FAKULTA MECHATRONIKY,
INFORMATIKY A MEZIOBOROVÝCH
STUDIÍ TUL

Aproximace funkcí, regrese

SEM2 – LS 2023/2024

Pavel Exner, Petr Rálek
NTI, FM TUL

pavel.exner@tul.cz, petr.ralek@tul.cz

[edit: 3. dubna 2024]

Aproximace – lineární regrese I

Předpokládejme, že jsou dány dvojice čísel $(x_1, y_1), \ldots, (x_n, y_n)$. Víme-li, že mezi veličinami x a y je lineární závislost, pak řešíme úlohu nalezení čísel $k, q \in \mathbb{R}$ tak, aby zadané dvojice čísel ležely co nejblíže přímky $y = kx + q$. V ideálním případě tedy k, q řeší soustavu

$$\begin{aligned} y_1 &= kx_1 + q \\ &\vdots \\ y_n &= kx_n + q, \end{aligned} \qquad \text{neboli} \qquad \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} k \\ q \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Máme však více rovnic (n) než neznámých (2). Můžeme se na problém dívat geometricky: Chceme, aby vzdálenost bodů $[x_i, y_i]$ od přímky $y = kx + q$ byla co nejmenší:

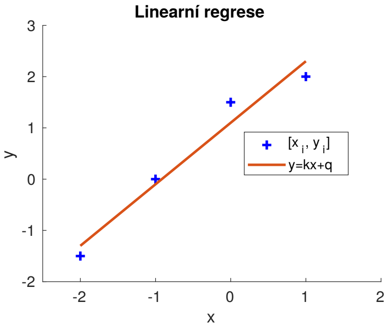
$$\arg\min S(k, q) = ?, \qquad S(k, q) = \sum_{i=1}^{n} (kx_i + q - y_i)^2$$

Aproximace – lineární regrese II

Hledáme tedy minimum funkce $S(k, q)$:

$$0 = \frac{\partial S}{\partial k} = 2\sum_{i=1}^{n}(kx_i + q - y_i)x_i,$$

$$0 = \frac{\partial S}{\partial q} = 2\sum_{i=1}^{n}(kx_i + q - y_i),$$



což nám dává soustavu 2 rovnic pro k, q :

$$k = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}, \qquad q = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2}.$$

Pozn: Minimalizace součtu druhých mocnin dává název Úloze nejmenších čtverců.

Úlohy nejmenších čtverců I

Označme soustavu $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, kde $\boldsymbol{x} = [k, q]$ byl vektor hledaných parametrů přímky.
Obecně takové soustavy nemají jednoznačné řešení.
Pro matici $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ a pravou stranu $\boldsymbol{b} \in \mathbb{R}^n$ budeme hledat vektor $\boldsymbol{x} \in \mathbb{R}^m$ tak, aby

$$\boldsymbol{A}\boldsymbol{x} \approx \boldsymbol{b}$$

platilo v jistém smyslu.
Je-li $m < n$, pak má soustava více rovnic než neznámých a nazývá se **přeurčená**.
Pro takové soustavy často neexistuje žádné řešení. V případě, že $m > n$, říkáme, že soustava je **nedourčená**, a obvykle má nekonečně mnoho řešení.
Cílem metody nejmenších čtverců je nalézt vektor $\boldsymbol{x}$, který minimalizuje součet kvadrátů odchylek:

$$\sum_{i=1}^{n} (b_i - (\boldsymbol{A}\boldsymbol{x})_i)^2 = (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \cdot (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) = (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})^2$$

Úlohy nejmenších čtverců II

$$\boxed{\arg\min_{\boldsymbol{x}} (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})^2 = ?}$$

Nutná podmínka na minimum (nulová derivace podle $\boldsymbol{x}$) vede na soustavu rovnic

$$\boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{0}.$$

Motivací pro úlohy nejmenších čtverců může být také jejich fyzikální význam. Pokud $\boldsymbol{b}$ reprezentuje naměřená nebo jinak získaná data, pak téměř jistě tato data obsahují nějakou chybu. Cílem pak je nalézt co nejmenší změnu $\boldsymbol{f}$ pravé strany (reprezentující chybu v datech) tak, aby $\boldsymbol{x}$ bylo řešením soustavy

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} + \boldsymbol{f}.$$

Požadavek minimalizovat normu $\|\boldsymbol{f}\|$ pak vede k definici následující úlohy.

