3 Metoda nejmenších čtverců

Břetislav Fajmon, UMAT FEKT, VUT Brno

Téma je podrobně zpracováno ve skriptech [1], kapitola 6, strany 73-80.

Jedná se o třetí možnou metodu aproximace, která se od předchozích dvou metod (interpoleční polynom a splajn) liší tím, že hledaná funkce NEMUSÍ procházet zadanými body $[x_0, y_0]$, $[x_1, y_1]$, ..., $[x_n, y_n]$. To je výhodné v tom, že mezi naměřenými body mohou mít dva různé body stejnou souřadnici x (např. může nastat situace $x_2 = x_3$).

Důležitá poznámka: U metody nejmenších čtverců (MNČ, anglicky least squares method = LSM) z charakteru naměřených dat (nejlépe z obrázku) předem stanovíme, jakého typu bude funkce, kterou chceme najít:

- Lineární aproximace: hledáme funkci $y = a + bx$.
- Kvadratická aproximace: hledáme funkci typu $y = a + bx + cx^2$.
- Exponenciální aproximace: hledáme funkci typu $y = l \cdot a^x$, kde a, l jsou neznámé konstanty.
- Jak řešit posunutou exponenciální aproximaci, když potřebujeme najít funkci typu $y = k + l \cdot a^x$?

3.1 Metoda nejmenších čtverců pro přímku

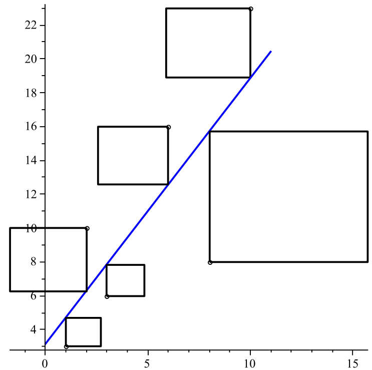
Jsou dány body $[x_0, y_0]$, $[x_1, y_1]$, $\ldots$, $[x_n, y_n]$. Hledáme funkci typu $y = a + bx$, která aproximuje tyto body.

Hledejme přímku $y = a + bx$ tak, že součet čtverců odchylek naměřené (zadané) souřadnice y_i a souřadnice $a + bx_i$ vypočtené pomocí hledané přímky

$$S = \sum_{0}^{n} (y_i - (a + bx_i))^2$$

je minimální možný.

Název metoda nejmenších čtverců se vžil díky grafickému názoru, že totiž opravdu hledáme takovou přímku, pro kterou je součet ploch čterců odchylek minimální – viz obr.:



(uvedené obrazce jsou skutečně čtverce – o obdélníky se jedná jen zdánlivě, díky různým měřítkům na obou osách).

$S(a,b) = \sum_0^n (y_i - (a + bx_i))^2$ **je funkce dvou neznámých proměnných** a, b. **Najít její minimum znamená najít stacionární bod (o kterém lze dokázat, že je minimem této funkce** S**), tj. řešit systém rovnic** $S'_a = 0$, $S'_b = 0$. **Derivujeme tedy podle proměnných** a, b, **ostatní písmena považujeme za konstanty:**

$$\begin{aligned} \sum_0^n 2(y_i - a - bx_i)(-1) &= 0; \\ \sum_0^n 2(y_i - a - bx_i)(-x_i) &= 0. \end{aligned}$$

Roznásobením závorek, vydělením rovnic dvěma a

rozdělením sumy k jednotlivým členům dostaneme

$$
\begin{aligned}
a(n+1) + b\sum x_i &= \sum y_i; \\
a\sum x_i + b\sum (x_i^2) &= \sum (x_i y_i).
\end{aligned}
$$

Tyto rovnice se nazývají normální rovnice. Jsou to vlastně dvě LINEÁRNÍ rovnice o dvou neznámých a, b, **které vypočteme, a hledanou funkci máme nalezenu.**

Příklad 3.1. Metodou nejmenších čtverců nalezněte přímku aproximující zadané body:

x_i	1	1	2	3
y_i	2	3	1	4

.

Výsledek:

$$y = 1{,}5455 + 0{,}5455 \cdot x.$$

Tatáž úloha, ale v maticovém zápisu: Hledáme přímku $y = a + bx$. Až ji najdeme, bude v bodech x_i platit

$$
\begin{aligned}
y'_0 &= a + b \cdot x_0; \\
y'_1 &= a + b \cdot x_1; \\
&\vdots \\
y'_n &= a + b \cdot x_n;
\end{aligned}
$$

kde y'_i nejsou derivace, ale funkční hodnoty vypočtené dosazením x_i do rovnice přímky.

