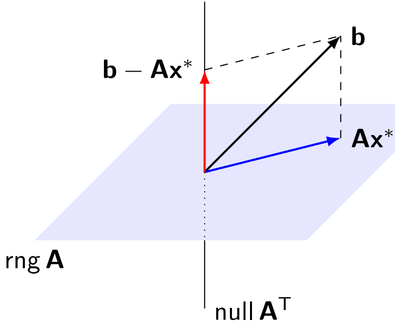
Optimalizace

Aplikace metody nejmenších čtverců

Tomáš Kroupa Tomáš Werner

2024

Fakulta elektrotechnická
ČVUT v Praze



Tyto výroky jsou ekvivaletní pro vektor $\mathbf{x}^*$.

1. $\|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|^2 = \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$.
2. $\mathbf{A}^T\mathbf{A}\mathbf{x}^* = \mathbf{A}^T\mathbf{b}$.
3. $\mathbf{A}\mathbf{x}^* = \mathbf{P}\mathbf{b}$, kde $\mathbf{P}$ je ortogonální projektor na rng $\mathbf{A}$.

- Podmínka 2. je podle obrázku ekvivalentní $\mathbf{b} - \mathbf{A}\mathbf{x}^* \in \operatorname{null} \mathbf{A}^T$.
- Podmínka 3. říká, že nejbližší vektor v rng $\mathbf{A}$ je ortogonální projekcí vektoru $\mathbf{b}$.

Jak na úlohu nejmenších čtverců

Optimální řešení úlohy

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2$$

je řešením soustavy normálních rovnic $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$.

- Má-li $\mathbf{A}$ LN sloupce, pak jediné řešení je

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

- Obecně může mít úloha nekonečně mnoho řešení a my volíme třeba řešení s nejmenší normou, tedy

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

Numerický výpočet řešení úlohy nejmenší čtverců

- Je žádoucí vyhnout se přímému výpočtu matice $\mathbf{A}^\mathsf{T}\mathbf{A}$ a následnému řešení soustavy normálních rovnic
- Numericky stabilní algoritmy využívají QR rozklad matice $\mathbf{A}$
- V implementacích se použije operátor `A\b`

Příklad (Läuchli, 1961)

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

Pro $\varepsilon = 10^{-8}$ je matice $\mathbf{A}^\mathsf{T}\mathbf{A}$ numericky singulární. V Julii dává příkaz `A\b` řešení $(0.9999999999999998, 1.0000000000000002)$.

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

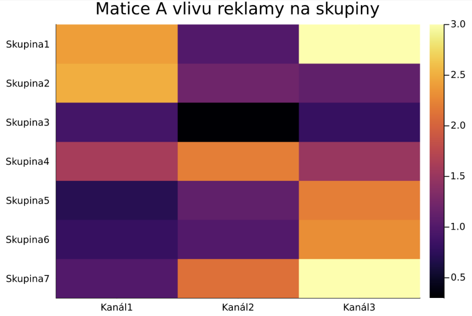
$$\mathbf{Ax} \approx \mathbf{b}$$

Úloha nejmenších čtverců

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

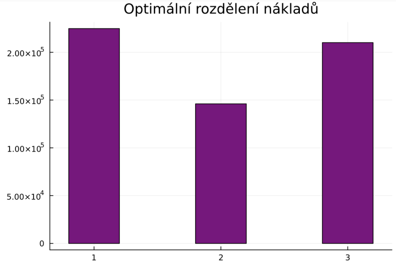
Omezení $x_1, x_2, x_3 \geq 0$ ani limit na náklady zde neuvažujeme.

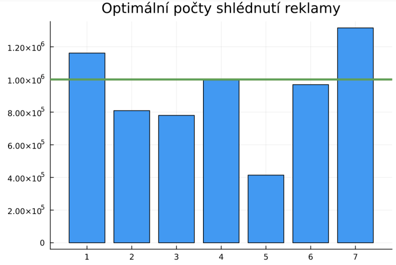
Příklad – nákup reklamy (2)



Matice **A** má LN sloupce, rank $\mathbf{A} = 3$, 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 $(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_m, \mathbf{y}_m)$ v obrázcích, kde $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^2$
- Afinní zobrazení $t(\mathbf{z}, \mathbf{p}) = \begin{bmatrix} p_1 & p_2 \\ p_4 & p_5 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} p_3 \\ p_6 \end{bmatrix}$ s parametry $\mathbf{p} = (p_1, \ldots, p_6) \in \mathbb{R}^6$
- Chceme řešit úlohu

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

Jde o úlohu nejmenších čtverců?

Příklad – transformace obrazu (3)

Řešení

- Můžeme psát $t(\mathbf{z}, \mathbf{p}) = \begin{bmatrix} z_1 & z_2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & z_1 & z_2 & 1 \end{bmatrix} \mathbf{p}$
- Potom dostaneme úlohu nejmenších čtverců, protože

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

kde

$$\mathbf{A} := \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_m \end{bmatrix} \quad \mathbf{A}_i := \begin{bmatrix} x_{i1} & x_{i2} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{i1} & x_{i2} & 1 \end{bmatrix} \quad \mathbf{y} := \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_m \end{bmatrix}$$

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 $(x_1, y_1), \ldots, (x_m, y_m)$ 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 $\sqrt{\frac{\sum_{i=1}^{m} r_i^2}{m}}$

