Optimalizace

3. Metoda nejmenších čtverců

Tom Werner

FEL ČVUT

Řešitelnost lineárních soustav

Soustava m lineárních rovnic o n neznámých ($\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$)

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

- pro $\mathbf{b} = \mathbf{0}$ je **homogenní**, pro $\mathbf{b} \neq \mathbf{0}$ je **nehomogenní**.
- má řešení, právě když $\mathbf{b} \in \operatorname{rng} \mathbf{A}$, tj. $\operatorname{rank}[\mathbf{A} \;\; \mathbf{b}] = \operatorname{rank} \mathbf{A}$ (Frobeniova věta)
- množina řešení je afinní podprostor $\mathbb{R}^n$

Tři případy:

- nemá řešení ($\mathbf{b} \notin \operatorname{rng} \mathbf{A}$): **přeurčená** soustava
- právě jedno řešení ($\mathbf{b} \in \operatorname{rng} \mathbf{A}$, $\operatorname{rank} \mathbf{A} = n$)
- nekonečně mnoho řešení ($\mathbf{b} \in \operatorname{rng} \mathbf{A}$, $\operatorname{rank} \mathbf{A} < n$): **nedourčená** soustava

Úloha nejmenších čtverců

Úloha nejmenších čtverců

Přibližné řešení přeurčené soustavy $\mathbf{A}\mathbf{x} = \mathbf{b}$: minimalizuj součet

$$\|\mathbf{r}\|^2 = \mathbf{r}^T\mathbf{r} = r_1^2 + \cdots + r_m^2$$

čtverců reziduí $(r_1, \ldots, r_m) = \mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$. Tedy

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

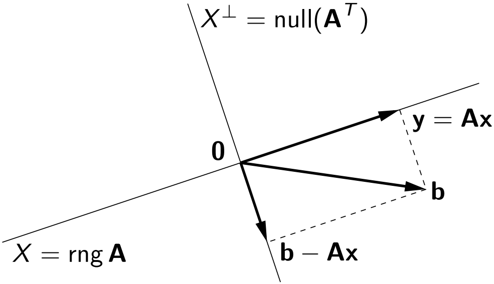
Příklad: Pro soustavu

$$\begin{aligned} x + 2y &= 6 \\ -x + y &= 3 \\ x + y &= 4 \end{aligned}$$

hledáme přibližné řešení $x, y \in \mathbb{R}$, které minimalizuje

$$(x + 2y - 6)^2 + (-x + y - 3)^2 + (x + y - 4)^2.$$

Řešení úlohy nejmenších čtverců



Věta

Vektor $\mathbf{x}$ je optimální řešení úlohy

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = \min_{\mathbf{y}\in\mathrm{rng}\,\mathbf{A}} \|\mathbf{y} - \mathbf{b}\|^2$$

právě když splňuje soustavu **normálních rovnic**

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}.$$

Důkaz: Víme, že $\mathbf{x}$ je optimální řešení právě když $(\mathbf{b} - \mathbf{A}\mathbf{x}) \perp \mathrm{rng}\,\mathbf{A}$, tj. $\mathbf{A}^T(\mathbf{b} - \mathbf{A}\mathbf{x}) = \mathbf{0}$.

Řešení pomocí derivací

Stejnou podmínku dostaneme z analýzy: minimalizujeme funkci

$$
\begin{aligned}
f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2 &= (\mathbf{Ax} - \mathbf{b})^T(\mathbf{Ax} - \mathbf{b}) \\
&= \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} - \mathbf{b}^T\mathbf{Ax} + \mathbf{b}^T\mathbf{b} \\
&= \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} - 2\mathbf{b}^T\mathbf{Ax} + \mathbf{b}^T\mathbf{b}
\end{aligned}
$$

