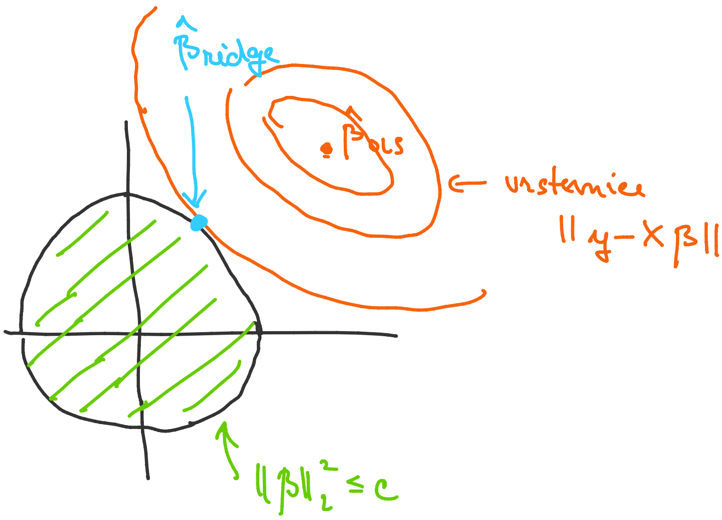
Minule: Hrebeňová regresia

$y = X\beta + \varepsilon$

$$\left[\ \|y - X\beta\|_2^2 + \lambda \cdot \|\beta\|_2^2 \to \min_{\beta}\ \right]$$

ako úloha s ohraničením:

$$\left\{ \begin{array}{l} \min\limits_{\beta} \|y - X\beta\|_2^2 \\ \|\beta\|_2^2 \le c \end{array} \right\}$$



- Ridge: napr. rieši situáciu, keď X_i, X_j sú vysoko korelované
  - → $\hat\beta$: jedna hodnota veľmi kladná, druhá veľmi záporná
  - → $\hat\beta_{ridge}$: zmierni ich
- $\hat\beta_{ridge}$: závisí od škálovanie ⇒ zvyčajne: normalizácia $X_1, \dots, X_{k-1}$ najprv
- $\hat\beta_{ridge}$ je vychýlený (aj keď model platí)

→ ale: pri zvyšovaní λ ↗ zvyšuje sa vychýlka, ↘ znižuje sa variancia } môže sa zlepšovať kvalita predikcie (MSE)

formálne: $MSE(\hat\beta_{ridge}) = E\left[(\hat\beta_\lambda - \beta)^T(\hat\beta_\lambda - \beta)\right] = \sum_{j=1}^{k-1} E\left[(\hat\beta_j^{ridge} - \beta_j)^2\right]$

$$\Rightarrow MSE(\hat\beta_{ridge}) = Var + Bias^2$$

: bias-variance tradeoff

$Var = \sum_{j=1}^{k-1} D(\hat\beta_j^\lambda)$ — klesá s λ

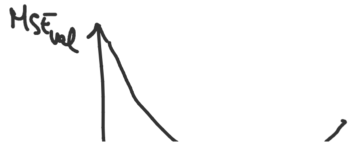
$\sum_{j=1}^{k-1} E(\hat\beta_j^\lambda - \beta_j) = Bias$ — rastie s λ

Voľba λ:

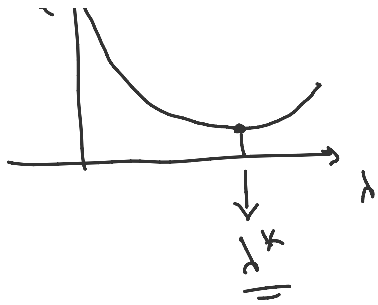
(kros)validácie

- validácia: kvalita predikcie na val. vzorke:

$$\forall \lambda: \ MSE_{val}(\lambda) = \frac{1}{} \| y - X\hat\beta_{..}(\lambda)\|^2$$



$$\forall \lambda: \quad MSE_{val}(\lambda) = \frac{1}{n_{val}} \| y - X \hat{\beta}_{ridge}(\lambda) \|^2$$



→ vyberieme λ, kt. minimalizuje $MSE_{val}(\lambda)$

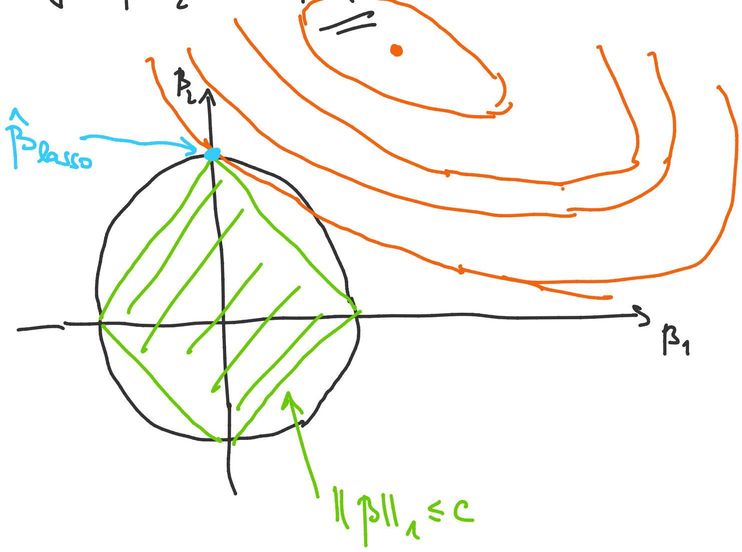
– krosvalidácia: to isté, akurát val. chybu meriame krosvalidáciou

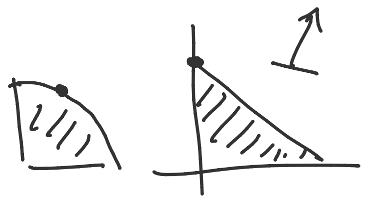
2. Lasso – least absolute shrinkage and selection operator

$$\hat{\beta}_{lasso} = \arg\min_{\beta} \|y - X\beta\|_2^2 + \lambda \underbrace{\sum_{j=1}^{k-1} |\beta_j|}_{\|\beta\|_1} = \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

v tvare s obmedzením:

$$\begin{Bmatrix} \min_{\beta} \|y - X\beta\|_2^2 \\ \|\beta\|_1 \le c \end{Bmatrix}$$





- tvar obmedzenia ⇒ Lasso má tendenciu kádzať niektoré $\hat{\beta}_j$ do nuly
- čím vyššie λ, tým viac nulových $\hat{\beta}_j$

⇒ kombinuje vlastnosti Ridge a kombinatorickej selekcie: odolnejšie voči perturbáciam (postupne zmenšuje $\hat{\beta}_j$)
: ale predsa – niektoré $\hat{\beta}_j$ dá na nulu ⇒ robí selekciu prem.

- L_1-norma $\| \ \|_1$: nie je hladká ⇒ $\nexists$ formulka pre $\hat{\beta}_{lasso}$
  – ale stále je lasso konvexná úloha ⇒ $\exists$ efektívne algoritmy

⇒ Ridge, Lasso sú často podstatne menej náročné alternatívy ku komb. selekcii

- opäť: začína sa s centrovanými a štandardizovanými dátami
- voľba λ: (kros)validácia

- ∃ Group lasso pre skupiny premenných (dummy)
- ∃ aj iné penalizačné metódy → napr. : $p(\beta) = \sum_{j=1}^{k-1} |\beta_j|^q$ pre $q \in (0, \infty)$

  : kombinácia $\|\beta\|_1$ a $\|\beta\|_2$