

Statističko učenje

Ridge regresija

listopad, 2022

Model

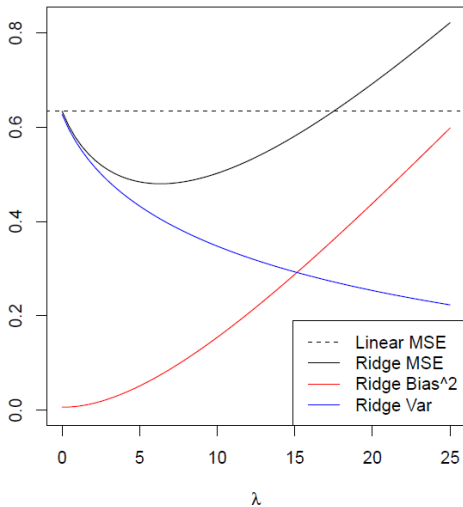
Pretpostavljamo da je $Y_i = \beta^t x^{(i)} + \epsilon$, $i = 1, \dots, n$ uz $n = 50$, pri čemu je

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

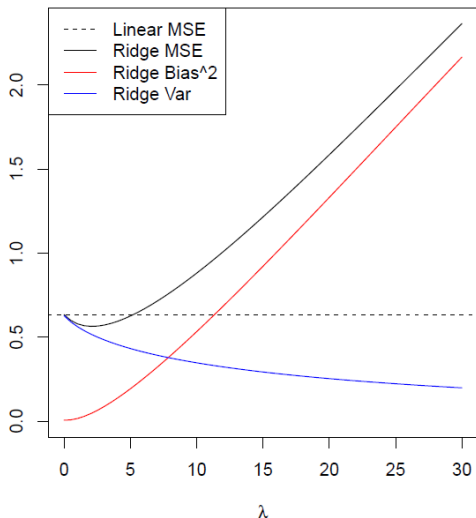
– ϵ_i njd iz $N(0, 1)$ (dakle, $\sigma = 1$);

Dakle, p je velik u odnosu na n .

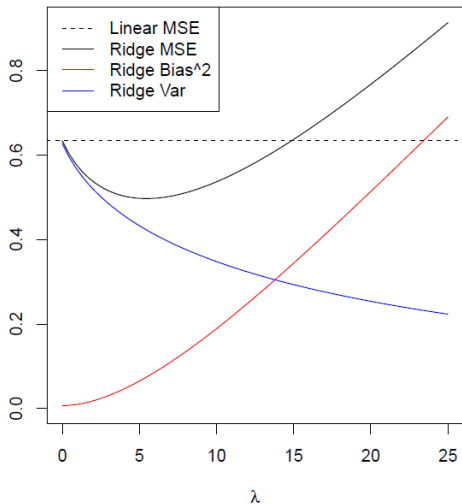
Simulacijom ćemo ispitati kako se ponaša očekivana greška unutar uzorka ("in-sample error") kada je \hat{f} dobivena (i) linearnom regresijom koristeći metodu najmanjih kvadrata, te (ii) ridge regresijom.



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. 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.