Přednáška 21 – problém nejmenších čtverců

Motivace: Maximum likelihood estimate (MLE)

předpokládejme, že máme model předpovídající jistý fenomén / fyzikální jev/děj → model v závislosti na vstupních datech a kalibraci vrací aproximaci/predikci výsledného stavu:

$$\underbrace{\vec{x}}_{\text{vstupní data}} \;\&\; \underbrace{\vec{\Theta}}_{\text{parametry modelu}} \rightsquigarrow \underbrace{F(\vec{x},\vec{\Theta})}_{\text{predikce modelu}} \approx \underbrace{\vec{y}}_{\text{realita}}$$

Výběr modelu ≡ problém pro odborníky dané oblasti

Volba parametrů (≡ kalibrace modelu) na základě dat ≡ problém matematiky/data-science

Např.: zajímá mě pohyb planet → modelem je pohyb po elipse (Keplerův zákon) ⇝ po které? → problém nalezení správných parametrů $\vec{\Theta}$.

- volba modelu (tj. funkce $\vec{F}(\vec{x},\vec{\theta})$) je u každého problému jiná, ale volbu $\vec{\theta}$ lze analyzovat relativně obecně

Zjednodušený set-up: předpokládejme, že máme „dokonalý model", tedy, $\exists \vec{\theta}$ takové, že pro libovolné přesné $\vec{x}$ existuje přesné $\vec{y}$ t. že: $\vec{y} = F(\vec{x}, \vec{\theta})$

$y_i \equiv F(x_i, \vec{\theta})$ zjevně v praxi téměř nedosažitelné

Ale i za tohoto předpokladu budou naše pozorování/měření/data nepřesná $\rightarrow$ pokud jsme pro inputy $x_1, \dots, x_m$ naměřili outputy $z_1, \dots, z_m$ pak bohužel $\delta z_i := z_i - F(x_i, \vec{\theta}) \not\approx 0$. Standardně modelujeme měřená data, jako náhodnou veličinu z normálního rozdělení vycentrovaného okolo y_i (y_i okolo „přesného měření") $\Longrightarrow Z_i \sim \mathcal{N}(y_i, \sigma_i^2)$ s nějakým rozptylem σ_i

často odhadnutelné podle typu/kvality měření.

Tudíž i ta chyba $\delta Z_i := Z_i - F(x_i, \vec{\theta})$ je náhodná veličina, zjevně $\delta Z_i \sim \mathcal{N}(0, \sigma_i^2)$, jejíž realizace jsou hodnoty $\delta z_1, \dots, \delta z_m$

$\rightarrow$ víme díky předpokladu, že y_i existují, ale zjevně je neznáme

v praxi δz_i samozřejmě nemáme, protože neznáme „přesné hodnoty" $y_i = F(x_i, \vec{\theta}) \rightarrow$ ale opět víme, že existují

Otázka: Máme-li $(x_1, z_1), \dots, (x_m, z_m)$, kde $Z_i \sim \mathcal{N}(y_i, \sigma_i^2)$ jako výše, jaké „musely být ty parametry $\vec{\theta}$" toho modelu $F(\cdot, \cdot)$?

Musely být $\equiv$ pro které $\vec{\theta}$ je největší šance, že jsme pak pozorovali $(x_1, z_1) \dots \dots (x_m, z_m)$

$\rightarrow$ tzv. maximum likelihood estimate parametrů $\vec{\theta}$

Odpověď:

pro ε malé („infinitezimálně")

$$\mathbb{P}\left(z_i - \varepsilon \leq Z_i < z_i + \varepsilon\right) = \int_{z_i-\varepsilon}^{z_i+\varepsilon} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{s - F(x_i, \vec{\theta})}{\sigma_i}\right)^2} ds \approx 2\varepsilon \cdot \frac{1}{\sigma_i \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{z_i - F(x_i, \vec{\theta})}{\sigma_i}\right)^2}$$

předpokládáme-li, že Z_i jsou iid, pak $\mathbb{P}\left(z_i - \varepsilon \leq Z_i < z_i + \varepsilon, \forall i\right) = \prod_{i=1}^{m} \mathbb{P}\left(z_i - \varepsilon \leq Z_i < z_i + \varepsilon\right)$

a tedy $\mathbb{P}\left(z_i - \varepsilon \leq Z_i < z_i + \varepsilon, \forall i\right) = \left(\frac{2\varepsilon}{\sqrt{2\pi}}\right)^m \cdot \prod_{i=1}^{m} \frac{1}{\sigma_i} \cdot \exp\left(-\frac{1}{2} \cdot \sum_{i=1}^{m} \left[\frac{z_i}{\sigma_i} - \frac{1}{\sigma_i} F(x_i, \vec{\theta})\right]^2\right)$

$\Longrightarrow$ čím menší $\sum_{i=1}^{m} \left[\frac{z_i}{\sigma_i} - \frac{1}{\sigma_i} F(x_i, \vec{\theta})\right]^2$ tím větší pravděpodobnost, že pro tyto parametry $\vec{\theta}$ jsme mohli získat tyto měření

$\Downarrow$ tzn. ten MLE odhad dopadne tak, že

$\vec{\theta}$ volíme jako $\operatorname{argmin} \left\| \left[\frac{1}{\sigma_i} F(x_i, \vec{\theta}) - z_i/\sigma_i\right]_{i=1,\dots,m} \right\|^2$

$b_i \equiv z_i/\sigma_i$

předpokládáme, že „máme"

$$\boxed{\min_{\vec{\theta}} \left\| \left[\frac{1}{\sigma_i} F(x_i, \vec{\theta}) - b_i\right]_{i=1,\dots,m} \right\|^2 \quad \dots \text{tzv. problém nejmenších čtverců}}$$

Ačkoliv $z_i \sim N(y_i, \sigma_i^2)$ není vždy oprávněný předpoklad, na základě CLT (central limit theorem) lze analogicky postupovat i v případě kdy δz_i jsou „pouze" iid (independent, identically distributed).