Maticově tento zápis můžeme psát

$$\begin{pmatrix} y'_0 \\ y'_1 \\ \vdots \\ y'_n \end{pmatrix} = \begin{pmatrix} 1 & x_0 \\ 1 & x_1 \\ \vdots & \\ 1 & x_n \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} .$$

což označením vektorů a matice je

$$\vec{y'} = Z \cdot \vec{a}.$$

Je možné ukázat, že normálním rovnicím pro určení koeficientů a, b odpovídá maticový zápis

$$Z^T \cdot \vec{y} = Z^T \cdot Z \cdot \vec{a},$$

kde

$$\vec{y} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{pmatrix}$$

jsou původní naměřené y-ové souřadnice zadávaných bodů. Odtud lze maticově psát přímo vzorec pro výpočet hledaných koeficientů a, b:

$$\vec{a} = \begin{pmatrix} a \\ b \end{pmatrix} = \left(Z^T \cdot Z\right)^{-1} \cdot Z^T \cdot \vec{y}.$$

Příklad 3.2. V našem příkladu aproximace bodů

x_i	1	1	2	3
y_i	2	3	1	4

pomocí přímky máme

$$Z = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix},$$

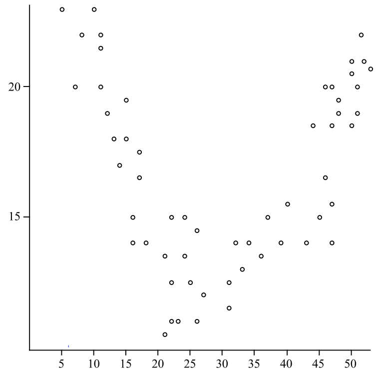
(v prvním sloupci matice Z jsou jedničky proto, že hledáme přímku $y = a \cdot 1 + b \cdot x$ (koeficient a stojí u funkce konstantně rovné jedné). Ve druhém sloupci matice Z jsou hodnoty x_i proto, že koeficient b stojí ve vzorci lineární funkce u x)

$$Z^T \cdot Z = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 4 & 7 \\ 7 & 15 \end{pmatrix},$$

$$\vec{a} = \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 4 & 7 \\ 7 & 15 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 3 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 1{,}5455 \\ 0{,}5455 \end{pmatrix}.$$

3.2 Metoda nejmenších čtverců pro parabolu

Samozřejmě že vztah mezi dvěma veličinami nemusí být lineární. Například pokud hledáme model (= funkci) popisující závislost věku člověka na době, za jakou uběhne sto metrů:



Z obrázku je zřejmé, že parabola uvedenou závislost popíše lépe než přímka, tj. hledáme koeficienty a, b, c tak, aby funkce $y = a + bx + cx^2$ byla nejvhodnějším modelem ze všech těchto parabolických funkcí.

Hledejme tedy kvadratickou funkci $y = a + bx + cx^2$ tak, aby součet čtverců odchylek naměřené (zadané) souřadnice y_i a souřadnice $a + bx_i + cx_i^2$ vypočtené pomocí hledané kvadratické funkce

$$S = \sum_{0}^{n} (y_i - (a + bx_i + cx_i^2))^2$$

byl minimální možný.

$S(a,b,c)=\sum_0^n(y_i-(a+bx_i-cx_i^2))^2$ je funkce tří neznámých proměnných a, b, c. Najít její minimum znamená najít stacionární bod (o kterém lze dokázat, že je minimem této funkce S), tj. řešit systém rovnic $S'_a=0$, $S'_b=0$, $S'_c=0$. Derivujeme tedy podle proměnných a, b, c ostatní písmena považujeme za konstanty:

$$\begin{aligned}
\sum_0^n 2(y_i-a-bx_i-cx_i^2)(-1) &= 0;\\
\sum_0^n 2(y_i-a-bx_i-cx_i^2)(-x_i) &= 0;\\
\sum_0^n 2(y_i-a-bx_i-cx_i^2)(-x_i^2) &= 0.
\end{aligned}$$

Roznásobením závorek, vydělením dvěma a rozdělením sumy k jednotlivým členům dostaneme tvar normálních rovnic pro parabolu:

$$\begin{aligned} a(n+1) + b\sum x_i + c\sum(x_i^2) &= \sum y_i; \\ a\sum x_i + b\sum(x_i^2) + c\sum(x_i^3) &= \sum(x_i y_i); \\ a\sum(x_i^2) + b\sum(x_i^3) + c\sum(x_i^4) &= \sum(x_i^2 y_i) \end{aligned}$$

Příklad 3.3. **Parabolou aproximujte tytéž zadané body:**

x_i	1	1	2	3
y_i	2	3	1	4

.

