Zad. 11. Program ilustrujący przykładowe zastosowanie rozkładu SVD macierzy

e-mail: andrzej.kedziorski@fizyka.umk.pl
tel.: 56611-3274
pokój: 485
http://www.fizyka.umk.pl/~tecumseh/EDU/MNII/

Zadanie 11

Napisać program rozwiązujący nadokreślony układ równań liniowych

$$\mathbf{Aa} \approx \mathbf{b}$$

dla zagadnień najmniejszych kwadratów korzystając z rozkładu SVD macierzy $\mathbf{A}$.

Przypomnienie: Ogólna liniowa metoda najmniejszych kwadratów

- Postać danych (x_i, y_i, σ_{y_i}), $i = 1, ..., N$, gdzie σ_{y_i} jest oszacowaniem błędu y_i [1]; zakładamy, że błąd x_i jest zaniedbywalnie mały
- Do danych chcemy dopasować funkcję (model liniowy ze względu na parametry)

$$y(x) = \sum_{k=1}^{M} a_k F_k(x),$$

gdzie $F_k(x)$ są ustalonymi funkcjami, a_k - parametry modelu, $N \geqslant M$
- Wartości parametrów a_k wyznaczamy z warunku

$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{y_i - \sum_{k=1}^{M} a_k F_k(x_i)}{\sigma_{y_i}} \right)^2 = \min,$$

co prowadzi do układu M równań normalnych (liniowych) na parametry a_k

$$\frac{\partial(\chi^2)}{\partial a_k} = 0, \quad k = 1, ..., M$$

[1] Wprowadzam oznaczenie σ_{y_i} na błąd wielkości y_i, żeby odróżnić je od wartości szczególnych σ_i w rozkładzie SVD macierzy, który będzie wykorzystywany w dalszej części zajęć

Przypomnienie: Równania normalne

(Wyprowadzenie równań normalnych - patrz zad. 7)
Układ M równań liniowych na parametry a_k, $k = 1, ..., M$

$$\boldsymbol{\alpha}\mathbf{a} = \boldsymbol{\beta},$$

gdzie $\alpha = \mathbf{A}^{\mathbf{Tr}}\mathbf{A}$, $\beta = \mathbf{A}^{\mathbf{Tr}}\mathbf{b}$ oraz

$$\mathbf{A} = \begin{pmatrix} \frac{F_1(x_1)}{\sigma_{y1}} & \frac{F_2(x_1)}{\sigma_{y1}} & \dots & \frac{F_M(x_1)}{\sigma_{y1}} \\ \frac{F_1(x_2)}{\sigma_{y2}} & \frac{F_2(x_2)}{\sigma_{y2}} & \dots & \frac{F_M(x_2)}{\sigma_{y2}} \\ \vdots & \vdots & \dots & \vdots \\ \frac{F_1(x_N)}{\sigma_{yN}} & \frac{F_2(x_N)}{\sigma_{yN}} & \dots & \frac{F_M(x_N)}{\sigma_{yN}} \end{pmatrix} \quad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} \frac{y_1}{\sigma_{y1}} \\ \frac{y_2}{\sigma_{y2}} \\ \vdots \\ \frac{y_N}{\sigma_{yN}} \end{pmatrix}$$

Widać, że

$$\mathbf{A}_{N\times M},\ \mathbf{a}_{M\times 1},\ \mathbf{b}_{N\times 1},\ \boldsymbol{\alpha}_{M\times M},\ \boldsymbol{\beta}_{M\times 1},\ cond(\boldsymbol{\alpha}) = cond(\mathbf{A})^2$$

Nadokreślony układ równań liniowych

$$\mathbf{A}\mathbf{a} \approx \mathbf{b} \quad \Leftrightarrow \quad y_i \approx \sum_{k=1}^{M} a_k F_k(x_i), \quad i = 1, 2, ..., N$$

prowadzi do innego sposobu rozwiązania problemu...

Uwagi

- Zagadnienie optymalizacji parametrów a_k

$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{y_i - \sum_{k=1}^{M} a_k F_k(x_i)}{\sigma_{y_i}} \right)^2 = \min,$$

sprowadza się do minimalizacji (kwadratu) normy euklidesowej (zob. zad. 6 i 7)

$$\left(\|\mathbf{b} - \mathbf{A}\mathbf{a}\|_2 \right)^2 = \min \quad \Leftrightarrow \quad \|\mathbf{b} - \mathbf{A}\mathbf{a}\|_2 = \min$$

gdzie $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{N} x_i^2}$

- Wynika stąd, że $\mathbf{b} - \mathbf{A}\mathbf{a} \approx \mathbf{0}$, czyli $\mathbf{A}\mathbf{a} \approx \mathbf{b}$
- Rozwiązanie w sensie minimalizowania normy euklidesowej zagadnienie $\|\mathbf{A}\mathbf{a} - \mathbf{b}\|_2 = \min$ jest równe $\mathbf{a} = \mathbf{A}^{+}\mathbf{b}$

