

Optimalizace

Aplikace metody nejmenších čtverců

Tomáš Kroupa Tomáš Werner

2025

Fakulta elektrotechnická
ČVUT v Praze

Jak na úlohu nejmenších čtverců s LN sloupci

Pro každou matici $\mathbf{A} \in \mathbb{R}^{m \times n}$ a vektor $\mathbf{b} \in \mathbb{R}^m$ platí

$$\arg \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{Ax} - \mathbf{b}\|^2 = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}\} \neq \emptyset.$$

- Má-li $\mathbf{A}$ LN sloupce, pak je $\mathbf{A}^T \mathbf{A}$ regulární a jediné řešení $\mathbf{x}^*$ nalezneme pomocí levé pseudoinverze $\mathbf{A}^+$,

$$\mathbf{x}^* = \underbrace{(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T}_{\mathbf{A}^+} \mathbf{b}.$$

- Má-li $\mathbf{A}$ LZ sloupce, soustava normálních rovnic má nekonečně mnoho řešení.

Jak na úlohu nejmenších čtverců s LZ sloupci

1. **Regularizace** – pro nějaké $\mu > 0$ řešíme regularizovanou úlohu

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{Ax} - \mathbf{b}\|^2 + \mu \|\mathbf{x}\|^2,$$

jejímž jediným optimálním řešením je

$$\mathbf{x}^* = (\mathbf{A}^T \mathbf{A} + \mu \mathbf{I})^{-1} \mathbf{A}^T \mathbf{b}$$

2. **Řešení s nejmenší normou** vede na úlohu

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \|\mathbf{x}\|^2 \mid \mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b} \right\},$$

jejímž jediným optimálním řešením je $\mathbf{x}^* = (\mathbf{A}^T \mathbf{A})^+ \mathbf{A}^T \mathbf{b}$,
kde $(\mathbf{A}^T \mathbf{A})^+$ je Moore-Penrosova pseudoinverze

Numerický výpočet řešení úlohy nejmenší čtverců

- Je žádoucí vyhnout se přímému výpočtu matice $\mathbf{A}^T \mathbf{A}$
- Numericky stabilní algoritmy využívají QR rozklad matice $\mathbf{A}$
- V Julii nebo Matlabu se použije příkaz $\mathbf{A} \backslash \mathbf{b}$

Příklad (Läuchli, 1961)

Pro $\varepsilon > 0$ má matice $\mathbf{A}$ LN sloupce:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ \varepsilon & 0 \\ 0 & \varepsilon \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ \varepsilon \\ \varepsilon \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 + \varepsilon^2 & 1 \\ 1 & 1 + \varepsilon^2 \end{bmatrix}$$

Pro $\varepsilon = 10^{-8}$ je $\mathbf{A}^T \mathbf{A}$ numericky singulární.

Příklad – nákup reklamy (1)

- Na 7 skupin obyvatel cílí 3 placené reklamní kanály
- Složka a_{ij} matice $\mathbf{A} \in \mathbb{R}^{7 \times 3}$ je počet shlédnutí skupinou i kanálu j na jednotku nákladů
- Cílový počet shlédnutí obsahu skupinami je $\mathbf{b} \in \mathbb{R}^7$
- Hledáme rozdělení nákladů $\mathbf{x} \in \mathbb{R}^3$ na kanály tak, aby bylo přibližně dosaženo cílového počtu shlédnutí, tedy

