S = CC^{\mathrm{T}}$, $\quad \%\ \mathrm{C} = \mathrm{chol\_gauss(X,sigma,tol)}$
2. $D \approx diag(CC^{\mathrm{T}})\mathbf{e} = C(C^{\mathrm{T}}\mathbf{e}) = \mathrm{C} * \mathrm{sum}(\mathrm{C}', 2)$
3. $D^{-1/2} C = QR$, $\quad \%\ [Q, R] = qr(D^{-1/2} C, 0)$
   $Q \in \mathbb{R}^{m \times r}, R \in \mathbb{R}^{r \times r}$
4. $\mathcal{S} = QRR^{\mathrm{T}} Q^{\mathrm{T}}$
5. $RR^{\mathrm{T}} = U \Lambda U'$, $\quad \%\ [U, \Lambda] = \mathrm{eig}(RR^{\mathrm{T}})$, $RR^{\mathrm{T}} \in \mathbb{R}^{r \times r}$
6. $S = D^{-1/2} S D^{-1/2} = (QU)\Lambda(QU)^{\mathrm{T}}$,
   tzn. kolumny macierzy QU są wektorami własnymi $\mathcal{S}$.

Niepełny rozkład Choleskiego (3)

Aby wyznaczyć dominujące wektory własne macierzy $\mathcal{S} = D^{-1/2} S D^{-1/2}$, $S \succ 0$,:

1. $S = CC^{\mathrm{T}}$, $\quad \%\ \mathrm{C} = \mathrm{chol\_gauss(X,sigma,tol)}$
2. $D \approx diag(CC^{\mathrm{T}})\mathbf{e} = C(C^{\mathrm{T}}\mathbf{e}) = \mathrm{C} * \mathrm{sum}(\mathrm{C}', 2)$
3. $D^{-1/2}C = QR$, $\quad \%\ [Q, R] = qr(D^{-1/2}C, 0)$
   $Q \in \mathbb{R}^{m \times r}, R \in \mathbb{R}^{r \times r}$
4. $\mathcal{S} = QRR^{\mathrm{T}}Q^{\mathrm{T}}$
5. $RR^{\mathrm{T}} = U\Lambda U'$, $\quad \%\ [U, \Lambda] = \mathrm{eig}(RR^{\mathrm{T}})$, $RR^{\mathrm{T}} \in \mathbb{R}^{r \times r}$
6. $\mathcal{S} = D^{-1/2} S D^{-1/2} = (QU)\Lambda(QU)^{\mathrm{T}}$,
   tzn. kolumny macierzy QU są wektorami własnymi $\mathcal{S}$.

Rozszerzenie Nyströma (Fowlkes, *et al.*, 2004)

Zarys oryginalnej metody (1): Aproksymacją rozwiązania problemu własnego

$$\int_1^0 K(x,y)\phi(y)dy = \lambda\phi(x)$$

jest

$$\frac{1}{m}\sum_{j=1}^{m} K(x,\xi_j)\hat{\phi}(\xi_j) = \lambda\hat{\phi}(x) \qquad (*)$$

Rozszerzenie Nyströma (Fowlkes, *et al.*, 2004)

Zarys oryginalnej metody (2): Aby wyznaczyć $\hat{\phi}$ wybierz się m punktów $\{x_1, \ldots, x_m\} \subset [0,1]$ i rozwiązuje układ m równań postaci

$$\frac{1}{m}\sum_{j=1}^{m} \mathrm{K}(\xi_i, \xi_j)\hat{\phi}(\xi_j) = \lambda\hat{\phi}(\xi_i), \quad i = 1, \ldots, m \qquad (**)$$

lub równoważnie: $K\hat{\Phi} = m\hat{\Phi}\Lambda^{-1}$, $K = [k_{ij}]$, $k_{ij} = \mathrm{K}(\xi_i, \xi_j)$, $\hat{\Phi} = (\hat{\phi}_1, \ldots, \hat{\phi}_m)$ wektory własne macierzy A odpowiadające wartościom własnym $\lambda_1, \ldots, \lambda_m$. Ostatecznie

$$\hat{\phi}_i(x) = \frac{1}{m\lambda_i}\sum_{j=1}^{m} K(x, \xi_j)\hat{\phi}_i(\xi_j), \quad x \in [0,1]$$