Výsledek:

$$y = 8{,}5 - 8{,}25 \cdot x + 2{,}25 \cdot x^2.$$

Kdybychom chtěli použít (s počítačem) vzorec

$$\vec{a} = \begin{pmatrix} a \\ \\ b \\ \\ c \end{pmatrix} = \left(Z^T \cdot Z\right)^{-1} \cdot Z^T \cdot \vec{y},$$

tak

$$Z = \begin{pmatrix} 1 & 1 & 1^2 \\ 1 & 1 & 1^2 \\ 1 & 2 & 2^2 \\ 1 & 3 & 3^2 \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} 2 \\ 3 \\ 1 \\ 4 \end{pmatrix}$$

(první sloupce matice Z je určen hodnotami 1, druhý hodnotami x_i, třetí hodnotami x_i^2).

3.3 Metoda nejmenších čtverců pro exponenciální funkci typu $y = l \cdot a^x$

Pokud bychom postupovali „klasicky", máme

$$S(a, l) = \sum_{0}^{n} (y_i - l \cdot a^{x_i})^2$$

a řešením rovnic $S'_a = 0$, $S'_l = 0$ dostaneme

$$\begin{aligned} \sum_{0}^{n} 2(y_i - l \cdot a^{x_i})(-a^{x_i}) &= 0; \\ \sum_{0}^{n} 2(y_i - l \cdot a^{x_i})(-l \cdot x_i \cdot a^{x_i - 1}) &= 0, \end{aligned}$$

což je systém nelineárních rovnic, jehož řešení je stejně obtížné jako řešení původní úlohy (např. Newtonovou metodou pro dvě nelineární rovnice – viz kapitola 5 skript).

Proto se tímto způsobem hledání paraboly neprovádí, nýbrž jistým způsobem převedeme úlohu na úlohu nalézt lineární funkci: Zlogaritmováním rovnice

$$y = l \cdot a^x$$

dostaneme

$$\ln y = \ln l + x \cdot \ln a.$$

Označením $z = \ln y$**,** $A = \ln l$**,** $B = \ln a$ **máme najít**

funkci typu

$$z = A + x \cdot B$$

pro neznámé koeficienty A**,** B **– a zde lze užít MNČ pro lineární funkci. Dostaneme tedy normální rovnice tvaru**

$$\begin{aligned} A(n+1) + B\sum x_i &= \sum \ln y_i; \\ A\sum x_i + B\sum(x_i^2) &= \sum(x_i \cdot \ln y_i), \end{aligned}$$

které jsou lineární. Po nalezení A**,** B **musíme užít zpětné transformace pro nalezení původních konstant:** $l = e^A$**,** $a = e^B$**.**

Příklad 3.4. V našem stále stejném příkladu aproximujte zadané čtyři body funkcí typu $y = l \cdot a^x$:

x_i	1	1	2	3
y_i	2	3	1	4
$\ln y_i$	0,69315	1,0986	0	1,3863

(do zadání byl přidán řádek hodnot $\ln y_i$, které jsou při výpočtu potřeba; tento řádek v zadání příkladu nebývá, a musíme jej tam dodat výpočtem!!!).

Výsledek: **Dostaneme** $A = 0{,}5473$**,** $B = 0{,}1413$**, a odtud lze určit**

$$l = e^A = 1{,}729; \quad a = e^B = 1{,}152.$$

Tedy hledaná funkce předem určeného typu má tvar

$$y = 1{,}729 \cdot 1{,}152^x.$$

3.4 Poznámka k možnostem hledání exponenciální funkce při aproximaci pomocí MNČ i jinak

Následující důležité poznámky jsou vzaty ze skript [2], včetně ilustračního příkladu. Některé z těchto věcí nejsou součástí skript [1]. Obsah je nepovinný, ale užitečný. Vše bude vysvětleno na následujícím příkladu:

Příklad 3.5. V letech 1987–1995 byly v jistém hudebním vydavatelství prodány (v tisících kusů) tyto počty hudebních CD:

rok	**1987**	**1988**	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**
x_i	1	2	3	4	5	6	7	8	9
y_i	3	10	15	21	35	42	58	81	110

Aproximujte tyto hodnoty exponenciální funkcí (místo skutečných let byla volena menší čísla x_i, která představují uvedených devět let).