- f je konvexní, kvadratická, zdola omezená funkce n proměnných
- Gradient f (vrátíme se k němu později):

$$\nabla f(\mathbf{x}) = 2\mathbf{A}^T\mathbf{Ax} - 2\mathbf{A}^T\mathbf{b}$$

- stacionární podmínka $\nabla f(\mathbf{x}) = \mathbf{0}$ jsou normální rovnice

Řešitelnost soustavy normálních rovnic

Věta

- $\mathrm{rng}(\mathbf{A}^T\mathbf{A}) = \mathrm{rng}(\mathbf{A}^T)$
- $\mathrm{null}(\mathbf{A}^T\mathbf{A}) = \mathrm{null}\,\mathbf{A}$

Důkaz druhé rovnosti: Zřejmě $\mathbf{Ax} = \mathbf{0} \;\Rightarrow\; \mathbf{A}^T\mathbf{Ax} = \mathbf{0}$. Ale také
$\mathbf{A}^T\mathbf{Ax} = \mathbf{0} \;\Rightarrow\; \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} = (\mathbf{Ax})^T(\mathbf{Ax}) = \|\mathbf{Ax}\|^2 = 0 \;\Rightarrow\; \mathbf{Ax} = \mathbf{0}$.
První rovnost pak plyne z druhé rovnosti a rank-nullity theoremu.

Důsledky pro řešitelnost soustavy $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$:

- Soustava má řešení pro každé $\mathbf{A}, \mathbf{b}$ (neboť vždy $\mathbf{A}^T\mathbf{b} \in \mathrm{rng}(\mathbf{A}^T) = \mathrm{rng}(\mathbf{A}^T\mathbf{A})$).
- Matice $\mathbf{A}^T\mathbf{A}$ je regulární, právě když $\mathbf{A}$ má lin. nezávislé sloupce.
  Pak řešení soustavy lze napsat pomocí inverze:
  $$\mathbf{x} = \underbrace{(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T}_{\mathbf{A}^+}\mathbf{b}$$
  Matice $\mathbf{A}^+$ se nazývá **pseudoinverze** matice $\mathbf{A}$ (s l.n. sloupci).
  Je to jedna z levých inverzí matice $\mathbf{A}$ (neboť $\mathbf{A}^+\mathbf{A} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{A} = \mathbf{I}$).

Řešení pomocí QR rozkladu

Řešíme soustavu normálních rovnic

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$$

kde sloupce $\mathbf{A}$ jsou l.n.

- řešení pomocí inverze matice $\mathbf{A}^T\mathbf{A}$ zanáší zaokrouhlovací chyby
- lépe pomocí (redukovaného) QR rozkladu $\mathbf{A} = \mathbf{QR}$:

$$\begin{aligned}
\mathbf{A}^T\mathbf{A}\mathbf{x} &= \mathbf{A}^T\mathbf{b} \\
\mathbf{R}^T\mathbf{Q}^T\mathbf{Q}\mathbf{R}\mathbf{x} &= \mathbf{R}^T\mathbf{Q}^T\mathbf{b} \\
\mathbf{R}^T\mathbf{R}\mathbf{x} &= \mathbf{R}^T\mathbf{Q}^T\mathbf{b} \\
\mathbf{R}\mathbf{x} &= \mathbf{Q}^T\mathbf{b}
\end{aligned}$$

V Matlabu: `x=A\b`

Ortogonální projekce na podprostor zadaný obecnou bází

Zopakujme úlohu nejmenších čtverců

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{Ax}-\mathbf{b}\|^2 = \min_{\mathbf{y}\in\operatorname{rng}\mathbf{A}} \|\mathbf{y}-\mathbf{b}\|^2$$

- Když jsou sloupce $\mathbf{A}$ lin. nezávislé, tvoří bázi podprostoru $X = \operatorname{rng}\mathbf{A}$.
- Pak optimální řešení

$$\mathbf{y} = \mathbf{Ax} = \mathbf{AA}^+\mathbf{b} = \underbrace{\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T}_{\mathbf{P}}\mathbf{b}$$

  je ortogonální projekce vektoru $\mathbf{b}$ na podprostor X
  (a matice $\mathbf{P}$ je ortogonální projektor).
- Když $\mathbf{A}$ má ortonormální sloupce ($\mathbf{A}^T\mathbf{A} = \mathbf{I}$), tak $\mathbf{P} = \mathbf{AA}^T$. To už známe!

