

A04 – Predikcia kvality vína, lineárna regresia pomocou L^1 , L^∞

Piatí proti optimalizácii

Tomáš Antal, Erik Božík, Róbert Kendereš,

Teo Pazera, Andrej Špitalský

2DAV

Január 2024

Predstavenie projektu - lineárna regresia

- ▶ lineárna regresia - predikcia závislej premennej $y \in \mathbb{R}^n$ pomocou nezávislých $x_1, \dots, x_n \in \mathbb{R}^n$

$$\min ||y - \hat{y}||$$

$$\hat{y} = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

pozorovania	x_1	x_2	\dots	x_k	y
atribút 1	1	2	\dots	4	5
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
atribút n	1	2	\dots	4	5

- ▶ vyjadriteľné ako úloha lineárneho programovania - L^1, L^∞

Predstavenie projektu - obsah

- ▶ formulácia LP úloh a dokázanie optimality
- ▶ implementácia v Python-e a predikcia kvality vína
- ▶ počítanie a interpretácia R^2 koeficientu
- ▶ implementácia všeobecnej triedy na počítanie L^1 a L^∞ lineárnej regresie
- ▶ minimalizácia váženej sumy noriem

A

...

B

...

...

D

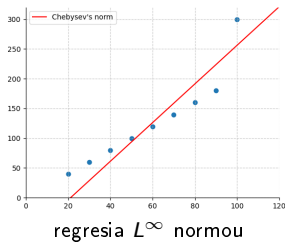
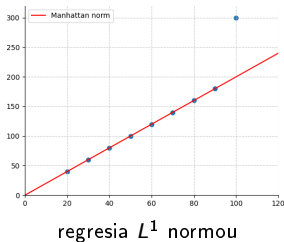
...

E-model

...

Porovnanie L^1 a L^∞ lineárnej regresie

- ▶ L^1 - veľmi dobre zachytáva lineárny vzťah, môže viesť k *overfittingu*
- ▶ L^∞ - príliš ovplyvňovaná outliermi



Minimalizácia váženého súčtu noriem

- ▶ redukcia *overfittingu* L^1 regresie váženým súčtom s L^∞ normou
- ▶ stále implementovateľné ako úloha lineárneho programovania
- ▶ nadobúda optimum

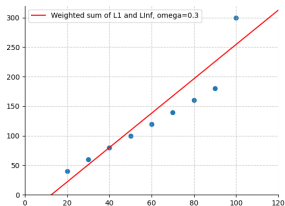
$$\min \left(0_{k+1}^T \mid \omega 1_n^T \mid (1-\omega) \right) \begin{pmatrix} \beta \\ t \\ \gamma \end{pmatrix}, \omega \in [0; 1]$$

$$\left(\begin{array}{c|c|c} A & \mathbb{I}_n & 0_n \\ \hline -A & \mathbb{I}_n & 0_n \\ \hline A & 0_{n \times n} & 1_n \\ \hline -A & 0_{n \times n} & 1_n \end{array} \right) \begin{pmatrix} \beta \\ t \\ \gamma \end{pmatrix} \geq \begin{pmatrix} y \\ -y \\ y \\ -y \end{pmatrix}$$

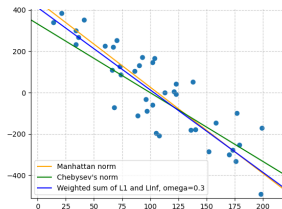
$$\beta \in \mathbb{R}^{k+1}, t \geq 0_n, \gamma \geq 0$$

Minimalizácia váženého súčtu noriem

► implementované ako WeightedL1LInfModel



regresia váženým súčtom noriem



porovnanie troch regresíí

Výsledky

...