Úlohy nejmenších čtverců III

Definition

Nechť $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ a $\boldsymbol{b} \in \mathbb{R}^{n}$. Problém nejmenších čtverců (LS) je úloha nalézt $\boldsymbol{x} \in \mathbb{R}^{m}$ takové, aby byla minimální

$$\|\boldsymbol{f}\| \text{ za podmínky } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} + \boldsymbol{f}.$$

Zkratka LS pochází z anglického **least squares** — minimalizuje se zde euklidovská norma, tj. odmocnina ze součtu kvadrátů prvků rezidua $\boldsymbol{f} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$. Poznamenejme ještě, že lze také uvažovat tzv. **úplný problém nejmenších čtverců**, kde předpokládáme, že chyba je obsažena také v koeficientech matice $\boldsymbol{A}$ a hledáme tedy korekce $\boldsymbol{E}$ a $\boldsymbol{f}$ s minimálními normami tak, aby

$$(\boldsymbol{A} + \boldsymbol{E})\boldsymbol{x} = \boldsymbol{b} + \boldsymbol{f}.$$

Úlohy nejmenších čtverců IV

Pro řešení problému LS se často používá tzv. soustava normálních rovnic.

Věta

Nechť $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ *a* $\boldsymbol{b} \in \mathbb{R}^{n}$. *Pak* $\boldsymbol{x} \in \mathbb{R}^{m}$ *je řešení problému LS právě tehdy, když je řešením soustavy normálních rovnic*

$$\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^\top \boldsymbol{b}.$$

Pozn.: Má-li $\boldsymbol{A}$ plnou sloupcovou hodnost, pak je matice $\boldsymbol{A}^\top \boldsymbol{A}$ regulární a platí:

$$\boldsymbol{x} = (\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top \boldsymbol{b} = \boldsymbol{A}^\dagger \boldsymbol{b},$$

kde $\boldsymbol{A}^\dagger$ je tzv. pseudoinverze $\boldsymbol{A}$. Vztah $\boldsymbol{x} = \boldsymbol{A}^\dagger \boldsymbol{b}$ má platnost i tehdy, když $\boldsymbol{A}$ nemá plnou sloupcovou hodnost.

Pseudoinverze – příklad I

levá pseudoinverze ($m < n$):

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 3 \end{bmatrix}$$

$$\boldsymbol{A}^\top \boldsymbol{A} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \qquad (\boldsymbol{A}^\top \boldsymbol{A})^{-1} = \begin{bmatrix} 2.\overline{3} & -1 \\ -1 & 0.5 \end{bmatrix}$$

$$(\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top = \begin{bmatrix} 0.\overline{3} & 1.\overline{3} & -0.\overline{6} \\ 0 & -0.5 & 0.5 \end{bmatrix} = \boldsymbol{A}^\dagger$$

$$\boldsymbol{A}^\dagger \boldsymbol{A} = \begin{bmatrix} 0.\overline{3} & 1.\overline{3} & -0.\overline{6} \\ 0 & -0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \boldsymbol{I}, \qquad \boldsymbol{A}\boldsymbol{A}^\dagger \neq \boldsymbol{I}$$

Pseudoinverze – příklad II

pravá pseudoinverze ($m > n$):

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 5 & 7 & 7 & 9 \end{bmatrix}$$

$$\boldsymbol{A}\boldsymbol{A}^\top = \begin{bmatrix} 4 & 28 \\ 28 & 204 \end{bmatrix} \qquad (\boldsymbol{A}\boldsymbol{A}^\top)^{-1} = \begin{bmatrix} 6.375 & -0.875 \\ -0.875 & 0.125 \end{bmatrix}$$

$$\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1} = \begin{bmatrix} 2 & -0.25 \\ 0.25 & 0 \\ 0.25 & 0 \\ -1.5 & 0.25 \end{bmatrix} = \boldsymbol{A}^\dagger$$

$$\boldsymbol{A}\boldsymbol{A}^\dagger = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 5 & 7 & 7 & 9 \end{bmatrix} \begin{bmatrix} 2 & -0.25 \\ 0.25 & 0 \\ 0.25 & 0 \\ -1.5 & 0.25 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \boldsymbol{I}, \qquad \boldsymbol{A}^\dagger\boldsymbol{A} \neq \boldsymbol{I}$$