Lineární regrese

Regrese je modelování funkční závislosti proměnné $y \in \mathbb{R}$ na proměnné $x \in X$ (kde X je libovolná množina) **regresní funkcí**

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

Odhad parametrů: Hledáme parametry $\boldsymbol{\theta} \in \mathbb{R}^n$ tak, aby funkce f dobře modelovala naměřená data $(x_1, y_1), \dots, (x_m, y_m)$:

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

- **Lineární regrese:** pro každé $x \in X$ je f lineární funkcí θ. Tedy

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

  kde $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_n)\colon X \to \mathbb{R}$ jsou dané "bázové funkce".
- Pak

  $$\sum_{i=1}^{m} (y_i - f(x_i, \boldsymbol{\theta}))^2 = \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|^2$$

  kde $\mathbf{y} = (y_1, \dots, y_m)$ a $\mathbf{A} = [a_{ij}] = [\varphi_j(x_i)]$.

Příklad: Aproximace polynomiální funkcí

Nechť $X = \mathbb{R}$ a $\varphi_j(x) = x^{j-1}$. Pak

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

a

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

(**Vandermondova matice**).

- Pro $n = 1$ je $f(x, \theta) = \theta$ a $f(x, \theta) = \sum_{i=1}^m (x_i - \theta)^2$. Řešení: $\theta^* = \frac{1}{m} \sum_{i=1}^m x_i$.
- Zobecnění na polynomy více proměnných: $X = \mathbb{R}^d$ a φ_j jsou monomy proměnných $x_1, \ldots, x_d$

Pozor! Pořád jde o lineární regresi, protože f je lineární v parametrech θ.

Statistický pohled na úlohu nejmenších čtverců

Lineární model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon}$$

- $\mathbf{x} = (x_1, \ldots, x_n)$ jsou odhadované (nepozorovatelné) parametry
- $\mathbf{y} = (y_1, \ldots, y_m)$ jsou naměřené (pozorované) parametry
- $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_m)$ jsou náhodné chyby (šum, chyby měření, neznalost modelu)

Předpoklady na náhodné chyby:

- nulové střední hodnoty: $\mathrm{E}[\epsilon_i] = 0$ pro každé i
- stejné variance: $\mathrm{E}[\epsilon_i^2] = \sigma^2$ pro každé i
- náhodné chyby jsou nekorelované: $\mathrm{E}[\epsilon_i \epsilon_j] = 0$ pro každé $i \neq j$

Neboli

$$\mathrm{E}[\boldsymbol{\epsilon}] = \mathbf{0}$$
$$\mathrm{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = \sigma^2 \mathbf{I}$$

Lineární estimátor

Lineární model:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon}$$

Lineární estimátor (‘odhadovač’): Odhad $\mathbf{x}$ jako lineární funkce $\mathbf{y}$:

$$\hat{\mathbf{x}} = \mathbf{B}\mathbf{y}$$

- Střední hodnota odhadu:

  $$\mathrm{E}[\hat{\mathbf{x}}] = \mathrm{E}[\mathbf{B}\mathbf{y}] = \mathrm{E}[\mathbf{B}(\mathbf{A}\mathbf{x} + \boldsymbol{\epsilon})] = \mathbf{B}\mathbf{A}\mathbf{x} + \mathbf{B}\mathrm{E}[\boldsymbol{\epsilon}] = \mathbf{B}\mathbf{A}\mathbf{x}$$

  Estimátor je **nevychýlený**, jestliže $\mathrm{E}[\hat{\mathbf{x}}] = \mathbf{x}$ pro všechna $\mathbf{x}$.
  To platí právě když $\mathbf{B}\mathbf{A} = \mathbf{I}$, tj. $\mathbf{B}$ je levá inverze $\mathbf{A}$.
  Potom $\hat{\mathbf{x}} = \mathbf{B}(\mathbf{A}\mathbf{x} + \boldsymbol{\epsilon}) = \mathbf{x} + \mathbf{B}\boldsymbol{\epsilon}$.

