RDD - 12

10 May 2021 11:15

- dotazník - čo by ste zmenili, čo sa vám nepáčilo
- ANKETA
- projekt - deň pred skúškou
- skúška - online

- asi dnes nestihneme
  => nahradná video-prednáška na stránke

Minule: kombinatorická selekcia

$y = X\beta + \varepsilon$

$X_{n\times k} = (1_n, X_1, \dots, X_{k-1})$ → výber prem. z $X_1, \dots, X_{k-1}$

- všetky podmnožiny: 2^{k-1}
- dopredná selekcia
- spätná sel.
- "dojistramná" sel.

→ AIC, BIC → MSE_{val}

Online News:

- všetky podmnožiny — manuálne (pre iba niektoré X_j)
  — regsubsets

Namiesto IC: validácia, krosvalidácia

validácia: dáta do 3 častí

1.) trénovacia vzorka: na nej odhadujeme β (fitujeme model) → $\hat\beta$ (train)

2.) validačná vzorka: výber modelu (val)

3.) testovacia (test): posúdenie kvality finálneho modelu

→ Hľ val: modelom predikujem $\hat y_i = x_i^T \hat\beta$ ⟷ porovnáme s realitou: y_i

$$MSE_{val} = \frac{1}{n_{val}} \sum_{i \in val} (y_i - x_i^T \hat\beta)^2 = \frac{1}{n_{val}} \| y_{val} - X_{val} \hat\beta \|^2 \to \min$$

n_{val} ↑ veľkosť val; min ↑ voľbou modelu

= → validačná chyba

- namiesto modelu s najmenším AIC/BIC skúšam model s najmenšou validačnou chybou — všetky podmnožiny
  — krokové met. — dopr. / spät. / dojistramná — selekcia

Alternatíva:

- $MSE_{train} = \| y_{train} - X_{train} \cdot \beta \|^2 \to \min_{\beta} \Rightarrow \hat\beta$: trénovacia chyba
- $MSE_{val} = \| y_{val} - X_{val} \hat\beta \|^2 \to \min_{M} \Rightarrow$ Model : val. chyba
- $MSE_{test} = \| y_{test} - X_{test} \cdot \hat\beta \|^2 \Rightarrow$ kvalita modelu : testovacia chyba

rozdelenie dát: napr. 80% - 10% - 10%
50% - 25% - 25%
(train - val - test)

Krosvalidácia:

málo dát ⇒ nechceme odčleniť val. vzorku
⇒ validácia trikrok pomocou trénovacej vzorky: krosvalidácia

- rozdelíme train na K častí (napr. $k=5$) rovnakej veľkosti: $y_{(1)}, \dots, y_{(K)}$, $X_{(1)}, \dots, X_{(K)}$

$$\begin{pmatrix} y_{(1)} \\ \vdots \\ y_{(K)} \end{pmatrix} = \begin{pmatrix} X_{(1)} \\ \vdots \\ X_{(K)} \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_{(1)} \\ \vdots \\ \varepsilon_{(K)} \end{pmatrix}$$

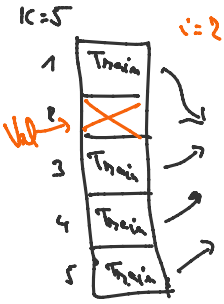
$\forall i = 1, \dots, k$: vezmem všetky časti okrem i-tej (4/5 dát)
→ na nich odhadnem model
→ predikujem $y_{(i)}$ pomocou toho modelu ⇒ MSE = chyba$_i$:

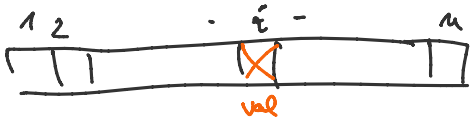
$$\text{chyba}_i = \frac{1}{n_i} \| y_{(i)} - X_{(i)} \cdot \hat{\beta}_{(i)} \|^2 \quad , \quad \hat{\beta}_{(i)} = (X_{-i}^T X_{-i})^{-1} X_{-i}^T y_{-i}$$

(n_i ↑ veľkosť i-tej časti; X_{-i}, y_{-i} ↑ všetko okrem i-tej)

$$\Rightarrow \text{val. chyba} = \frac{1}{k} \sum_{i=1}^{k} \text{chyba}_i = \frac{1}{k} \sum_{i=1}^{k} \frac{1}{n_i} \| y_{(i)} - X_{(i)} \hat{\beta}_{(i)} \|^2$$