Lineární regrese

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

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

Metoda nejmenších čtverců

Píšeme

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

a řešíme úlohu

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

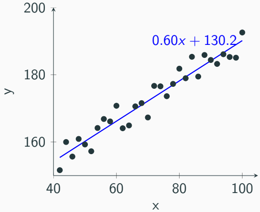
Příklad – prokládáme body přímkou

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

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

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

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



Jednoduchá regrese $f(x, \theta) = \theta$, tedy odhad hodnot $y_1, \ldots, y_m$ konstantou $\theta \in \mathbb{R}$. Jedná se o model, kde regresní přímka je rovnoběžná s osou x. Platí

$$\mathbf{A} = \begin{bmatrix} 1 \\ \ldots \\ 1 \end{bmatrix} = \mathbf{1} \in \mathbb{R}^{m \times 1}.$$

Dostaneme aritmetický průměr

$$\theta^* = (\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T \mathbf{y} = m^{-1} \sum_{i=1}^{m} y_i = \frac{y_1 + \cdots + y_m}{m} = \bar{\mathbf{y}}.$$

Optimální hodnota je

$$\|\mathbf{y} - \theta^* \mathbf{1}\|^2 = \sum_{i=1}^{m} (y_i - \theta^*)^2 = \sum_{i=1}^{m} (y_i - \bar{\mathbf{y}})^2$$

což je až na konstantu $\frac{1}{m}$ výběrový rozptyl.

Příklad – doprava (1)

- Úsek délky 130 m mezi semafory 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

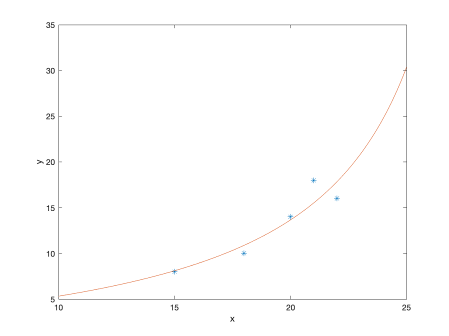
- Odhadujeme dobu průjezdu úseku, pokud je v něm 12 aut

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 $\boldsymbol{\theta}^* = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$ 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,

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

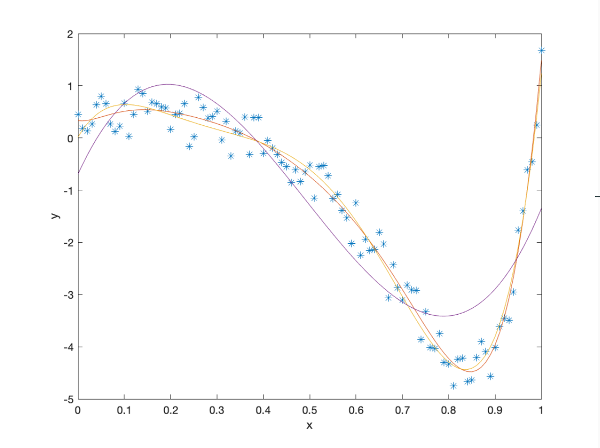
- Vandermondova matice $\mathbf{A} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ & & & \vdots & \\ 1 & x_m & x_m^2 & \dots & x_m^{n-1} \end{bmatrix}$

Interpolace polynomem

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

Regrese s polynomy (2) – úloha nejmenších čtverců

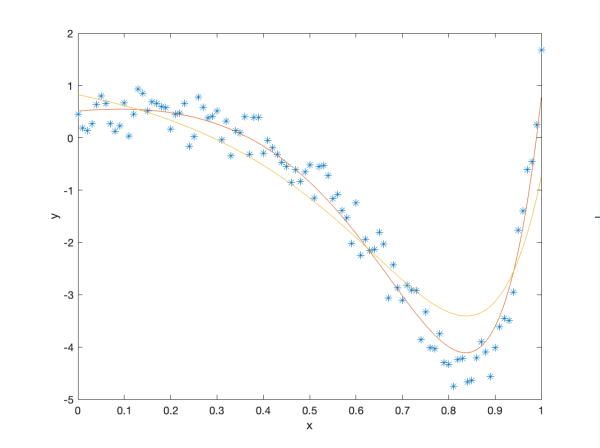
- Interpolace vede na polynomy velmi vysokého stupně
- Preferujeme jednoduché modely vystihující podstatné rysy dat
- Pro $m = 101$ měření (x_i, y_i) porovnáme 3 různé polynomy



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

Regrese s polynomy (3) – ridge regression

- Regularizace zmenší model a zvýší jeho robustnost
- Minimalizujeme kritérium $\|\mathbf{A}\boldsymbol{\theta} - \mathbf{y}\|^2 + \mu\|\boldsymbol{\theta}\|^2$ pro nějaké $\mu > 0$
- Jediné řešení je $\boldsymbol{\theta}^* = (\mathbf{A}^\mathsf{T}\mathbf{A} + \mu\mathbf{I})^{-1}\mathbf{A}^\mathsf{T}\mathbf{y}$
- Pro polynom stupně 10 (n=11):



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

Další využití regularizace

- Většinou předpokládáme lineární nezávislost sloupců matice

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

- V případě $m < n$ tento předpoklad není splněn

Postřehy

- Jedna z funkcí φ_i je na zadaných datech redundantní
- Soustava normálních rovnic $\mathbf{A}^\top\mathbf{A}\boldsymbol{\theta} = \mathbf{A}^\top\mathbf{y}$ má více řešení
- Protože je matice $\mathbf{A}^\top\mathbf{A} + \mu\mathbf{I}$ invertibilní pro libovolné $\mu > 0$, můžeme použít regularizované řešení $\boldsymbol{\theta}^* = (\mathbf{A}^\top\mathbf{A} + \mu\mathbf{I})^{-1}\mathbf{A}^\top\mathbf{y}$

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 $\|\mathbf{Ax} - \mathbf{b}\|^2$
- Regularizace pomocí součtové normy $\|\cdot\|_1$ vede na LASSO

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

- LASSO model umožnuje hledat řídká řešení a slouží k výběru příznaků, ovšem účelová funkce není všude diferencovatelná