Porovnanie funkcií na selekciu fixných efektov vo vysokodimenzionálnych lineárnych zmiešaných modeloch

Jozef Jakubík

Ústav merania SAV

23. januára 2014

Vysokodimenzionálne dáta

$$p > n$$

Problémy

```
Console ~/
> bi2 <- rep(d2,n1)
> bi <- rbind(bi1,bi2)
> z=x[,1:2,drop=FALSE]
> z1=x1[,1:2,drop=FALSE]
> x<-x[,c(1,3:(p+2))]
> x1<-x1[,c(1,3:(p+2))]
> epsilon=rnorm(120)
> epsilon1=rnorm(120)
> y <- numeric(n)
> y1 <- numeric(n)
> for (k in 1:n) y[k] <- x[k,]%*%beta + t(z[k,])%*%bi[,k] + epsilon[k]
>
> for (l in 1:n) y1[l] <- x1[l,]%*%beta + t(z1[k,])%*%bi[,k] + epsilon1[k]
>
> model1<-vector()
> model1=which(beta!=0)
> xorig=x[,model1]
> fitorig=lmer(y~ -1 + xorig+(-1+z|grp))
>
> fitplny=lmer(y~ -1 + x+(-1+z|grp))
Error in lme4::lFormula(formula = y ~ -1 + x + (-1 + z | grp), control = list( :
  rank of X = 120 < ncol(X) = 200
>
```

```
> daco1
[1] 13.17404
> daco2
[1] 87.52728
> daco3
[1] 12.23474
>
```


Jozef Jakubík
Ústav merania SAV

Očakávania

- Submodel dobre popisuje namerané pozorovania.
- Submodel je redší ako pôvodný model.
- Submodel poskytuje primerané predikcie.
- Testovať významnosť parametrov a konštruovať konfidenčné intervaly.

Hľadanie optimálneho modelu

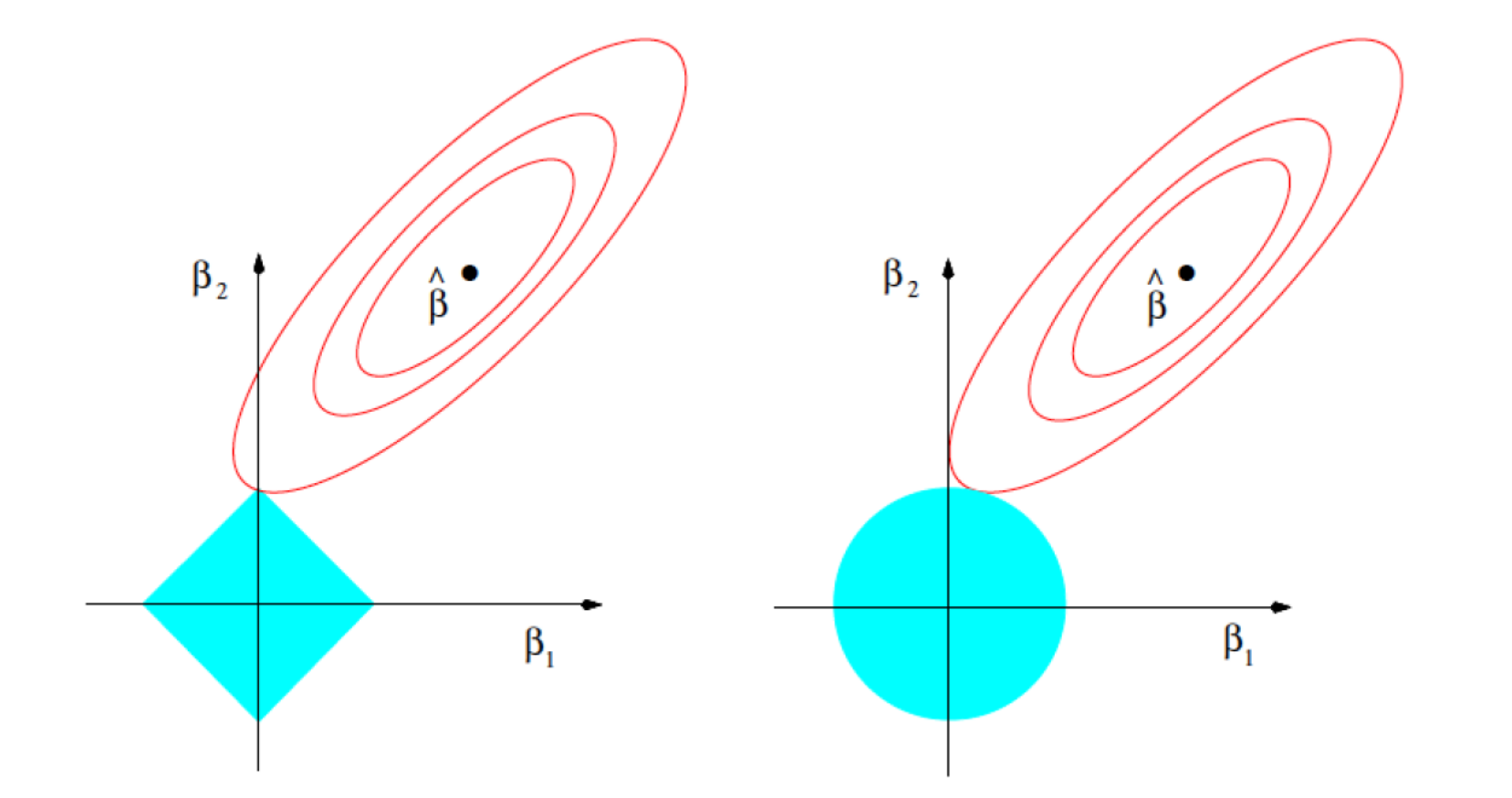
- Best-Subset výber
- Forward- a Backward-Stepwise výber
- Regularizačné metódy (Ridge regresia, Lasso, ...)
- ...