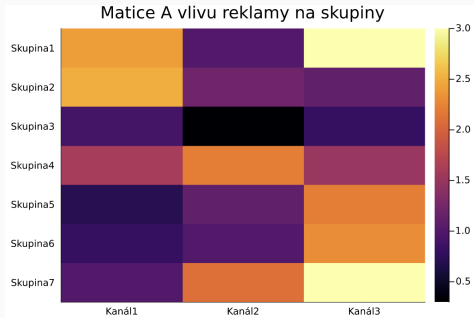
$$\mathbf{Ax} \approx \mathbf{b}$$

Úloha nejmenších čtverců

$$\min_{\mathbf{x} \in \mathbb{R}^3} \|\mathbf{Ax} - \mathbf{b}\|^2$$

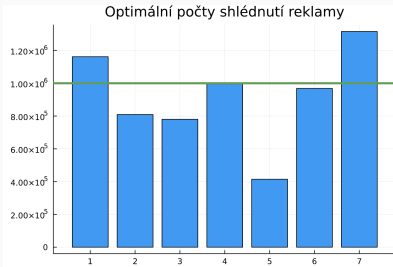
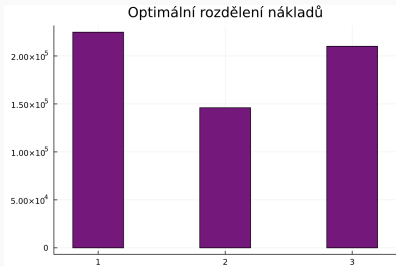
Omezení $x_1, x_2, x_3 \geq 0$ ani limit na náklady zde neuvažujeme.

Příklad – nákup reklamy (2)



Matice **A** má LN sloupce,
rank **A** = 3, optimální
řešení je tedy tvaru

$$\mathbf{x}^* = \mathbf{A}^+ \mathbf{b}.$$



Příklad – transformace obrazu (1)

Jak spojit dvě fotky dohromady?



Příklad – transformace obrazu (2)

Předpoklady

- Jsou dány referenční dvojice bodů $(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_m, \mathbf{y}_m)$ v obrázcích, kde $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^2$
- Transformace je afinní zobrazení $\mathbf{x} \in \mathbb{R}^2 \mapsto t(\mathbf{x}, \mathbf{p}) \in \mathbb{R}^2$ s parametry $\mathbf{p} = (p_1, \dots, p_6) \in \mathbb{R}^6$, kde

$$t(\mathbf{x}, \mathbf{p}) = \begin{bmatrix} p_1 & p_2 \\ p_4 & p_5 \end{bmatrix} \mathbf{x} + \begin{bmatrix} p_3 \\ p_6 \end{bmatrix}$$

Chceme řešit úlohu

$$\min_{\mathbf{p} \in \mathbb{R}^6} \sum_{i=1}^m \|t(\mathbf{x}_i, \mathbf{p}) - \mathbf{y}_i\|^2,$$

ale nevíme, zda jde o úlohu nejmenších čtverců.

Příklad – transformace obrazu (3)

Řešení

- Můžeme psát $t(\mathbf{x}_i, \mathbf{p}) = \mathbf{A}_i \mathbf{p} = \begin{bmatrix} x_{i1} & x_{i2} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{i1} & x_{i2} & 1 \end{bmatrix} \mathbf{p}$
a potom platí

$$\sum_{i=1}^m \|t(\mathbf{x}_i, \mathbf{p}) - \mathbf{y}_i\|^2 = \sum_{i=1}^m \|\mathbf{A}_i \mathbf{p} - \mathbf{y}_i\|^2$$

- Poslední součet je roven $\|\mathbf{A} \mathbf{p} - \mathbf{y}\|^2$, pokud označíme

$$\mathbf{A} := \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_m \end{bmatrix} \in \mathbb{R}^{2m \times 6}, \quad \mathbf{y} := \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_m \end{bmatrix} \in \mathbb{R}^{2m \times 1}$$

Regresní model

Závislost proměnné y na x modelujeme funkcí f tak, aby

$$y \approx f(x, \boldsymbol{\theta}),$$

kde $\boldsymbol{\theta} \in \mathbb{R}^n$ je vektor parametrů.

