

Optimalizace

6. Aplikace metody nejmenších čtverců

Tomáš Kroupa Tomáš Werner

2021 LS

Fakulta elektrotechnická
ČVUT v Praze

- **Regresní model** vyjadřuje závislost proměnných x a y funkcí

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

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

- Proměnná y je reálná, proměnná x může mít hodnoty v \mathbb{R}^k
- Pro data $(x_1, y_1), \dots, (x_m, y_m)$ hledáme $\boldsymbol{\theta}$ minimalizující

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

Pro jaké funkce f umíme efektivně spočítat $\boldsymbol{\theta}$?

Lineární regrese

Funkce $f(x, \boldsymbol{\theta})$ je lineární v *parametrech*:

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

kde $\varphi_1, \dots, \varphi_n$ jsou zadané funkce.

Metoda nejmenších čtverců

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^n} \sum_{i=1}^m \left(y_i - \boldsymbol{\varphi}(x_i)^\top \boldsymbol{\theta} \right)^2 = \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|^2,$$

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

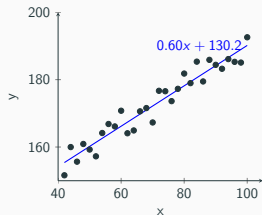
Příklad – prokládáme body přímkou

- 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:

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

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

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



Příklad – doprava (1)

- Úsek délky $130m$ mezi semaforey projede automobil za čas [s]

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

kde x je počet všech aut v úseku, $N = 30$ je kapacita úseku a θ_1, θ_2 jsou neznámé parametry

- Na základě $m = 5$ měření odhadujeme průměrnou rychlost auta [km/h] v úseku, pokud je v něm 12 aut

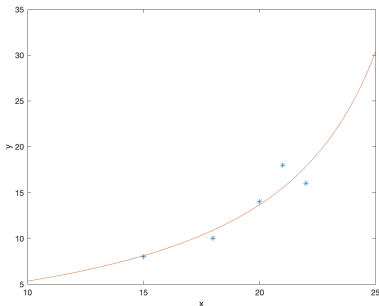
x_j	20	18	22	21	15
y_j	14	10	16	18	8

Příklad – doprava (2)

- Lineární regrese $y = \theta_1 + \frac{\theta_2 x}{1 - \frac{x}{N}}$, kde $\varphi_1 = 1$, $\varphi_2(x) = \frac{x}{1 - \frac{x}{N}}$,

$$\mathbf{y} \in \mathbb{R}^5 \text{ a } \mathbf{A} = \begin{bmatrix} 1 & \varphi_2(x_1) \\ \vdots & \vdots \\ 1 & \varphi_2(x_5) \end{bmatrix}$$

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



Pro $x = 12$ dostaneme průjezd za $y = 6.252$ s, tedy rychlost

$$\frac{N}{6.252} \cdot 3.6 = 74.861 \text{ km/h}$$

Příklad – doprava (3)

```
1 % Data
2 - x = [20;18;22;21;15];
3 - y = [14;10;16;18;8];
4 - plot(x,y,'*')
5 - xlabel('x')
6 - ylabel('y')
7 - hold on
8
9 % MNČ
10 - N =30;
11 - f = x./(1-x/N); % hodnoty druhé bázové funkce
12 - A = [ones(5,1) f];
13 - theta = A\y; % výpočetně lepší než inv(A'*A)*A'*y
14
15 % Graf výsledného modelu
16 - X = 10:0.01:25;
17 - Y = theta(1) + theta(2)*X./(1-X/N);
18 - plot(X,Y)
19 - hold off
20
21
```

Proložení bodů polynomem (1)

- Fitování polynomu na základě m pozorování (x_i, y_i)
- Regresní funkce je zde polynom stupně $n - 1$ proměnné x ,

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

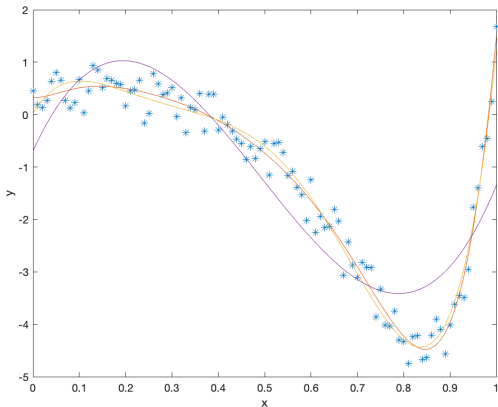
- Vandermondova matice $\mathbf{A} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ & & & \vdots & \\ 1 & x_m & x_m^2 & \dots & x_m^{n-1} \end{bmatrix}$

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 $y_i = f(x_i, \boldsymbol{\theta})$, $i = 1, \dots, m$.

Proložení bodů polynomem (2)

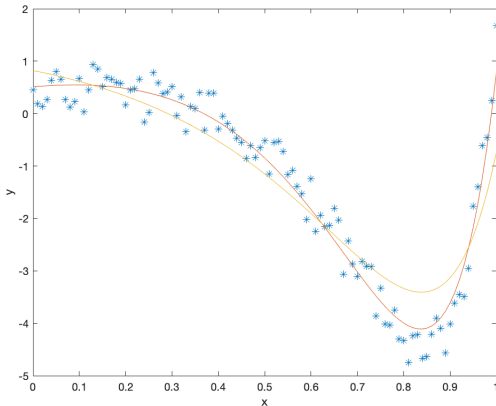
- Přesná interpolace vede na polynomy velmi vysokého stupně
- Preferujeme jednoduché modely vystihující podstatné rysy dat
- Pro $m = 101$ měření (x_i, y_i) porovnáme 3 různé polynomy



