Pokročilé numerické metody I
Metoda nejmenších čtverců

Jiří Zelinka

Metoda nejmenších čtverců

Systém lineárních rovnic

$$Ax = b$$

Reziduum:

$$r_x = b - Ax$$

Definice

Vektor $\hat{x}$ se nazývá *řešením ve smyslu NČ*, jestliže

$$\forall x : \|r_x\| \geq \|r_{\hat{x}}\|$$

Věta 1

Řešení ve smyslu NČ vždy existuje a je řešením systému normálních rovnic

$$A^* Ax = A^* b.$$

Důkaz: Pomocí ortogonálního rozkladu vektoru $b = b_1 + b_2$, kde $b_1 \in \mathcal{R}(A)$.

Důsledek

Systému normálních rovnic je vždy řešitelný.

Věta 2

Každé řešení systému normálních rovnic je řešením ve smyslu NČ.

Věta 3

Řešení ve smyslu NČ je jediné právě tehdy, když sloupce A jsou LNZ.

Numerický přístup k MNČ

Aproximace dat daným typem funkce

$\boldsymbol{x} = (x_0, \ldots, x_N)$ – uzly
$\boldsymbol{f} = (f_0, \ldots, f_N)$ – funkční hodnoty v uzlech neznáme funkce f

$\varphi_0, \ldots, \varphi_M$ – zadané bázové funkce (např. $\varphi_k = x^k$)

Data (x_i, f_i) aproximujeme funkcí $\varphi = \sum\limits_{k=0}^{M} c_k \varphi_k$

Hledáme minimum výrazu

$$\sigma^2(c_0, \ldots, c_M) = \sum_{j=0}^{N} (f_j - \varphi(x_j))^2$$

Vektor $c = (c_0, \ldots, c_M)^T$ je řešením normálního systému

$A^T A x = A^T \boldsymbol{f}$ pro matici $A = \begin{pmatrix} \varphi_0(x_0) & \cdots & \varphi_M(x_0) \\ \vdots & & \vdots \\ \varphi_0(x_N) & \cdots & \varphi_M(x_N) \end{pmatrix}$.

Statistický přístup k MNČ

Lineární regresní model:

$$Y = \beta_0 X_0 + \dots \beta_M X_M + \varepsilon$$

Y, ε – náhodné veličiny, β_j – neznámé parametry.
Pro N pozorování:

$$\begin{aligned} Y_0 &= \beta_0 X_{00} + \dots \beta_M X_{M0} + \varepsilon_0 \\ &\vdots \\ Y_N &= \beta_0 X_{0N} + \dots \beta_M X_{MN} + \varepsilon_N \end{aligned}$$

t.j.

$$Y = X\beta + \varepsilon$$

X – matice plánu, β – neznámé parametry

Minimalizujeme výraz (rozptyl odhadu β)

$$(Y - X\beta)^T (Y - X\beta)$$

Řešení je odhad parametru β (BLUE)

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Mooreova-Penroseova pseudoinverze

Definice

Buď A matice typu $m \times n$ Matice X se nazývá její pseudoinverzní maticí, jestliže platí následující vlastnosti (Penroseovy axiomy)

(1) $AXA = A$

(2) $XAX = X$

(3) $(AX)^* = AX$ (AX je hermitovská)

(4) $(XA)^* = XA$ (XA je hermitovská)

Věta 4

Pseudoinevrzní matice vždy existuje a je určena jednoznačně.

Označení

Pseudoinverzní matici k matici A značíme A^+.

Lemma (skeletní rozklad)
Nechť A je typu $m \times n$, $r(A) = r$. Pak existují matice B typu $m \times r$ a C typu $r \times n$, $r(B) = r(C) = r$, že platí $A = BC$.
Důkaz: Jako B vybereme nezávislé sloupce matice A, C pak dopočítáme.

Důkaz vety o existenci a jednoznačnosti

Jednoznačnost dostaneme z Penroseových axiomů.

Pro nulovou matici A je $A^+ = A^*$.

Pokud má A nezávislé sloupce, pak $A^+ = (A^*A)^{-1}A^*$, pro matici s nezávislými řádky je $A^+ = A^*(AA^*)^{-1}$.