- Variance (přesněji: kovarianční matice) nevychýleného odhadu:

  $$\mathrm{E}[(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^T] = \mathrm{E}[\mathbf{B}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\mathbf{B}^T] = \mathbf{B}\mathrm{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T]\mathbf{B}^T = \sigma^2\mathbf{B}\mathbf{B}^T$$

Estimátor metodou nejm. čtverců (least-squares estimator)

Pro LS (least-squares) estimátor máme

$$\mathbf{B} = \mathbf{A}^{+} = (\mathbf{A}^{T}\mathbf{A})^{-1}\mathbf{A}^{T}$$

- LS estimátor je nevychýlený, neboť $\mathbf{A}^{+}\mathbf{A} = \mathbf{I}$.
- Variance odhadu:

  $$\mathrm{E}[(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^{T}] = \sigma^2 \mathbf{A}^{+}(\mathbf{A}^{+})^{T} = \sigma^2 (\mathbf{A}^{T}\mathbf{A})^{-1}\mathbf{A}^{T}\mathbf{A}(\mathbf{A}^{T}\mathbf{A})^{-1} = \sigma^2 (\mathbf{A}^{T}\mathbf{A})^{-1}$$

Příklad

Motivace: Máme m stejných kuliček a studnu. Chceme odhadnout čas x, jak dlouho padá kulička do studny.

Model pro $n = 1$, $\mathbf{A} = \mathbf{1}$:

$$\mathbf{y} = \mathbf{1}x + \boldsymbol{\epsilon} \qquad \text{neboli} \qquad y_i = x + \epsilon_i$$

- LS odhad je aritmetický průměr:

$$\hat{x} = \mathbf{A}^{+}\mathbf{y} = \frac{1}{m}\mathbf{1}^T\mathbf{y} = \frac{1}{m}\sum_{i=1}^{m} y_i$$

- variance:

$$\mathrm{E}[(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^T] = \sigma^2(\mathbf{A}^T\mathbf{A})^{-1} = \sigma^2(\mathbf{1}^T\mathbf{1})^{-1} = \frac{\sigma^2}{m}$$

  Intuice: Provedu-li m nepřesných měření veličiny x a výsledek zprůměruju, variance ('neurčitost') odhadu veličiny klesne m-krát.

Příklad

Motivace: Mám pružinu a předpokládám Hookeův zákon: síla F tahu pružiny závisí na délce d pružiny jako $F = kd$. Chci odhadnout k.

Zde je opět $n = 1$.

LS estimátor je BLUE (Gauss-Markov theorem)

Částečné uspořádání $\preceq$ pro čtvercové symetrické matice $n \times n$:

- $\mathbf{K} \preceq \mathbf{L}$ značí, že matice $\mathbf{K} - \mathbf{L}$ je positivně semidefinitní
- Tj. pro každé $\mathbf{z} \in \mathbb{R}^{n \times n}$ platí $\mathbf{z}^T \mathbf{K} \mathbf{z} \leq \mathbf{z}^T \mathbf{L} \mathbf{z}$.
- Tj. $\mathbf{K}$ je 'menší' než $\mathbf{L}$.
- Tj. náhodný vektor s variancí $\mathbf{K}$ je 'méně nerčitý' než náh. vektor s variancí $\mathbf{L}$.

Věta (Gauss-Markov)

Pro všechny matice $\mathbf{A}, \mathbf{B}$ splňující $\mathbf{B}\mathbf{A} = \mathbf{I}$ platí $\mathbf{A}^+ (\mathbf{A}^+)^T \preceq \mathbf{B}\mathbf{B}^T$.

