

Statističko učenje

Pristranost i varijanca za ridge regresiju i lasso

prosinac, 2021

Model

Pretpostavljamo da je $Y_i = \beta^t x^{(i)} + \epsilon$, $i = 1, \dots, n$ uz $n = 50$, pri čemu je

– $x^{(i)} \in \mathbb{R}^p$ za $p = 30$ (generirani njd iz $N(0, 1)$);

– ϵ_i njd iz $N(0, 1)$ (dakle, $\sigma = 1$);

Dakle, p je velik u odnosu na n .

Prisjetimo se,

$$\hat{\beta}^r := \arg \min_{\beta \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_2^2, \text{ (ridge)}$$

$$\hat{\beta}^{las} := \arg \min_{\beta \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1, \text{ (lasso)}.$$

Očito, u oba slučaja $\hat{\beta}^{ls}$ dobijemo za $\lambda = 0$. Ako je $\hat{\beta}$ procjena za β (bilo kojom metodom), $\hat{f}(x) := \hat{\beta}^t x$ je procjena za $f(x) = \beta^t x$, $x \in \mathbb{R}^p$.

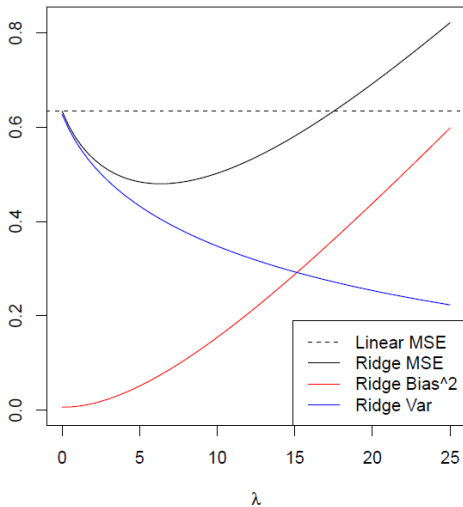
Pristranost i varijanca

Simulacijom ćemo ispitati kako se ponaša tzv. očekivana greška unutar uzorka ("in-sample error")

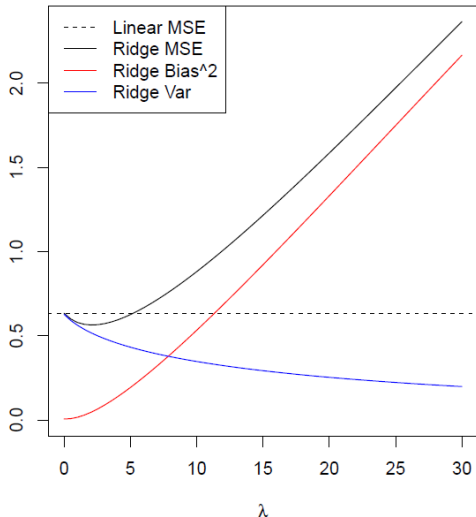
$$\begin{aligned} err &:= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_T[L(\hat{f}, x^{(i)})] \\ &= \sigma^2 + \frac{1}{n} \sum_{i=1}^n \text{Var}(\hat{f}(x^{(i)})) + \frac{1}{n} \sum_{i=1}^n (\mathbb{E}_T[\hat{f}(x^{(i)})] - \beta^t x^{(i)})^2 \\ &=: \sigma^2 + \underbrace{\text{varijanca} + \text{pristranost}^2}_{=: \text{MSE}}, \end{aligned}$$

kada je \hat{f} dobivena (i) linearnom regresijom koristeći metodu najmanjih kvadrata, (ii) ridge regresijom, te (iii) lasso metodom.