V obecném případě uděláme skeletní rozklad $A = BC$ a dokážeme, že $A^+ = C^+B^+$.

Věta 5 Vlastnosti pseudoinverzní matice

1. pro regulární A je $A^+ = A^{-1}$
2. $(A^+)^+ = A$
3. $(A^*)^+ = (A^+)^*$
4. $(A^*A)^+ = A^+A^{+*}$, $(AA^*)^+ = A^{+*}A^+$
5. $A^+ = (A^*A)^+A^* = A^*(AA^*)^+$
6. $h(A) = h(A^+) = h(A^*) = h(A^+A) = h(A^*A) = \dots$
7. $\mathcal{R}(A^+) = \mathcal{R}(A^*)$
8. $\mathcal{N}(A^+) = \mathcal{N}(A^*)$
9. obecně neplatí $(AB)^+ = B^+A^+$
10. $c \neq 0 \Rightarrow (cA)^+ = \frac{1}{c}A^+$
11. $A^+ = (QA^*A)^+QA^*$ pro regulární Q

M–P pseudoinverze a MNČ

Lemma

Buď $P = AA^+$, $Q = A^+A$. Pro libovolné vektory x, y platí

- $\|Ax + (I - P)y\|^2 = \|Ax\|^2 + \|(I - P)y\|^2$
- $\|A^+y + (I - Q)x\|^2 = \|A^+y\|^2 + \|(I - Q)x\|^2$

Důkaz: dokážeme kolmost.

Věta 6

Buď A typu $m \times n$, b typu $n \times 1$, $\hat{x} = A^+b$. Pak pro každé $x \neq \hat{x}$ platí: $\|b - Ax\| \geq \|b - A\hat{x}\|$, a v případě, že $\|b - Ax\| = \|b - A\hat{x}\|$, pak $\|x\| > \|\hat{x}\|$.

Interpretace:

$\hat{x} = A^+b$ je řešení ve smyslu NČ s minimální normou a je jediné s touto vlastností.

Důkaz

$Ax - b = Ax - AA^+b + AA^+b - b = A(x - A^+b) + (I - AA^+)(-b)$
Podle předchozího lemmatu
$\|Ax - b\|^2 = \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 \geq \|A\hat{x} - b\|^2$,
přičemž rovnost nastane pouze pro
$\|A(x - \hat{x})\|^2 = 0$, tj. $Ax = A\hat{x}$.

Pokud $Ax = A\hat{x}$, pak $A^+Ax = A^+A\hat{x} = A^+AA^+b = A^+b = \hat{x}$.
Potom $x = \hat{x} + (x - \hat{x}) = A^+b + (I - A^+A)x$.
Podle druhé rovnosti předchozího lemmatu platí
$\|x\|^2 = \|\hat{x}\|^2 + \|x - \hat{x}\|^2$, takže $\|x\|^2 \geq \|\hat{x}\|^2$ a rovnost nastane pouze pro $x = \hat{x}$.

M–P pseudoinverze a systém lineáních rovnic

Věta 7 (Penrose 1955)
Systém rovnice $AXB = C$ je řešitelný právě tehdy, když platí $AA^+CB^+B = C$, V tom případě je obecné řešení systému ve tvaru

$$X = A^+CB^+ + Y - A^+AYBB^+$$

pro libovolnou matici Y.

Důsledek

Systém $Ax = b$ je řešitelný právě tehdy, když platí $AA^+b = b$. Obecné řešení tohoto systému je pak

$$x = A^+b + (I - A^+A)y$$

pro libovolný vektor y.

Grevilleův algoritmus

Výpočet Mooreovy–Penroseovy pseudoinverze po řádcích (sloupcích)

A: matice, A_k: prvních k řádků, a_k: k-tý řádek, známe A^+_{k-1}

Položme:

$$
\begin{aligned}
d_k &= a_k A^+_{k-1} \\
c_k &= a_k - d_k A_{k-1} \\
b_k &= \begin{cases} c_k^+ & c_k \neq 0 \\[2ex] \frac{1}{1+d_k d_k^*} A^+_{k-1} d_k^* & c_k = 0 \end{cases} \\
A^+_k &= \left(A^+_{k-1} - b_k d_k | b_k \right)
\end{aligned}
$$

