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:

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

Výběr modelu ≡ problém pro odborníky v 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 jiné.

Překvapivě, volba $\vec{\theta}$ lze analyzovat relativně obecně

Máme dáno: • model, tj. funkci $f : (x, \vec{\theta}) \longmapsto y$

• měření, tj. páry (x_i, y_i), t. že v praxi x_i vede k y_i $\forall i$

$\rightsquigarrow$ měření nikdy není přesné $\to$ v hodnotách y_i máme chyby standardně modelujeme jako náhodné veličiny $\Rightarrow$ i kdybychom měli naprosto správný model $f(\cdot,\cdot)$ a správné parametry $\vec{\theta}$, tak

$\delta y_i := f(x_i, \vec{\theta}) - y_i \neq 0$ .... standardně modelujeme jako náhodnou veličinu $\forall i$ $\Big\} \to$

$\to$ např.: $\delta y_i \sim N(\mu_i, \sigma_i^2)$

Otázka: jaká volba parametrů $\vec{\theta}$ činí pozorované chyby δy_i nejpravděpodobnější?

(= volba takových parametrů se nazývá MLE)

Odpověď: $\mathbb{P}(0 \leq |\delta y_i| \leq \varepsilon) = \varepsilon \cdot \frac{1}{\sigma_i \sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2} \frac{(\delta y_i - \mu_i)^2}{\sigma_i^2}\right)$

& $\mathbb{P}(0 \leq |\delta y_i| \leq \varepsilon, i = 1, \dots, m) = \left(\varepsilon \cdot \frac{1}{\sigma_i \sqrt{2\pi}}\right)^m \cdot \exp\left(-\frac{1}{2} \sum_{i=1}^{m} \frac{(\delta y_i - \mu_i)^2}{\sigma_i^2}\right) =$

$$= \left(\varepsilon \cdot \frac{1}{\sigma_i \sqrt{2\pi}}\right)^m \cdot \exp\left(-\frac{1}{2} \cdot \sum_{i=1}^{m} \left[\frac{1}{\sigma_i} f(x_i, \vec{\theta}) - \frac{y_i - \mu_i}{\sigma_i}\right]^2\right)$$

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

$\vec{\theta}$ vybereme pomocí MLE

$\Updownarrow$

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

známe $\to$ lze spočítat

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

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

Problém nejmenších čtverců (least-squares problem) může být
- lineární $\iff f(x_i, \vec{\theta}) = \vec{a}_i^T \cdot \vec{\theta} + c_i$ pro nějaké $\vec{a}_i, c_i$
- nelineární $\iff$ není lineární

Jak obecně řešíme? $\to$ jako v práku $\to$ najdeme „kandidáty"

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

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

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

Lineární LS:

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

$$\Rightarrow \left\| \left[\frac{1}{\sigma_i} f(x_i, \vec{\theta}) - b_i \right]_{i=1,\dots,m} \right\|^2 = \left\| \underbrace{\frac{1}{\sigma_i} A}_{=: A} \cdot \vec{\theta} + \underbrace{\vec{c} - \vec{b}}_{=: \vec{b}} \right\|^2$$

$\rightsquigarrow$ přeznačíme:

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

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

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

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

$\Rightarrow$ A je obdélníková

Pozorování $\|A\vec{\theta} - \vec{b}\|^2 \geq 0 \;\; \forall \vec{\theta} \Rightarrow$ hledáme $\vec{\theta}$ t. , že $A\vec{\theta} = \vec{b}$ $\rightsquigarrow$ řešení lin. sous. rovnic s obdélníkovou maticí.

ty už „umíme"

Lze převést na systém se čtvercovou?

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

(lze přímo vypočítat $\rightarrow$ „jednoduché" cviko z analýzy)

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

tzv. systém normálových rovnic

$A^T A \sim$ 

ale sestavit $A^T A$ vyžaduje násobení matic (drahé)