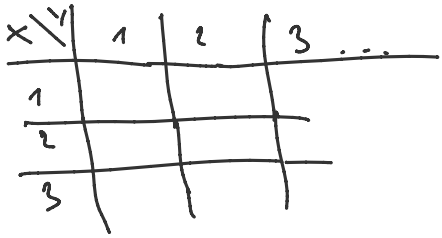
⇓ výber modelu





- voľba k: napr. $k=5$, $k=10$, $K=n$
↓ leave-one-out crossvalidation

4 Rko - Forest fires:



validácia, krosvalidácia - manuálne
↳ v Rku: balíček caret

validácia, krosvalidácia – manuálne

↳ v R-ku: balíček caret – istá forma spätnej selekcie

– funguje trochu inak – na každej z K sád nejako hodnotí premenné, a potom odhadzuje premenné s najhorším celkovým skóre

---

Penalizované regresie

$y = X\beta + \varepsilon$, $X = (1\, x_1, \dots, x_{k-1}) = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}$

– komb. selekcia – pre každú prem. $*$: v modeli < Áno / Nie

– veľmi skokové: nestabilné voči perturbácii dát

iný prístup: model so všetkými prem., ale zmenšíme hodnoty $\hat\beta$ → spojito, nie iba buď nula, alebo nenula

⟶ k LS pridáme penaltu:

$$\hat\beta \in \operatorname{argmin}\left(\|y - X\beta\|^2 + \lambda \cdot p(\beta)\right)$$

penalizačný parameter $\lambda \geq 0$; zvolená penalta (zvyčajne norma)

- $\lambda = 0 \Rightarrow$ OLS
- $\lambda \to \infty \Rightarrow \hat\beta = \vec{0}$
- β_0 nechceme skracovať: najprv všetko vycentrujeme: $y \mapsto y - \bar{y}$, $X_j \mapsto X_j - \bar{X}_j \quad \forall j$

: potom už nepotrebujem β_0

1. Ridge regression – Hrebeňová regresia

$p(\beta) = \|\beta\|_2^2 \quad \Rightarrow \quad \hat\beta_{ridge} \in \underset{\beta}{\operatorname{argmin}} \|y - X\beta\|^2 + \lambda \|\beta\|^2 = \sum_{i=1}^{n} (y_i - x_i^T\beta)^2 + \lambda \cdot \sum_{j=1}^{k-1} \beta_j^2$

$= (y - X\beta)^T (y - X\beta) + \lambda \beta^T\beta$

$= \underbrace{y^T y - 2y^T X\beta + \beta^T X^T X \beta}_{f(\beta)} + \lambda \beta^T \beta \longrightarrow \min_\beta$

$\frac{\partial}{\partial\beta} a^T\beta = a$

$\frac{\partial}{\partial\beta} \beta^T A \beta = 2A\beta$

$\frac{\partial f}{\partial \beta} = 0 - 2(y^T X)^T + 2X^T X\beta + 2\lambda\beta = 0 \qquad /:2$

$$\frac{\partial \mathcal{L}}{\partial \beta} = 0 - 2(y^T X)^T + 2X^T X\beta + 2\lambda\beta = 0 \qquad /\cdot 2$$

$$(X^T X + \lambda I)\,\beta = X^T y$$

$$\hat{\beta}_{ridge} = (X^T X + \lambda I)^{-1} X^T y \;//$$

$$\lambda I = \begin{pmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda \end{pmatrix}$$

- pôvodná motivácia: $X^T X + \lambda I$ je lepšie podmienená ako $X^T X$ (lepšie sa invertuje)
  (Hoerl & Kennard, 1970)

$\hat{\beta}_{ridge}$ ako riešenie úlohy s obmedzením:

$$\left\{ \begin{array}{l} \min \|y - X\beta\|_2^2 \\ \|\beta\|_2^2 \le c \end{array} \right\} \quad (\acute{u})$$

Lagrange: $\mathcal{L} = \|y - X\beta\|_2^2 + \mu(\beta^T\beta - c)$ ← to je presne $f(\beta)$, akurát namiesto μ máme λ (a je tu navyše $-\mu c$, čo nezohľadňuje na riešenie)

riešenie $(\acute{u})$: $\frac{\partial \mathcal{L}}{\partial \beta} = 0$

$\Rightarrow$ $\hat{\beta}_{ridge}$ je riešením $(\acute{u})$

$\Rightarrow$ $\hat{\beta}_{ridge}$: minimalizácia RSS, ale máme obmedzenie na to, aký veľký môže β byť