10.
Metody obliczeniowe najmniejszych kwadratów

1. Dowód twierdzenia o faktoryzacji macierzy

Twierdzenie 1 *Każdą dodatnio określoną i symetryczną macierz* $\mathbf{M}$ *można przedstawić w postaci*

$$\mathbf{M} = \mathbf{P}\mathbf{P}^T$$

gdzie $\mathbf{P}$ *jest macierzą nieosobliwą (nazywaną pierwiastkiem macierzy* $\mathbf{M}$*).*

Dowód

oznaczmy

$$\begin{aligned} w_i &- \text{wektory własne macirzy } M, & i &= 1, 2, ..., s \\ \lambda_i &- \text{wartości własne macierzy } M, & i &= 1, 2, ..., s \end{aligned}$$

własność

$$Mw_i = \lambda_i w_i$$

oznaczając

$$W = [w_1, w_2, ..., w_s] \text{ oraz } \Lambda = diag(\lambda_i) = \begin{bmatrix} \lambda_1 & 0 & .. & 0 \\ 0 & \lambda_2 & 0 & .. \\ .. & 0 & .. & 0 \\ 0 & .. & 0 & \lambda_s \end{bmatrix}$$

ponieważ macierz M jest symetryczna, to

$$\begin{aligned} &\lambda_i - \text{rzeczywiste} \\ &w_i - \text{parami ortogonalne} \end{aligned}$$

ponieważ macierz M jest dodatnio określona ($M > 0$), to

$$\lambda_i > 0 \text{ dla każdego } i = 1, 2, ..., s$$

wniosek

$$\begin{aligned} & W - \text{macierz ortogonalna} \\ W^T W &= I \Longrightarrow W^{-1} = W^T \end{aligned}$$

a zatem

$$\begin{aligned} MW &= W\Lambda \\ M &= W\Lambda W^{-1} = W\Lambda W^T \text{ (tzw. rozkład spektralny macierzy)} \end{aligned}$$

oznaczając

$$\Lambda = \Lambda^{1/2}\Lambda^{1/2}, \text{ gdzie } \Lambda^{1/2} = diag(\sqrt{\lambda_i}) = \begin{bmatrix} \sqrt{\lambda_1} & 0 & .. & 0 \\ 0 & \sqrt{\lambda_2} & 0 & .. \\ .. & 0 & .. & 0 \\ 0 & .. & 0 & \sqrt{\lambda_s} \end{bmatrix}$$

otrzymujemy

$$W\Lambda W^T = W\Lambda^{1/2}\left(W\Lambda^{1/2}\right)^T$$

stąd

$$M = PP^T, \text{ gdzie } P = W\Lambda^{1/2} \text{ (c.k.d.) } \blacksquare$$

2. Klasyfikacja metod obliczeniowych NK

1) Algorytmy równań normalnych
2) Algorytmy ortogonalizacji (przekształceń ortogonalnych)

Ad 1) podejście trywialne – ogromny problem z odwracaniem macierzy

$$\left.\begin{matrix} X_N \\ Y_N \end{matrix}\right\} \qquad \left.\begin{matrix} X_N^T X_N \rightarrow \left(X_N^T X_N\right)^{-1} \\ X_N^T Y_N \end{matrix}\right\} \qquad a_N = \left(X_N^T X_N\right)^{-1} X_N^T Y_N$$

zadania źle uwarukowane, np.

$$\begin{aligned} X_N^T X_N &= \begin{bmatrix} 1 & 0 \\ 0 & \varepsilon \end{bmatrix}, \qquad \det(X_N^T X_N) = \varepsilon \text{ – bliskie zeru} \\ \text{dla } \varepsilon &= 0.001 \rightarrow \left(X_N^T X_N\right)^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1000 \end{bmatrix} \\ \text{dla } \varepsilon &= 0.0001 \rightarrow \left(X_N^T X_N\right)^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 10000 \end{bmatrix} \text{ – różnica rzędu wielkości} \end{aligned}$$

3. Metoda eliminacji Gaussa

mamy równanie

$$Ma = b$$

$$\begin{aligned}[M, b, I] &\rightarrow \text{ operacje liniowe na wierszach i kolumnach (mnożenie przez } M^{-1}) \rightarrow \\ &\rightarrow [I, a_N, M^{-1}]\end{aligned}$$