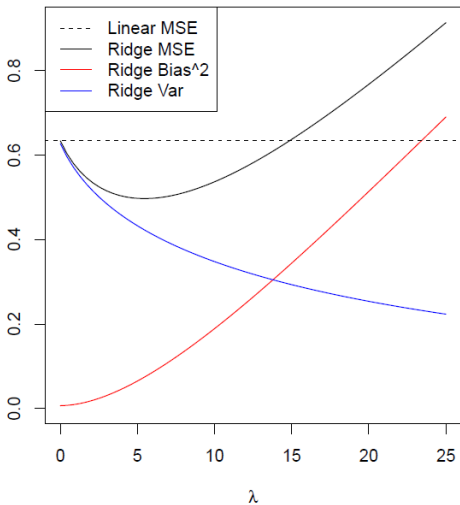
Glavna poruka je da iako ridge regresija i lasso uvode pristranost, u nekim slučajevima značajno smanjuju varijancu, a samim time i ukupnu grešku.



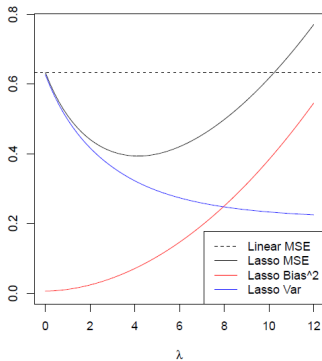
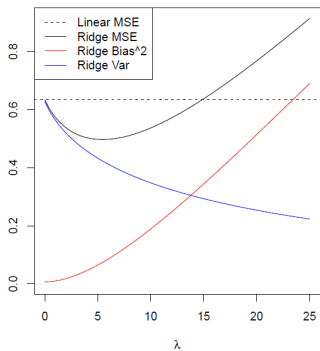
Slika: β sadrži 10 velikih (između 0.5 i 1) te 20 malih (između 0 i 0.3), koeficijenata. (Ova i sve ostale slike preuzete od Ryan Tibshirani, kolegij Data mining na CMU)



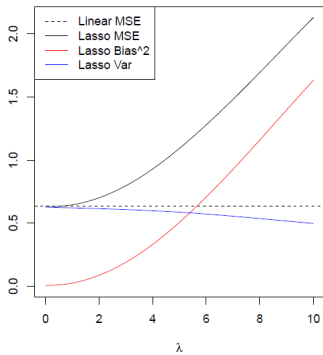
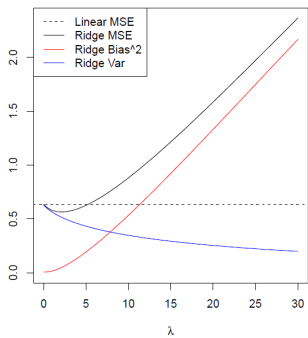
Slika: β sadrži 30 velikih (između 0.5 i 1) koeficijenata. MSE za LS metodu se nije promijenio jer ne ovisi o veličini koeficijenata – vidi (5.4) s predavanja. Ridge regresija ipak daje malo bolji rezultat za male vrijednosti od λ , a za velike vrijednosti pristranost je značajna.



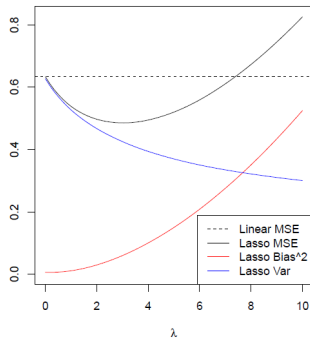
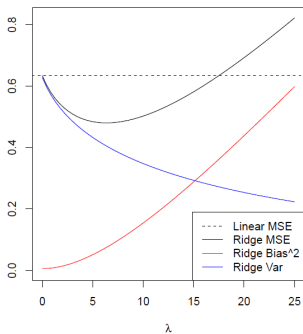
Slika: β sadrži 10 velikih koeficijenata, a ostalih 20 su jednaki 0.



Slika: β sadrži 10 velikih koeficijenata, a ostalih 20 su jednaki 0. Očekivano, lasso daje bolje rezultate od ridge regresije.



Slika: β sadrži 30 velikih (između 0.5 i 1) koeficijenata. Kod Lasso metode povećanje u pristranosti nije kompenzirano dovoljnim smanjenjem varijance (za razliku od ridge regresije).



Slika: β sadrži 10 velikih (između 0.5 i 1) te 20 malih (između 0 i 0.3), koeficijenata. Slični rezultati.