- Na základě dat $(x_1, y_1), \dots, (x_m, y_m)$ řešíme úlohu:

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^n} \sum_{i=1}^m (y_i - f(x_i, \boldsymbol{\theta}))^2$$

- **Reziduum** je hodnota $r_i := y_i - f(x_i, \boldsymbol{\theta})$

- Kvalitu modelu hodnotí např. **RMS kritérium** $\sqrt{\frac{\sum_{i=1}^m r_i^2}{m}}$

Volíme **bázové funkce** $\varphi_1, \dots, \varphi_n$ a předpokládáme

$$f(x, \boldsymbol{\theta}) = \theta_1 \varphi_1(x) + \dots + \theta_n \varphi_n(x) = \boldsymbol{\varphi}(x)^\top \boldsymbol{\theta}.$$

Úloha nejmenších čtverců

Píšeme

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \dots & \varphi_n(x_1) \\ \vdots & \vdots & & \vdots \\ \varphi_1(x_m) & \varphi_2(x_m) & \dots & \varphi_n(x_m) \end{bmatrix}, \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

a řešíme úlohu

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^n} \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|^2.$$

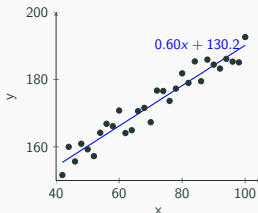
Příklad – prokládáme body přímkou

- Máme m měření $(x_i, y_i) \in \mathbb{R}^2$ váhy a výšky
- Vztah mezi proměnnými vyjadřuje funkce $f(x, \theta) = \theta_1 + \theta_2 x$
- Metoda nejmenších čtverců:

$$\min_{\theta \in \mathbb{R}^2} \|\mathbf{y} - \mathbf{A}\theta\|^2, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}$$

Protože existují alespoň dvě různá měření váhy ($x_i \neq x_j$), tak má matice $\mathbf{A}$ LN sloupce a existuje jediné řešení $\theta^* = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}$.

$$\theta_1^* = 130.2, \quad \theta_2^* = 0.6$$



Příklad – doprava (1)

- Úsek délky 130 m mezi semaforey projede automobil za

$$y = \theta_1 + \frac{\theta_2 x}{1 - \frac{x}{30}}$$

sekund, kde x je počet všech aut v úseku a 30 je kapacita úseku, θ_1, θ_2 jsou parametry

- Máme $m = 5$ měření:

x_i	20	18	22	21	15
y_i	14	10	16	18	8

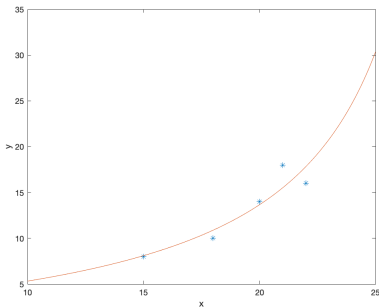
- Odhadujeme dobu průjezdu úseku, pokud je v něm 12 aut

Příklad – doprava (2)

- Lineární regrese $y = \theta_1 + \frac{\theta_2 x}{1 - \frac{x}{30}}$,

kde $\varphi_1(x) = 1$, $\varphi_2(x) = \frac{x}{1 - \frac{x}{30}}$, $\mathbf{y} \in \mathbb{R}^5$ a $\mathbf{A} = \begin{bmatrix} 1 & \varphi_2(x_1) \\ \vdots & \vdots \\ 1 & \varphi_2(x_5) \end{bmatrix}$

- Jelikož $\mathbf{A}$ má LN sloupce, jediné řešení je $\theta^* = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}$
a dostaneme tak $\theta_1^* = 2.546$, $\theta_2^* = 0.185$



Pro $x = 12$ aut dostaneme dobu průjezdu $y = 6.252$ s.

Regrese s polynomy (1) – přesná interpolace

- Funkce f je polynom stupně $n - 1$ proměnné x ,

$$f(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2 + \dots + \theta_n x^{n-1}$$

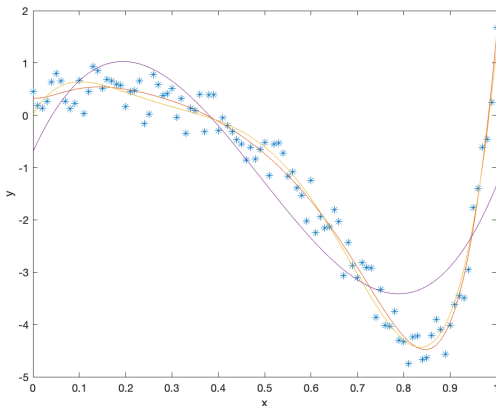
- Vandermondova matice $\mathbf{A} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ & & & \ddots & \\ 1 & x_m & x_m^2 & \dots & x_m^{n-1} \end{bmatrix}$

Interpolace polynomem

Nechť jsou všechny hodnoty $x_1, \dots, x_m$ různé. Potom má Vandermondova matice LN sloupce a existuje jediný polynom $f(x, \theta^*)$ stupně $\leq m - 1$ splňující $y_i = f(x_i, \theta^*)$, $i = 1, \dots, m$.