Lasso - Least Absolute Selection and Shrinkage Operator

Regularizácia za pomoci ℓ_1 normy.

$$\begin{aligned}\hat{\beta}^{lasso} &= \underset{\beta \in \mathbb{R}^p}{argmin} \|y - X\beta\|_2^2 + \lambda \sum_{j=1}^{p} |\beta_j| \\ &= \underset{\beta \in \mathbb{R}^p}{argmin} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \\ &= \underset{\beta \in \mathbb{R}^p, \|\beta\|_1 \leq t}{argmin} \|y - X\beta\|_2^2\end{aligned}$$

Lasso vs. Ridge



Jozef Jakubík
Ústav merania SAV

Čo ideme riešiť

$$y = X\beta + Zb + \varepsilon$$

1. $\varepsilon \sim \mathcal{N}(0, R = (\sigma^2 I))$ sú nekorelované
2. $b \sim \mathcal{N}(0, D(\text{blokovo diagonálna}))$
3. ε, b sú nezávisé

- skupinové dáta
- jednoduchá, ale opodstatnená kovariančná štruktúra
- veľký počet fixných efektov a malý počet náhodných efektov

Aktuálne články

- Schelldorfer, Bühlmann, van de Geer - 2011
- Rohart, San-Cristobal, Laurent - 2012

Teoretické predpoklady

- Pre konzistenciu odhadu potrebujem technickú podmienku na vlastné čísla matice $Z^T Z$ v zmysle horného ohraničenia.
- Ak predpokladáme, že skutočný model je riedky a je splnená podmienka pre konzistenciu odhadu, potom obe metódy dokážu odhadnúť skutočný model.

Schelldorfer - ℓ_1-penalizovaný ML odhad

$$y \sim \mathcal{N}(X\beta, V(= ZDZ^T + R))$$

$$\min_{\beta,V} \left(\frac{1}{2} log|V| + \frac{1}{2}(y - X\beta)^T V^{-1}(y - X\beta) + \lambda \ \|\beta\|_1 \right),$$

kde λ je nezáporný penalizačný parameter, fixný počas celej minimalizácie.

Coordinate Gradient Descent algoritmus

CGD algoritmus spočíva v cyklickom prechádzaní cez jednotlivé premenné a minimalizovaní účelovej funkcie vzhľadom na jednu premennú, pričom ostatné sa berú ako fixné (podobne ako Gauss-Seidlova metóda).

`R - lmmlasso`

Rohart - rovnaká úloha, iný prístup

Vektor b považujme za chýbajúce dáta. Kompletné dáta označme $x = (y^T, b^T)^T$.

$$\min_{\beta,D,R} \left(log|R| + (y - X\beta - Zb)^T R^{-1} (y - X\beta - Zb) + \right.$$

$$\left. + log|D| + b^T D^{-1} b + \lambda \ \|\beta\|_1 \right),$$

kde n je počet pozorovaní a λ je opäť nezáporný penalizačný parameter, fixný počas celej minimalizácie.

Multicyklický Expectation Conditional Maximization algoritmus

Účelovú funkciu minimalizujeme multicyklickým ECM algoritmom:

- Prvý E krok určí Hendersonov (BLUP) odhad $b^{[t+1/2]}$ na základe aktuálnych hodnôt parametrov $\beta^{[t]}$, $D^{[t]}$, $R^{[t]}$.
- M krok odhadne najskôr $\beta^{[t+1]}$.
- E krok rovnako odhadne $b^{[t+1]}$.
- M krok Odhadne variačné komponenty $D^{[t+1]}$ a $R^{[t+1]}$.

`R - MMS`

Výber regularizačného parametra λ

`R - lmer`

Bayesovo informačné kritérium (BIC):

$$BIC_\lambda := -2\ell + log\, n \cdot \hat{df}_\lambda,$$

kde $\hat{df}_\lambda = |\{1 \leq k \leq p : \hat{\beta}_k \neq 0\}| + dim(D)$.

Porovnanie

- takmer žiadne rozdiely v čase behu funkcií (Rohart takmer vždy rýchlejší)
- $\pm$ podobné výsledky
- momentálne sa nedajú použiť ak je počet fixných efektov veľký
- problémy ak sú fixné efekty závislé
- ...

Ďakujem za pozornosť.