Sloupcová verze

A: matice, A_k: prvních k sloupců, a_k: k-tý sloupec, známe A_{k-1}^+

Položme:

$$\begin{aligned} d_k &= A_{k-1}^+ a_k \\ c_k &= a_k - A_{k-1} d_k \\ b_k &= \begin{cases} c_k^+ & c_k \neq 0 \\ \frac{1}{1+d_k d_k^*} d_k^* A_{k-1}^+ & c_k = 0 \end{cases} \\ A_k^+ &= \begin{pmatrix} A_{k-1}^+ - d_k b_k \\ - - - - - - - \\ b_k \end{pmatrix} \end{aligned}$$

Využití: MNČ – přidání pozorování nebo vysvětlující proměnné

Nelineární MNČ

Aproximace dat nelineární funkcí závislé na parametrech, které se snažíme určit

Příklad:

Michaelisův-Mentenové model kinetiky enzymů

$$f(x, \boldsymbol{\beta}) = \frac{\beta_1 x}{\beta_2 + x}$$

$x_0, \ldots, x_N$ – uzly
$y_0, \ldots, y_N$ – funkční hodnoty, $y_j \approx f(x_j, \boldsymbol{\beta})$

Rezidua: $r_j(\boldsymbol{\beta}) = y_j - f(x_j, \boldsymbol{\beta})$ Hledáme minimum výrazu

$$S(\boldsymbol{\beta}) = \sum_{j=0}^{N} r_j(\boldsymbol{\beta})^2$$

Nutná podmínka pro minimum:

$$0 = \frac{\partial S}{\partial \beta_i} = 2\sum_{j=0}^{N} r_j(\boldsymbol{\beta})\frac{\partial r_j(\boldsymbol{\beta})}{\partial \beta_i} = -2\sum_{j=0}^{N} J_{ji} r_j(\boldsymbol{\beta}) \quad \forall i,$$

pro $\frac{\partial r_j(\boldsymbol{\beta})}{\partial \beta_i} = -\frac{\partial f(x_j,\boldsymbol{\beta})}{\partial \beta_i} = -J_{ji}$,
$J = (J_{ji})$ je Jacobiova matice, $J = J(\boldsymbol{\beta})$.

Máme systém nelineárních rovnic – řešíme iterační metodou:

$\boldsymbol{\beta}^0$ - počáteční iterace, $\boldsymbol{\beta}^k + \Delta\boldsymbol{\beta}^k = \boldsymbol{\beta}^{k+1} \approx \boldsymbol{\beta}$

Taylorův rozvoj v parametrech $\boldsymbol{\beta}$:

$$f(x_j,\boldsymbol{\beta}) \approx f(x_j,\boldsymbol{\beta}^k) + \sum_i \frac{\partial f(x_j,\boldsymbol{\beta}^k)}{\partial \beta_i}(\beta_i - \beta_i^k) \approx$$
$$\approx f(x_j,\boldsymbol{\beta}^k) + \sum_i J_{ji}\Delta\beta_i^k.$$

Označení: $\Delta y_j^k = y_j - f(x_j, \boldsymbol{\beta}^k)$

$$r_j(\boldsymbol{\beta}) = y_j - f(x_j, \boldsymbol{\beta}) = (y_j - f(x_j, \boldsymbol{\beta}^k)) + (f(x_j, \boldsymbol{\beta}^k) - f(x_j, \boldsymbol{\beta})) \approx$$
$$\approx \Delta y_j^k - \sum_i J_{ji} \Delta\beta_i^k = \Delta y_j^k - \sum_l J_{jl} \Delta\beta_l^k$$

Dosazením do nutné podmínky pro minimum dostaneme

$$-2 \sum_j J_{ji} \left(\Delta y_j^k - \sum_l J_{jl} \Delta\beta_l^k \right) = 0, \quad \forall i$$

Po úpravě dostaneme systém normálních rovnic

$$J^T J \Delta\beta^k = J^T \Delta y^k, \quad \text{pro } J = J(\boldsymbol{\beta}^k).$$

Gaussův-Newtonův algoritmus: volíme počáteční iteraci β^0, z předchozího systému normálních rovnic určíme přírůstek $\Delta\beta^0$, z něj iteraci β^1 atd.