Zalety metody:

1) stabilna numerycznie

2) oprócz a_N uzyskujemy także ocenę jakości a_N, czyli macierz kowariancji

$$cov(a_N) = \left(X_N^T X_N\right)^{-1} \sigma_z^2 = M^{-1} \sigma_z^2$$

4. Metoda Cholesky'ego, rozkład LU

$$
\begin{aligned}
Ma &= b \\
M \text{ – dodatnio określona } (M &> 0), \text{ symetryczna } (M^T = M)
\end{aligned}
$$

na podstawie twierdzenia o faktoryzacji

$$
\begin{aligned}
M &= PP^T \\
& P \text{ – macierz nieosobliwa, tzw. pierwiastek macierzy } M
\end{aligned}
$$

dowiedliśmy, że

$$
\begin{aligned}
M &= W\Lambda W^T \text{ (tzw. rozkład spektralny macierzy)} \\
\Lambda &= diag(\lambda_i),\ \lambda_i > 0
\end{aligned}
$$

isnieje inny rozkład macierzy M, tzw. rozkład LU

$$M = LDL^T$$

gdzie

$$
\begin{aligned}
L \text{ – macierz trójkątna dolna, tzn. dla } i &< j \text{ zachodzi } L[i,j] = 0 \\
D \text{ – macierz diagonalna, } D &= diag(d_i),\ d_i > 0
\end{aligned}
$$

stąd mamy

$$
\begin{aligned}
M &= LD^{1/2}D^{1/2}L^T \\
\text{gdzie } D^{1/2} &= diag(\sqrt{d_i})
\end{aligned}
$$

przyjmując oznaczenie

$$\overline{L} = LD^{1/2}$$

otrzymujemy

$$\begin{aligned} M &= \overline{L}\overline{L}^T \\ & \quad \overline{L} \text{ – macierz trójkątna dolna} \\ & \quad \overline{L}^T \text{ – macierz trójkątna górna} \\ \det M &= \det \overline{L}\overline{L}^T = \det \overline{L} \det \overline{L}^T > 0 \end{aligned}$$

Schemat metody obliczeniowej

$$\overline{L}\overline{L}^T a = b$$

Etap 1.
oznaczając $\alpha = \overline{L}^T a$ rozwiązujemy równanie "zewnętrzne"

$$\overline{L}\alpha = b$$

ponieważ $\det \overline{L} > 0$ to elementy diagonalne $\overline{L}$ są różne od zera, rozwiązanie α_N istnieje i jest jednoznaczne
Etap 2.
wartość α_N jest znana z etapu 1
rozwiązujemy równanie "wewnętrzne"

$$\overline{L}^T a = \alpha_N$$

ponieważ $\det \overline{L}^T > 0$ to elementy diagonalne $\overline{L^T}$ są różne od zera, rozwiązanie a_N istnieje i jest jednoznaczne

Zalety metody
1) prostota operacji przy rozwiązywaniu równań
Etap 1 – tzw. metoda podstawiania "od góry"
Etap 2 – tzw. metoda podstawiania "od dołu"

2) dobre uwarunkowanie zadania, gdyż

$$\begin{aligned} \det M &= \left(\det \overline{L}\right)^2 \\ \det \overline{L} &= \sqrt{\det M} \end{aligned}$$

zatem dla $0 < \det M < 1$

$$\det \overline{L} = \det \overline{L}^T > \det M$$

czyli zadania z etapów 1 i 2 są lepiej uwarunkowane

Ad 2) metody przekształceń ortogonalnych – nie wymagają tworzenia równań normalnych

5. Metoda odbić Householdera

Definicja 1 *Macierzą odbicia Householdera nazywamy macierz postaci*

$$P = I - 2ww^T$$

gdzie $\|w\| = 1$*, tj.* $w^T w = 1$*.*

Interpretacja

$$Pw = (I - 2ww^T)w = w - 2ww^T w = w - 2w = -w$$

Własności:
(i) $P = P^T$ (symetria)
(ii) $P^T P = PP^T = I$ (ortogonalność)
(iii) dla każdego wektora x istnieje macierz P_j, taka że

$$P_j x = \pm \|x\| e_j, \text{ gdzie } e_j \text{ jest } j\text{-tym wersorem } e_j = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