Problém nejmenších čtverců (least-squares problem) může být

- lineární $\Leftrightarrow f(x_i, \vec{\theta}) = \vec{a}_i^T \cdot \vec{\theta} + c_i$ pro nějaké $\vec{a}_i, c_i$
- nelineární $\Leftrightarrow$ není lineární

Jak obecně řešíme? → jako v práku → najdeme „kandidáty"

tj. $\vec{\theta}_\ell$, $\ell = 1, 2, \ldots$ pro které

$$\nabla_\theta \left(\left\| \left[\frac{1}{\sigma_i} F(x_i, \vec{\theta}) - b_i \right]_{i=1,\ldots,m} \right\|^2 \right) \Bigg|_{\vec{\theta} = \vec{\theta}_\ell} \overset{(*)}{=} 0$$

- pro lineární LS → lze „explicitně" spočítat
- pro nelineární LS → musíme použít numerické metody k aproximaci takových bodů. Některé numerické metody ale s ∇ vůbec nepracují a uplatňují jiné postupy.

Lineární LS:

$$F(x_i, \vec{\theta}) = \vec{a}_i^T \cdot \vec{\theta} + c_i \Rightarrow \begin{bmatrix} F(x_1, \vec{\theta}) \\ \vdots \\ F(x_m, \vec{\theta}) \end{bmatrix} = \begin{bmatrix} \vec{a}_1^T \cdot \vec{\theta} + c_1 \\ \vdots \\ \vec{a}_m^T \cdot \vec{\theta} + c_m \end{bmatrix} = \underset{\begin{bmatrix} - \vec{a}_1 - \\ - \vec{a}_2 - \\ \vdots \\ - \vec{a}_m - \end{bmatrix}}{\underset{\uparrow}{A}} \cdot \vec{\theta} + \underset{\begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}}{\underset{\uparrow}{\vec{c}}}$$

$$\Rightarrow \left\| \left[\frac{1}{\sigma_i} F(x_i, \vec{\theta}) - b_i \right]_{i=1,\ldots,m} \right\|^2 = \left\| \underbrace{\begin{bmatrix} - \vec{a}_1/\sigma_1 - \\ - \vec{a}_2/\sigma_2 - \\ \vdots \\ - \vec{a}_m/\sigma_m - \end{bmatrix}}_{=: A} \cdot \vec{\theta} - \underbrace{\begin{bmatrix} b_1 - c_1/\sigma_1 \\ b_2 - c_2/\sigma_2 \\ \vdots \\ b_m - c_m/\sigma_m \end{bmatrix}}_{=: \vec{b}} \right\|^2$$

⇝ přeznačíme: $=: A$ $\qquad =: \vec{b}$

$\Rightarrow$ lineární LS $\sim \min_{\vec{\theta}\in\mathbb{R}^n} \|A\vec{\theta}-\vec{b}\|^2$

kde $A\in\mathbb{R}^{m\times n}$, $\vec{b}\in\mathbb{R}^m$

$\rightsquigarrow$ $m \equiv$ počet měření & $n \equiv$ počet parametrů

$\Rightarrow$ $m > n$ nebo dokonce $m >> n$

$\Rightarrow$ A je obdélníková

<u>Pozorování</u> $\|A\vec{\theta}-\vec{b}\|^2 \geq 0 \ \forall \vec{\theta} \Rightarrow$ hledáme $\vec{\theta}$ t.ž. $A\vec{\theta}=\vec{b}$ $\rightsquigarrow$ řešení lin. soust. rovnic s obdélníkovou maticí.

ty už „umíme"

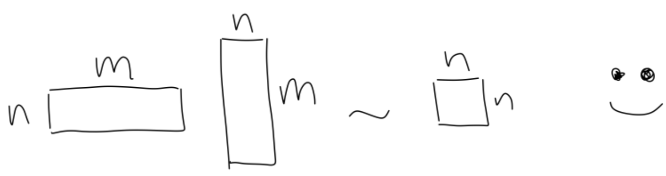
Lze převést na systém se čtvercovou?

<u>Lemma:</u> $\nabla_{\vec{\theta}}\left(\|\vec{b}-A\vec{\theta}\|^2\right) = 2\cdot A^T(A\vec{\theta}-\vec{b})$

(lze přímo vypočítat → „jednoduché" cviko z analýzy)

$$\Rightarrow \left\{ \nabla_{\vec{\theta}}\left(\|\vec{b}-A\vec{\theta}\|^2\right) = 0 \iff \boxed{A^TA\vec{\theta} = A^T\vec{b}} \right.$$

tzv. systém normálových rovnic

$A^TA \sim$  , ale

- sestavit A^TA vyžaduje násobení matic (drahé)
- $\kappa(A^TA) = \kappa(A)^2 \rightsquigarrow$ zhoršení podmíněnosti