Řešení: dosazením do MNČ pro exponenciální funkci (normální rovnice na str. 19) hledáme funkci typu $y = l \cdot a^x$ **a najdeme**

$$y = 3{,}6052 \cdot 1{,}4938^x.$$

Při výpočtu jsme vlastně nehledali minimum funkce
$S = \sum(y_i - l \cdot a^{x_i})^2$**, ale minimum součtu čtverců po linearizaci, tj. minimum funikce**

$$S_l = \sum \left(\ln y_i - \ln l - x_i \cdot \ln a\right)^2$$

(kde tedy $A = \ln l$ **a** $B = \ln a$ **a využili jsme normální rovnice ze strany 19). Nicméně, pokud už máme nalezenu funkci** y**, nic nám nebrání dosadit do původního**

„klasického“ součtu čtverců a spočítat

$$S = \sum_{0}^{8} (y_i - 3{,}6052 \cdot 1{,}4938^{x_i})^2 = 726{,}546.$$

Příklad 3.6. **V letech 1987–1995 byly v jistém hudebním vydavatelství prodány (v tisících kusů) tyto počty hudebních CD:**

rok	1987	1988	1989	1990	1991	1992	1993	1994	1995
x_i	1	2	3	4	5	6	7	8	9
y_i	3	10	15	21	35	42	58	81	110

Aproximujte tyto hodnoty pomocí MNČ modifikované „váženým součtem“ neboli „součtem s vahami“.

Řešení: **Ukazuje se, že místo hledání minima funkce**

$$S_l = \sum (\ln y_i - \ln l - x_i \cdot \ln a)^2$$

(což je vlastně minimalizovaná funkce v oddílu 3.3, tj. MNČ pro přímku aplikovaná na zlinearizovanou funkci) dává lepší výsledky hledání minima funkce

$$S_{lw} = \sum w_i \cdot (\ln y_i - \ln l - x_i \cdot \ln a)^2 \,,$$

tedy zlinearizovaná exponenciální funkce + navíc použijeme tzv. váhy $w_i = y_i^2$ (druhé mocniny y-ových souřadnic zadaných bodů), kde hodnotou y_i^2 násobíme i-tou závorku.

V případě minimalizace S_{lw} dostaneme normální rovnice ve tvaru

$$\begin{aligned}\ln l \cdot \sum y_i^2 + \ln a \cdot \sum x_i y_i^2 &= \sum y_i^2 \ln y_i;\\ \ln l \cdot \sum x_i y_i^2 + \ln a \cdot \sum x_i^2 y_i^2 &= \sum x_i y_i^2 \ln y_i\end{aligned}$$

(jedná se o tytéž rovnice jako na str. 19 s tím doplněním, že každý člen v sumách je vynásoben ještě hodnotou y_i^2).

Využitím těchto rovnic na naše zadání příkladu dostaneme $\ln l = 1{,}8483$**,** $\ln a = 0{,}317$**, tedy po odlogaritmování máme** $l = e^{1{,}8483} = 6{,}349$**,** $a = e^{0{,}317} = 1{,}373$**, a tak**

hledaná funkce je tvaru

$$y = 6{,}349 \cdot 1{,}373^x.$$

Pokud už máme nalezenu funkci y, nic nám nebrání dosadit do původního „klasického“ součtu čtverců a spočítat

$$S = \sum_0^8 (y_i - 6{,}349 \cdot 1{,}373^{x_i})^2 = 58{,}255.$$

Je vidět, že „modifikovaná“ netoda dává nižší součet odchylek než „linearizovaná“ metoda.

Skripta [2], str. 30-32, uvádí ještě jakousi jinou metodu, která už není typu metody nejmenších čtverců:

metoda částečných součtů hledá aproximační křivku tvaru $y = k + l \cdot a^x$ a nalezne funkci

$$y = -10{,}8313 + 11{,}81813 \cdot 1{,}292061^x;$$

pokud pro tuto funkci vypočteme „klasický“ součet odchylek S, dostaneme ještě nižší hodnotu, než u obou zmiňovaných varinat MNČ pro exponenciálu:

$$S = \sum_0^8 \left(y_i + 10{,}8313 - 11{,}81813 \cdot 1{,}292061^{x_i}\right)^2 = 29{,}797.$$

Tato metoda částečných součtů je velmi rozumnou alternativou MNČ, protože nemusí nutně mít osu x jako asymptotu grafu – asymptotou je přímka $y = k$. Omezení této metody:

- $a > 0$ **... tohle je velmi rozumný předpoklad a žádné ozemení moc nedává – i při MNČ nás zajímal zejména tento případ;**
- **metodu částečných součtů lze užít jen pro počet zadaných bodů dělitelný třemi, čehož lze lehko dosáhnout eventuelním vypuštěním prvních jedné nebo dvou hodnot z výpočtu;**