$n - 1$	RSS	$\ \theta\ $
3	57	77
5	8.93	713
10	7.76	$4 \cdot 10^4$

Proložení bodů polynomem (3)

- Regularizací zjednodušíme model s polynomem 10. stupně
- Minimalizujeme kritérium $\|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|^2 + \mu\|\boldsymbol{\theta}\|^2$ pro nějaké $\mu > 0$
- Jediné řešení je $\boldsymbol{\theta}^* = (\mathbf{A}^\top \mathbf{A} + \mu \mathbf{I})^{-1} \mathbf{A}^\top \mathbf{y}$



μ	RSS	$\ \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 $\mathbf{A}^T \mathbf{A} \boldsymbol{\theta} = \mathbf{A}^T \mathbf{y}$ má více řešení
- Protože je matice $\mathbf{A}^T \mathbf{A} + \mu \mathbf{I}$ invertibilní pro libovolné $\mu > 0$, můžeme použít regularizované řešení $\boldsymbol{\theta}^* = (\mathbf{A}^T \mathbf{A} + \mu \mathbf{I})^{-1} \mathbf{A}^T \mathbf{y}$

Statistický pohled na lineární model

Do modelu lineární regrese začleníme **náhodnost** spočívající v chybách měření nebo šumu:

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

- $\mathbf{A} \in \mathbb{R}^{m \times n}$ pozorovatelné hodnoty vysvětlujících proměnných
- $\boldsymbol{\theta} \in \mathbb{R}^n$ neznámé odhadované parametry
- $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_m)$ *náhodné* chyby
- $\mathbf{y} = (y_1, \dots, y_m)$ *náhodné* vysvětlované proměnné

Předpoklady lineárního modelu $\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\epsilon}$

- Každá chyba ϵ_i má nulovou střední hodnotu, tedy $\mathbb{E}[\epsilon_i] = 0$
- Všechny chyby ϵ_i mají stejný rozptyl σ^2 , tedy $\mathbb{E}[\epsilon_i^2] = \sigma^2$
- Různé chyby ϵ_i a ϵ_j jsou nekorelované, tedy $\mathbb{E}[\epsilon_i\epsilon_j] = 0$

Zkrácený zápis

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

Lineární estimátor $\hat{\theta}$ pro model $\mathbf{y} = \mathbf{A}\theta + \epsilon$

Lineární estimátor je náhodný vektor $\hat{\theta} := \mathbf{B}\mathbf{y}$, kde $\mathbf{B} \in \mathbb{R}^{n \times m}$.
Říkáme, že $\hat{\theta}$ je nevychýlený, pokud

$$\mathbb{E}[\hat{\theta}] = \theta.$$

Vlastnosti

- Střední hodnota $\mathbb{E}[\hat{\theta}] = \mathbf{B}\mathbf{A}\theta$
- Estimátor $\hat{\theta}$ je nevychýlený $\Leftrightarrow \mathbf{B}$ je levá inverze matice \mathbf{A}
- Kovarianční matice nevychýleného estimátoru je $\sigma^2\mathbf{B}\mathbf{B}^T$

LS estimátor pro model $y = \mathbf{A}\theta + \epsilon$

Pokud má \mathbf{A} LN sloupce, definujeme **LS estimátor** pomocí matice

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

a pozorujeme, že

- je nevychýlený, protože $\mathbf{A}^+ \mathbf{A} = \mathbf{I}$
- má kovarianční matici $\sigma^2 (\mathbf{A}^T \mathbf{A})^{-1}$

Věta (Gauss-Markov)

LS estimátor má ze všech nevychýlených lineárních estimátorů nejmenší čtvercovou chybu.

Jak dlouho padá kulička do studny?

Máme m kuliček a měříme dobu dopadu y_i kuličky i do studny.

- Lineární model $\mathbf{y} = \mathbf{A}\theta + \epsilon$, kde $\mathbf{A} = [1 \dots 1]^T$
- LS estimátor je **aritmetický průměr**

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

- Rozptyl LS estimátoru je

$$\sigma^2 (\mathbf{A}^T \mathbf{A})^{-1} = \frac{\sigma^2}{m}$$

Odhad metodou nejmenších čtverců je maximálně věrohodný



Předpoklady modelu $\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\epsilon}$ jsou nyní:

- nezávislost chyb $\epsilon_1, \dots, \epsilon_m$
- $\epsilon_1, \dots, \epsilon_m \sim N(0, \sigma)$

Hodnotíme $\mathbf{r}(\boldsymbol{\theta}) := \tilde{\mathbf{y}} - \mathbf{A}\boldsymbol{\theta}$, kde $\tilde{\mathbf{y}} \in \mathbb{R}^m$ jsou naměřené hodnoty.

Tyto optimalizační úlohy jsou ekvivalentní

1. $\min_{\boldsymbol{\theta} \in \mathbb{R}^n} \|\mathbf{r}(\boldsymbol{\theta})\|^2$
2. $\max_{\boldsymbol{\theta} \in \mathbb{R}^n} f_{\epsilon}(\mathbf{r}(\boldsymbol{\theta}))$, kde $f_{\epsilon}(\mathbf{r}(\boldsymbol{\theta}))$ je věrohodnost naměřených chyb pro parametry $\boldsymbol{\theta}$ a rozdělení $N(0, \sigma)$

-  T. Werner. *Optimalizace* (kapitola 5). Elektronická skripta. FEL ČVUT, 2020.
-  F. J. Aragón, M. A. Goberna, M. A. López, and M. M. Rodríguez. *Nonlinear optimization*. Springer, 2019.