Przypomnienie: macierz pseudoodwrotna

- Rozkład SVD macierzy $\mathbf{A}_{N\times M}$ i macierz pseudoodwrotna $\mathbf{A}^+_{M\times N}$

$$\mathbf{A} = \mathbf{U\Sigma V}^{Tr} \qquad \mathbf{A}^+ = \mathbf{V\Sigma}^+\mathbf{U}^{Tr},$$

gdzie $\mathbf{U}_{N\times N}$ i $\mathbf{V}_{M\times M}$ to macierze ortogonalne, macierz diagonalna $\mathbf{\Sigma}_{N\times M}$ ma na głównej diagonali wartości szczególne $\sigma_i \geqslant 0$ w kolejności od największej do najmniejszej, tzn. $\sigma_1 > \sigma_2 > ... > \sigma_{\min(N,M)}$, a macierz $\mathbf{\Sigma}^+{}_{M\times N}$ ma na głównej diagonali pseudoodwrotności σ_i^+, tzn.

$$\sigma_i^+ = \begin{cases} \sigma_i^{-1}, & \sigma_i > 0 \\ 0, & \sigma_i = 0 \end{cases}$$

- Dla ustalenia uwagi: $N \geqslant M$, jak w zagadnieniu aproksymacji χ^2
- Wartości szczególne σ_i oraz kolumny macierzy ortogonalnych $\mathbf{U}$ i $\mathbf{V}$

$$\left(\mathbf{A}^{Tr}\mathbf{A}\right)\,\mathbf{v}_k = \sigma_k^2\mathbf{v}_k,\ \ k = 1, ..., M \ \Rightarrow\ \mathbf{V} = [\mathbf{v}_1\ \mathbf{v}_2\ ...\ \mathbf{v}_M] \tag{1}$$

$$\mathbf{u}_j = \frac{1}{\sigma_j}\mathbf{A}\mathbf{v}_j,\ \ \sigma_j > 0 \Rightarrow \mathbf{U} = [\mathbf{u}_1\ \mathbf{u}_2\ ...\ \mathbf{u}_{j_{max}} \ldots],\ \ j_{max} \leqslant M \tag{2}$$

gdzie kolejność kolumn zdeterminowana jest uporządkowaniem
$\sigma_1 \geqslant \sigma_2 \geqslant \sigma_3 \geqslant ... \geqslant \sigma_{j_{max}} > 0$ oraz $j_{max} \leqslant M$

Jawne wyrażenie na parametry a_k (wektor $\mathbf{a}$)

- Przypomnienie: Rozkład SVD macierzy $\mathbf{A}_{\mathbf{N}\times\mathbf{M}}$, gdzie $N \geqslant M$, możemy zapisać następująco

$$\mathbf{A}_{\mathbf{N}\times\mathbf{M}} = \mathbf{U}_{N\times N}\mathbf{\Sigma}_{N\times M}\mathbf{V}^{Tr}_{M\times M} = \sum_{\sigma_i>0} \sigma_i \mathbf{E}_i,$$

  gdzie każda macierz $\mathbf{E}_i = \mathbf{u}_i\mathbf{v}_i^{Tr}$ jest rzędu 1. Analogicznie możemy przedstawić macierz pseudoodwrotną

$$\mathbf{A}^{+}_{\mathbf{M}\times\mathbf{N}} = \mathbf{V}_{M\times M}\mathbf{\Sigma}^{+}_{M\times N}\mathbf{U}^{Tr}_{N\times N} = \sum_{\sigma_i>0} \frac{1}{\sigma_i}\mathbf{v}_i\mathbf{u}_i^{Tr}$$

- Wobec tego rozwiązanie nadmiarowego układu równań liniowych $\mathbf{A}_{N\times M}\mathbf{a}_{M\times 1} \approx \mathbf{b}_{N\times 1}$ w sensie minimalizacji metodą najmniejszych kwadratów, czyli $\mathbf{a} = \mathbf{A}^{+}\mathbf{b}$, można zapisać w następującej postaci

$$\mathbf{a} = \sum_{\sigma_i>0} \frac{\mathbf{u}_i^{Tr}\mathbf{b}}{\sigma_i}\mathbf{v}_i,$$

  gdzie $\mathbf{u}_i$, to i-ta kolumna macierzy $\mathbf{U}_{N\times N}$, a $\mathbf{v}_i$, to i-ta kolumna macierzy $\mathbf{V}_{M\times M}$, oraz $\mathbf{u}_i^{Tr}\mathbf{b}$ jest iloczynem skalarnym wektorów (kolumn $N\times 1$)
- Zauważmy, że potrzebujemy co najwyżej M początkowych kolumn macierzy $\mathbf{U}$

Do zrobienia

- Utworzyć macierze $\mathbf{A}_{N\times M}$ i $\mathbf{b}_{N\times 1}$, $N \geqslant M$
- Dokonać rozkładu SVD macierzy $\mathbf{A}_{N\times M}$; obliczamy co najwyżej M początkowych kolumn macierzy $\mathbf{U}$ wg wzoru

$$\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i, \quad \sigma_i > \epsilon$$

  gdzie $\epsilon > 0$ i $\epsilon \sim 0$.
- Wyznaczamy parametry a_k (wektor $\mathbf{a}$)

$$\mathbf{a} = \sum_{\sigma_i > \epsilon} \frac{\mathbf{u}_i^{Tr}\mathbf{b}}{\sigma_i}\mathbf{v}_i,$$

- Lub po prostu

$$\mathbf{a} = \mathbf{A}^{+}\mathbf{b}$$