Rozszerzenie Nyströma (Fowlkes, *et al.*, 2004)

Przedstawiamy macierz podobieństw w postaci:

$$S = \begin{bmatrix} A & B \\ B^{\mathrm{T}} & C \end{bmatrix}, \quad A \in \mathbb{R}^{r\times r}, B \in \mathbb{R}^{r\times m-r}, C \in \mathbb{R}^{m-r\times m-r},$$

gdzie $r \ll m$, i niech $S = Y\Lambda Y^{\mathrm{T}}$, $A = U\Sigma U^{\mathrm{T}}$, $U = (u_1, \ldots, u_r)$. Rozszerzenie Nyströma ma postać

$$\tilde{u}_{ji} = \frac{1}{\lambda_i}\mathbf{b}_j^{\mathrm{T}} u_i, \quad i = 1, \ldots, p \leqslant r, j = r{+}1, \ldots, m$$

gdzie $\mathbf{b}_j^{\mathrm{T}}$ – j-ty wiersz macierzy B^{T} (odpowiednik $\mathrm{K}(x, \cdot)$). Inaczej $\tilde{u}_i = \frac{1}{\lambda_i} B^{\mathrm{T}} u_i$, $i = 1, \ldots, p \leqslant r$ (Williams & Seeger, 2000). Ogólnie:

$$\tilde{U} = (\tilde{u}_1, \ldots, \tilde{u}_p) = B^{\mathrm{T}} U\Lambda^{-1}, \tilde{U} \in \mathbb{R}^{(m-r)\times p}$$

Rozszerzenie Nyströma (Fowlkes, *et al.*, 2004)

$\hat{U} = (U; \tilde{U}) = (U; B^{\mathrm{T}} U \Lambda^{-1})$. $\|S - \hat{U}\Sigma\hat{U}^{\mathrm{T}}\|_F = \|C - B^{\mathrm{T}} A^{-1} B\|_F$.
Kolumny $\hat{U}$ nie są ortogonalne! Jeżeli $A \succeq 0$, to:

1. wyznacz macierz $Q = A + A^{-1/2} B B^{\mathrm{T}} A^{-1/2}$; niech $Q = U_Q \Lambda_Q U_Q^{\mathrm{T}}$
2. wyznacz $V = \begin{bmatrix} A \\ B^{\mathrm{T}} \end{bmatrix} A^{-1/2} U_Q \Lambda_Q^{-1/2}$

Fowlkes *et al.* (2000) pokazują, że: (a) $\widetilde{S} = \hat{U}\Lambda\hat{U}^{\mathrm{T}} = V\Lambda_Q V^{\mathrm{T}}$, oraz (b) $V^{\mathrm{T}} V = \mathbb{I}$.

Rozszerzenie Nyströma – implementacja

Niech

$$\hat{S} = \hat{U}\Lambda\hat{U}^{\mathrm{T}} = \begin{bmatrix} A & B \\ B^{\mathrm{T}} & B^{\mathrm{T}}A^{-1}B \end{bmatrix}$$

$$\mathbf{d} = \hat{S}\mathbf{e} = \begin{bmatrix} A\mathbf{e}_{\ell} + B\mathbf{e}_{m-\ell} \\ B^{\mathrm{T}}\mathbf{e}_{\ell} + B^{\mathrm{T}}A^{-1}B\mathbf{e}_{m-\ell} \end{bmatrix} = \begin{bmatrix} \mathbf{d}_A + \mathbf{d}_B \\ \mathbf{d}_{B^{\mathrm{T}}} + B^{\mathrm{T}}A^{-1} \cdot \mathbf{d}_B \end{bmatrix}$$

W algorytmie NJW należy znaleźć wektory własne normalizowanej macierzy $\mathcal{S}$. Bloki jej aproksymacji, $\hat{\mathcal{S}}$ mają postać $\hat{A}_{ij} = A_{ij}/\sqrt{d_i d_j}$, $i,j = 1,\ldots,\ell$, $\hat{B}_{ij} = B_{ij}/\sqrt{d_i, d_j}$, $i = 1,\ldots,m$, $j = \ell{+}1,\ldots,m$.

Dziękuję za uwagę