Regrese s polynomy (2) – úloha nejmenších čtverců

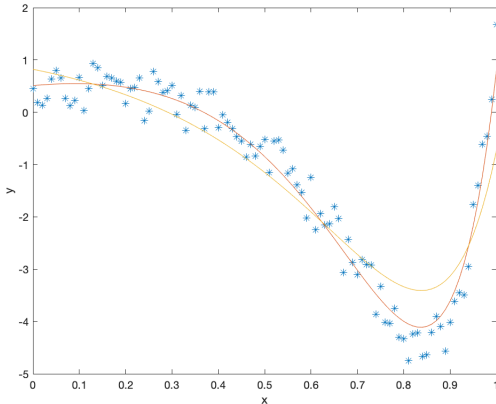
- Interpolace vede na polynomy velmi vysokého stupně
- Preferujeme jednoduché modely s vlastností generalizace
- Pro $m = 101$ měření (x_i, y_i) porovnáme 3 různé polynomy:



n	$\sum_{i=1}^m r_i^2$	$\frac{\ \theta^*\ }{\sqrt{n}}$
4	57	39
6	8.93	291
11	7.76	12 060

Regrese s polynomy (3) – hřebenová (ridge) regrese

- ℓ_2 -regularizace: $\min_{\boldsymbol{\theta} \in \mathbb{R}^n} \|\mathbf{A}\boldsymbol{\theta} - \mathbf{y}\|^2 + \mu \|\boldsymbol{\theta}\|^2$ pro nějaké $\mu > 0$
- Tato úloha má vždy jediné řešení $\boldsymbol{\theta}^* = (\mathbf{A}^\top \mathbf{A} + \mu \mathbf{I})^{-1} \mathbf{A}^\top \mathbf{y}$
- Pro polynom stupně 10 ($n=11$) dostaneme:



μ	$\sum_{i=1}^m r_i^2$	$\ \boldsymbol{\theta}^*\ $
0.1	11.29	10.84
1	39.7	5.91

Další využití ℓ_2 -regularizace

$$\mathbf{A} = \begin{bmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \dots & \varphi_n(x_1) \\ & & \vdots & \\ \varphi_1(x_m) & \varphi_2(x_m) & \dots & \varphi_n(x_m) \end{bmatrix}$$

Sloupce matice $\mathbf{A}$ nejsou lineárně nezávislé:

- lineární závislost/korelace mezi různými φ_i
- více prediktorů než dat, $m < n$

Analytické řešení ℓ_2 -regularizované úlohy NČ

Soustava normálních rovnic $\mathbf{A}^\top \mathbf{A} \boldsymbol{\theta} = \mathbf{A}^\top \mathbf{y}$ má více řešení, ale soustava $(\mathbf{A}^\top \mathbf{A} + \mu \mathbf{I}) \boldsymbol{\theta} = \mathbf{A}^\top \mathbf{y}$ už má jediné řešení díky regularitě matice $\mathbf{A}^\top \mathbf{A} + \mu \mathbf{I}$ pro libovolné $\mu > 0$.

Úloha nejmenších čtverců v kontextu

- Jedná se o úlohu **konvexní optimalizace**
- Pro velká m se používají numerické metody s garancí konvergence k minimu kvadratické funkce $\|\mathbf{Ax} - \mathbf{b}\|^2$
- ℓ_1 -regularizace také zvaná **LASSO** (least absolute shrinkage and selection operator) preferuje řídka řešení:

$$\min_{\mathbf{x} \in \mathbb{R}^n} (\|\mathbf{Ax} - \mathbf{b}\|^2 + \mu \|\mathbf{x}\|_1)$$