- **musí také platit, že** x_i **tvoří posloupnost bodů** $1, 2, \dots, n$ – **toto omezení některé situace vylučuje a omezuje její užití jen na situace, kde** $x_{i+1} - x_i$ **je konstantní; přesto je tato metoda užitečná, pokud zpracováváme posloupnost čísel, kdy každé udává hodnotu jisté veličiny za stále stejné období.**

Vysvětlení metody: označme

$$
\begin{aligned}
S_1 &:= \sum_{x=1}^{m} y(x) \doteq \sum_{i=1}^{m} y_i;\\
S_2 &:= \sum_{x=m+1}^{2m} y(x) \doteq \sum_{i=m+1}^{2m} y_i;\\
S_3 &:= \sum_{x=2m+1}^{3m} y(x) \doteq \sum_{i=2m+1}^{3m} y_i.
\end{aligned}
$$

Hodnoty $y(x)$ přesně neznáme, ale můžeme je přibližně aproximovat pomocí souřadnic y_i zadaných bodů $[1, y_1]$, $[2, y_2]$, $\ldots$, $[n, y_n]$ (v této matematické metodě výjimečně číslujeme indexy od hodnoty 1).

Využijme dále toho, že $y(i) = k+l\cdot a^i$ – nyní můžeme S_1, S_2, S_3 přepsat do tvaru

$$S_1 = mk + l \cdot \sum_{x=1}^{m} a^i = mk + l \cdot a \cdot \frac{a^m - 1}{a - 1}; \quad (1)$$

$$S_2 = mk + l \cdot \sum_{x=m+1}^{2m} a^i = mk + l \cdot a^{m+1} \cdot \frac{a^m - 1}{a - 1}; \quad (2)$$

$$S_3 = mk + l \cdot \sum_{x=2m+1}^{3m} a^i = mk + l \cdot a^{2m+1} \cdot \frac{a^m - 1}{a - 1}. \quad (3)$$

(sumy se sečetly podle vzorce pro součet m členů geometrické posloupnosti:

$$1 + a + a^2 + \cdots + a^{m-1} = \frac{a^m - 1}{a - 1})$$

Nyní odečtením 3 MINUS 2 máme

$$S_3 - S_2 = l \cdot \frac{a^m - 1}{a - 1} \cdot a^{m+1} \cdot (a^m - 1)\,,$$

a podobně odečtením 2 MINUS 1 máme

$$S_2 - S_1 = l \cdot \frac{a^m - 1}{a - 1} \cdot a \cdot (a^m - 1)\,.$$

Dohromady dělením obou rovnic máme (téměř vše se „potluče“!!!)

$$\frac{S_3 - S_2}{S_2 - S_1} = a^m \Longrightarrow a = \left(\frac{S_3 - S_2}{S_2 - S_1}\right)^{\frac{1}{n}}.$$

Dále např. z rovnice pro 2 MINUS 1 určíme

$$l = \frac{(S_2 - S_1)(a - 1)}{a(a^m - 1)^2}$$

a z rovnice 1 vypočteme poslední neznámou

$$k = \frac{1}{m} \cdot \left(S_1 - l \cdot a \cdot \frac{a^m - 1}{a - 1} \right).$$

Tato metoda, která hraním se se součty odvodila vzorce pro výpočet a, l, k (poslední tři vzorce v uvedeném pořadí), v našem příkladu prodeje CD dává překvapivě lepší výsledek, než MNČ s logaritmizací i než MNČ s vahami:

Pro $m = 3$ máme odhady $S_1 = 3 + 10 + 15 = 28$, $S_2 = 21 + 35 + 42 = 98$, $S_3 = 58 + 81 + 110 = 249$, odkud lze určit $a = 1{,}292061$, $l = 11{,}81813$, $k = -10{,}8313$ – to je výsledek ,který je včetně perfektně nejnižšího součtu čtverců uveden už na str. 32.

V rámci procvičení tohoto tématu můžete vypočíst ze skript [1], str. 83, příklady 6.8, 6.9, 6.10.

Literatura

[1] **Fajmon, B., Růžičková, I.: Matematika 3. Skriptum FEKT VUT v elektronické formě, Brno 2003. Počet stran 257 (identifikační číslo v informačním systému VUT: MAT103).** http://www.rozhovor.cz/souvislosti/matematika3.pdf.

[2] **Zapletal, J.: Úvod do analýzy ekonomických časových řad. Skriptum FEI VUT, PC-DIR Real, Brno 2000.**