Lineární regrese jako LS úloha I

Připomeňme úlohu lineární regrese s daty $[x_i, y_i]$ a regresní přímkou $y = kx + q$:

$$\begin{aligned} y_1 &= kx_1 + q \\ &\vdots \\ y_n &= kx_n + q, \end{aligned} \qquad \text{neboli} \qquad \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} k \\ q \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

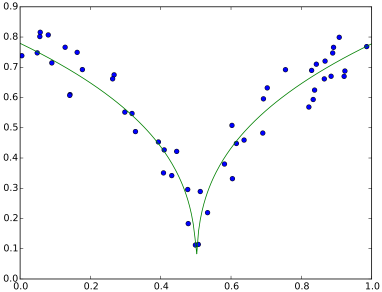
Tuto **přeurčenou** soustavu lze řešit ve smyslu nejmenších čtverců, tj. pomocí soustavy normálních rovnic ve tvaru

$$\begin{pmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{pmatrix} \begin{pmatrix} k \\ q \end{pmatrix} = \begin{pmatrix} \sum x_i y_i \\ \sum y_i \end{pmatrix}$$

(ve všech sumách sčítáme přes $i = 1, \ldots, n$).

Srovnejte se vztahy odvozenými dříve pomocí derivací.

Nelineární metoda nejmenších čtverců (NLS) I



Obrázek: Nelineární metoda nejmenších čtverců, $f(x, \beta_1, \beta_2) = |x + \beta_1|^{\beta_2}$.

Nelineární metoda nejmenších čtverců (NLS) II

Metoda LS předpokládá lineární vztah mezi daty $\{x_i\}$ a $\{y_i\}$. Toto omezení lze odstranit — hledáme pak obecně nelineární funkci $f(x, \boldsymbol{\beta})$, která má opět minimalizovat součet kvadrátů odchylek od zadaných dat:

$$\sum_{i=1}^{n} (y_i - f(x_i, \boldsymbol{\beta}))^2.$$

Vektor $\boldsymbol{\beta}$ zde představuje neznámé parametry, na nichž funkce f může záviset komplikovaným (nelineárním) způsobem.
Nelineární metoda LS úlohu nalezení optimálního $\boldsymbol{\beta}$ převádí na iterační proces, při kterém se postupně zpřesňuje počáteční odhad vektoru.

Nelineární metoda nejmenších čtverců (NLS) III

V každém kroce tohoto procesu se funkce f nahradí funkcí, která na $\boldsymbol{\beta}$ závisí lineárně: Je-li dáno $\boldsymbol{\beta}$, pak lze provést následující aproximaci:

$$f(x,\boldsymbol{\beta}) \approx f(x,\overline{\boldsymbol{\beta}}) + \sum_j \frac{\partial f}{\partial \beta_j}(x,\boldsymbol{\beta}) \cdot (\beta_j - \overline{\beta}_j),$$

kde výraz napravo od symbolu "$\approx$" je lineární vzhledem k prvkům vektoru $\boldsymbol{\beta}$. Iterační proces pak můžeme definovat následovně:

1. Zvolíme $\boldsymbol{\beta}_0$.
2. Použijeme LS na linearizovanou funkci (s $\boldsymbol{\beta} := \boldsymbol{\beta}_0$).
3. Získáme tak nový vektor $\boldsymbol{\beta}_1$.
4. Opět použijeme LS, tentokrát na funkci linearizovanou v bodě $\boldsymbol{\beta} := \boldsymbol{\beta}_1$.
5. Opakujeme postup do té doby, než se $\boldsymbol{\beta}$ ustálí.

Nelineární metoda nejmenších čtverců (NLS) IV

Poznamenejme, že úspěšnost procesu závisí na tom, jak dobře bylo zvoleno $\boldsymbol{\beta}_0$. Pokud by totiž $\boldsymbol{\beta}_0$ bylo příliš daleko od optimální hodnoty, může se stát, že cyklus nezkonverguje.