Důkaz:

- Každý projektor $\mathbf{P}$ splňuje $\mathbf{0} \preceq \mathbf{P} \preceq \mathbf{I}$.
- Tedy $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \preceq \mathbf{I}$, tj. pro každé $\mathbf{y}$ je $\mathbf{y}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y} \leq \mathbf{y}^T\mathbf{y}$.
- Speciálně, pro $\mathbf{y} = \mathbf{B}^T\mathbf{z}$ je $\mathbf{z}^T\mathbf{B}\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{B}^T\mathbf{z} = \mathbf{z}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{z} \leq \mathbf{z}^T\mathbf{B}\mathbf{B}^T\mathbf{z}$.

Tedy: LS estimátor je **best linear unbiased estimator (BLUE)**.

Další vhled:

- $\mathbf{A}^+(\mathbf{A}^+)^T \preceq \mathbf{B}\mathbf{B}^T$ implikuje $\|\mathbf{A}^+\| \leq \|\mathbf{B}\|$ (kde $\|\mathbf{B}\|^2 = \operatorname{tr}(\mathbf{B}\mathbf{B}^T) = \sum_{i,j} b_{ij}^2$)
- Intuice: Pro menší $\|\mathbf{B}\|$ mají chyby ϵ menší vliv na odhad $\hat{\mathbf{x}} = \mathbf{x} + \mathbf{B}\epsilon$.

Řešení lineární soustavy s nejmenší normou

Řešení lineární soustavy s nejmenší normou

Mezi řešeními soustavy $\mathbf{Ax} = \mathbf{b}$ najdeme to s nejmenší normou:

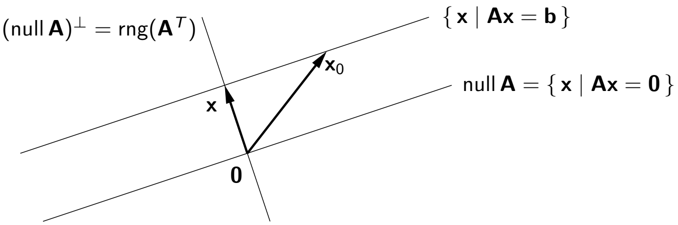
$$\min\{\, \|\mathbf{x}\|^2 \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{Ax} = \mathbf{b}\,\}$$

Příklad: Soustava

$$\begin{aligned} x + 2y + z &= 1 \\ -x + y + 2z &= 2 \end{aligned}$$

má nekonečně mnoho řešení. Hledáme řešení které minimalizuje číslo $x^2 + y^2 + z^2$.

Je-li $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$, pak $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b} \,\} = \mathbf{x}_0 + \operatorname{null} \mathbf{A}$.



Řešení $\mathbf{x}$ má nejmenší normu, právě když $\mathbf{x} \perp \operatorname{null} \mathbf{A}$, tj. $\mathbf{x} \in (\operatorname{null} \mathbf{A})^\perp = \operatorname{rng}(\mathbf{A}^T)$, tj. $\mathbf{x} = \mathbf{A}^T\mathbf{y}$ pro nějaké $\mathbf{y}$. Tedy řešíme soustavu

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$
$$\mathbf{A}^T\mathbf{y} = \mathbf{x}$$

Řešení: Dosazení dá $\mathbf{A}\mathbf{A}^T\mathbf{y} = \mathbf{b}$. Má-li $\mathbf{A}$ l.n. řádky, máme $\mathbf{y} = (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$. Tedy

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

kde $\mathbf{A}^+$ je **pseudoinverze** matice $\mathbf{A}$ (s l.n. řádky).

Je to jedna z pravých inverzí matice $\mathbf{A}$ (neboť $\mathbf{A}\mathbf{A}^+ = \mathbf{A}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} = \mathbf{I}$).

Aplikace úlohy nejmenších čtverců

Úloha

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

a příbuzné úlohy mají velmi mnoho aplikací. Mnoho jich je popsané v této knize (zdarma ke stažení, viz doplňující literatura na CW):