(iiii) dla ciągu macierzy Householdera $P_1, P_2, ..., P_s$ przekształcenie złożenia

$$\Psi = P_s \cdot P_{s-1} \cdot ... \cdot P_2 \cdot P_1$$

jest macierzą ortogonalną, tj.

$$\Psi^T \Psi = I$$

Dowody

(i) oczywiste – różnica macierzy symetrycznych jest symetryczna

(ii)

$$PP^T = (I - 2ww^T)(I - 2ww^T) = I - 2ww^T - 2ww^T + 4ww^Tww^T = I - 4ww^T + 4ww^T = I$$

(iii)

$$\begin{aligned} & x \text{ – dowolny wektor} \\ u_j &= x \pm \|x\|\, e_j \\ P_j &= I - \frac{u_j u_j^T}{H_j}, \text{ gdzie } H_j = \frac{1}{2}\|u_j\|^2 = \frac{1}{2}u_j^T u_j \end{aligned}$$

macierz P_j jest macierzą Householdera, ponieważ

$$P_j = I - 2\frac{u_j}{\|u_j\|}\frac{u_j^T}{\|u_j\|} \text{ (wektory o długości 1)}$$

$$\begin{aligned} P_j x &= (I - \frac{u_j u_j^T}{H_j})x = x - u_j\frac{2u_j^T x}{u_j^T u_j} = x - u_j\frac{2(x \pm \|x\|\, e_j)^T x}{(x \pm \|x\|\, e_j)^T (x \pm \|x\|\, e_j)} = \\ &= \ldots = x - u_j = x - x \pm \|x\|\, e_j = \pm \|x\|\, e_j \end{aligned}$$

(iiii)

$$\Psi^T \Psi = P_1^T P_2^T \ldots P_{s-1}^T \underbrace{P_s^T P_s} P_{s-1} \ldots P_2 P_1 = I$$

Lemat 1 *Dla każdej macierzy $X_{N\times s}$ istnieje taki ciąg macierzy Householdera $\{\widetilde{P}_1, \widetilde{P}_2, ..., \widetilde{P}_s\}$, że przekształcenie*

$$\Psi = \widetilde{P}_s \widetilde{P}_{s-1}, ... \widetilde{P}_2 \widetilde{P}_1$$

ma własność

$$\Psi X_{N\times s} = \begin{bmatrix} R \\ 0 \end{bmatrix}$$

gdzie R jest macierzą trójkątną górną o rozmiarach $s \times s$.

Twierdzenie 2 *Oszacowanie NK jest równoważne rozwiązaniu układu równań*

$$Ra = \eta_R$$

gdzie η_R jest wektorem zawierającym s pierwszych elementów wektora ΨY_N.

Rozwiązanie tego układu równań istnieje i jest jednoznaczne. Pytanie dotyczy samego oszacowania NK.

Dowód

$$\begin{aligned}
Q(a) &= \|X_N a - Y_N\|_e^2 \to \min_a \\
Q(a) &= (X_N a - Y_N)^T (X_N a - Y_N) = \\
&= (X_N a - Y_N)^T \Psi^T \Psi (X_N a - Y_N) = \\
&= [\Psi(X_N a - Y_N)]^T [\Psi(X_N a - Y_N)] = \\
&= [\Psi X_N a - \Psi Y_N]^T [\Psi X_N a - \Psi Y_N] = \\
&= [\begin{bmatrix} R \\ 0 \end{bmatrix} a - \begin{bmatrix} \eta_R \\ \eta_z \end{bmatrix}]^T [\begin{bmatrix} R \\ 0 \end{bmatrix} a - \begin{bmatrix} \eta_R \\ \eta_z \end{bmatrix}] = \\
&= \begin{bmatrix} Ra - \eta_R \\ -\eta_z \end{bmatrix}^T \begin{bmatrix} Ra - \eta_R \\ -\eta_z \end{bmatrix} = (Ra - \eta_R)^T (Ra - \eta_R) + \eta_z^T \eta_z = \\
&= \|Ra - \eta_R\|_e^2 + \|\eta_z\|_e^2 \to \min_a
\end{aligned}$$

zatem

$$Q(a) \to \min_a \Rightarrow Ra = \eta_R$$