Vliv multikolinearity a odlehlých pozorování

Tomáš Jurczyk

Univerzita Karlova v Praze,
Matematicko-fyzikální fakulta
Katedra pravděpodobnosti a matematické statistiky

ROBUST 2010, Králíky, 31.1.-5.2.2010

Struktura posteru

Nejmenší čtverce $\xrightarrow{\text{Multikolinearita}}$ *Metody bojující s multikolinearitou*

Nejmenší čtverce $\xrightarrow{\text{Odlehlá pozorování}}$ *Robustní metody*

Metody bojující s multikolinearitou $\xrightarrow{\text{Multikolinearita + odlehlá pozorování}}$ *Metody řešící oba problémy*

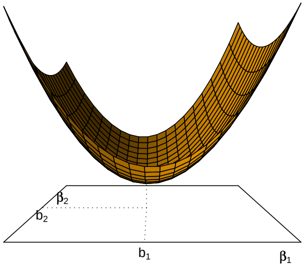
Robustní metody $\xrightarrow[\text{+ odlehlá pozorování}]{\text{Multikolinearita}}$ *Metody řešící oba problémy*

Klasické metody

Problémy jsou v posteru názorně ukázány pomocí grafů ztrátových funkcí:

Metoda nejmenších čtverců:

Příklad (Ztrátová funkce pro **LS**, 2 regresory, data **bez multikolinearity**)



- ← $\sum_{i=1}^{n}(Y_i - x_{1i}\beta_1 - x_{2i}\beta_2)^2 = \sum_{i=1}^{n} r_i^2(\boldsymbol{\beta})$
- ← typický tvar pro data bez multikolinearity
- ← $(b_1, b_2)' = \boldsymbol{b}$ značí LS odhad – nejmenší hodnota ztrátové funkce

Problémy s multikolinearitou

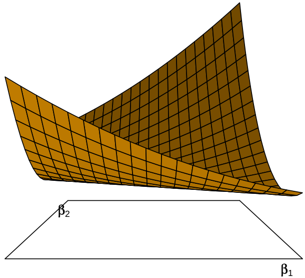
Multikolinearita

- situace, kdy jsou regresory "téměř" lineárně závislé

Důsledky pro nejmenší čtverce (Ztrátová funkce pro **LS**, 2 regresory, data **s multikolinearitou**)

→ tvar odpovidající silné multikolinearitě v datech

→ téměř stejné hodnoty ztrátové funkce podél jisté přímky v parametrickém prostoru ⇒ nestabilní chování ⇒ velký rozptyl odhadu



Řešení problémů s multikolinearitou

Abychom se vyhnuli velkým rozptylům statistik b_j, doporučuje se za přítomnosti multikolinearity používat následující vychýlený odhad.

Hřebenová regrese

Pro $\delta \geq 0$ definujeme hřebenový odhad vektoru $\boldsymbol{\beta}$ jako

$$\boldsymbol{b}_\delta = \operatorname*{argmin}_{\beta \in \mathcal{R}^p} \left(\sum_{i=1}^{n} r_i^2(\boldsymbol{\beta}) + \delta \sum_{j=1}^{p} \beta_j^2 \right).$$

Jak funguje hřebenová regrese

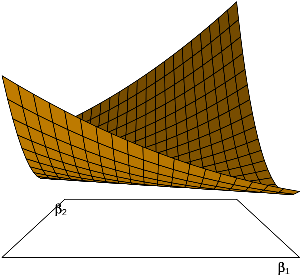
příklad (2 regresory, data **s multikolinearitou**)

Ztrátová funkce pro LS odhad

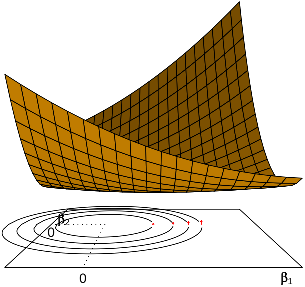
$$\sum_{i=1}^{n} r_i^2(\boldsymbol{\beta})$$

Ztrátová funkce pro hřebenový odhad

$$\sum_{i=1}^{n} r_i^2(\boldsymbol{\beta}) + \delta \sum_{j=1}^{p} \beta_j^2$$



→



Struktura posteru

Klasické metody

Multikolinearita
$\longrightarrow$

Metody bojující s multikolinearitou

Odlehlá pozorování
$\downarrow$

Multikolinearita
+ odlehlá pozorování
$\downarrow$

Robustní metody

Multikolinearita
$\longrightarrow$
+ odlehlá pozorování

Metody řešící oba problémy

Metoda řešící oba problémy

Nejmenší čtverce

Multikolinearita
→

Hřebenová regrese

Odlehlá pozorování
↓

Multikolinearita
+ odlehlá pozorování
↓

Nejmenší vážené čtverce

Multikolinearita
→
+ odlehlá pozorování

Robustní hřebenová regrese

Metoda řešící oba problémy

$$\sum_{i=1}^{n} r_i^2(\boldsymbol{\beta}) \longrightarrow \sum_{i=1}^{n} r_i^2(\boldsymbol{\beta}) + \delta \sum_{j=1}^{p} \beta_j^2$$

$$\downarrow \qquad\qquad \downarrow$$

$$\sum_{i=1}^{n} w_i r_{(i)}^2(\boldsymbol{\beta}) \longrightarrow \sum_{i=1}^{n} w_i r_{(i)}^2(\boldsymbol{\beta}) + \delta \sum_{j=1}^{p} \beta_j^2$$

Návrh robustní hřebenové regrese

Dospěli jsme tedy k návrhu metody, která by se měla umět vypořádat s oběma problémy najednou.

Hřebenový odhad nejmenších vážených čtverců

Pro $\delta \geq 0$ definujeme hřebenový odhad nejmenších vážených čtverců vektoru $\boldsymbol{\beta}$ jako

$$\boldsymbol{b}_{\delta}^{(LWS,w)} = \underset{\beta \in \mathcal{R}^p}{\operatorname{argmin}} \left(\sum_{i=1}^{n} w_i r_{(i)}^2(\boldsymbol{\beta}) + \delta \sum_{j=1}^{p} \beta_j^2 \right).$$

Děkuji
za
pozornost

Více informací na mém posteru