Stat 566
5-28-15

<u>Ridge regression</u> Hoerl & Kennard (1970)

①

Instead of solving $X'X\beta = X'Y$, $\hat\beta = (X'X)^{-1}X'Y$

they solved $(X'X + kI)\beta = X'Y$

$$\hat\beta_R = (X'X + kI)^{-1}X'Y$$

Note $\hat\beta_R = (X'X + kI)^{-1}X'X\hat\beta$ (since $\hat\beta$ solves $X'X\beta = X'Y$)

$$= Z_k\hat\beta$$

②

$$E(\hat\beta_R) = E(Z_k\hat\beta) = Z_kE(\hat\beta) = Z_k\beta$$

$$\text{Bias}(\hat\beta_R) = E(\hat\beta_R) - \beta = Z_k\beta - \beta = (Z_k - I)\beta$$

$$MSE(\hat\beta_R) = E\left[(\hat\beta_R - \beta)'(\hat\beta_R - \beta)\right]$$

$$= E\left[(\underbrace{\hat\beta_R - Z_k\beta}_{} + \underbrace{Z_k\beta - \beta}_{})'(\underbrace{\hat\beta_R - Z_k\beta}_{} + \underbrace{Z_k\beta - \beta}_{})\right]$$

$$= E\left[(\hat\beta_R - Z_k\beta)'(\hat\beta_R - Z_k\beta)\right] + \cancel{E\left[(\hat\beta_R - Z_k\beta)'(Z_k\beta - \beta)\right]}^{\,0}$$

$$+ \cancel{E\left[(Z_k\beta - \beta)'(\hat\beta_R - Z_k\beta)\right]}^{\,0} + E\left[(Z_k\beta - \beta)'(Z_k\beta - \beta)\right]$$

③

$$= E\left[(\hat{\beta}_R - Z_k\beta)'(\hat{\beta}_R - Z_k\beta)\right] + (Z_k\beta - \beta)'(Z_k\beta - \beta)$$

① ②

② : $\left[(Z_k - I)\beta\right]'(Z_k - I)\beta = \beta'(Z_k - I)'(Z_k - I)\beta$

$$= \beta'\left[(X'X + kI)^{-1}X'X - I\right]'\left[(X'X + kI)^{-1}X'X - I\right]\beta$$

$$= \beta'\left[(X'X+kI)^{-1}\left[\underbrace{X'X - (X'X+kI)}_{-kI}\right]\right]'\left[(X'X+kI)^{-1}\left[\underbrace{X'X - (X'X+kI)}_{-kI}\right]\right]\beta$$

$$= k^2\beta'(X'X+kI)^{-2}\beta$$

④

① : $E\left[(\hat{\beta}_R - Z_k\beta)'(\hat{\beta}_R - Z_k\beta)\right]$

$$= E\left[tr\left((\hat{\beta}_R - Z_k\beta)'(\hat{\beta}_R - Z_k\beta)\right)\right]$$

$$= E\left[tr\left((\hat{\beta}_R - Z_k\beta)(\hat{\beta}_R - Z_k\beta)'\right)\right] \qquad (tr\,AB = tr\,BA)$$

$$= tr\,E\left[(\hat{\beta}_R - Z_k\beta)(\hat{\beta}_R - Z_k\beta)'\right]$$

$$= tr\,Var(\hat{\beta}_R) = tr\,Var(Z_k\hat{\beta})$$

$$= tr\left(Z_k\,Var(\hat{\beta})\,Z_k'\right)$$

$$= tr\left(Z_k\,\sigma^2(X'X)^{-1}Z_k'\right)$$

$$= \sigma^2 \operatorname{tr}\left((X'X+kI)^{-1} \cancel{X'X}\, \cancel{(X'X)^{-1}}\, X'X\, (X'X+kI)^{-1} \right)$$ (5)

$$(1) \quad = \sigma^2 \operatorname{tr}\left(X'X\,(X'X+kI)^{-2} \right)$$

Write $X'X$ in its spectral decomposition

$$X'X = P\Lambda P'$$

$$\begin{aligned} X'X + kI &= P\Lambda P' + kPP' \\ &= P(\Lambda + kI)P' \end{aligned}$$

$$(1) = \sigma^2 \operatorname{tr}\left(P\Lambda P' P(\Lambda+kI)^{-2} P' \right)$$

(6)

$$\begin{aligned} &= \sigma^2 \operatorname{tr}\left(P\Lambda(\Lambda+kI)^{-2} P' \right) \\ &= \sigma^2 \operatorname{tr}\left(\Lambda(\Lambda+kI)^{-2} \cancel{P'P} \right) \\ &= \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i+k)^2} \end{aligned}$$

Now $\mathrm{MSE}(\hat{\beta}_R) = \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i+k)^2} + k^2 \beta'(X'X+kI)^{-2}\beta$

Hoerl & Kennard proved that

(1) $\exists k$ such that $\mathrm{MSE}(\hat{\beta}_R) < \sum_{i=1}^{p} V(\hat{\beta}_i)$

(2) as $k \uparrow$ $R^2 \downarrow$

(7)

H & K suggest using $k = \dfrac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$,

where p, $\hat{\sigma}^2 = MSE$, $\hat{\beta}$ all come from the OLS solution.

R: lm.ridge

SAS: PROC REG (ridge =)

HW#8 Use B21.xlsx Run OLS + ridge regression + compare results

(8)

Final exam: Take-home only.

Handed out Thursday June 4

due Tues June 9 at 5pm