Ridge regression

Y centered, $x_1, \dots, x_p$ centered & scaled, β $p \times 1$ estimated via:

$$\min_{\beta} \; (y - X\beta)^T (y - X\beta) + \lambda \beta^T \beta \qquad \lambda \geq 0$$

$\lambda \|\beta\|_2^2$

$\lambda \sum_{j=1}^{p} \beta_j^2$

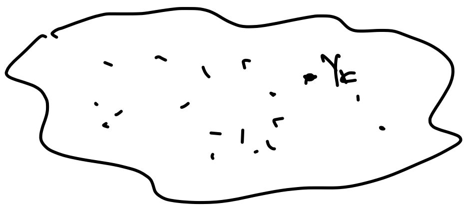
$$\Rightarrow \quad \hat{\beta}_{ridge} = (X^T X + \lambda I)^{-1} X^T y$$

- $\hat{\beta}_{ridge}$ is biased, but can have a lower MSE than OLS
- How to choose smoothing parameter λ?

Leave-one-out cross-validation (LOOCV)

(0) Fix $\lambda \geq 0$

(1)



Fit the model with all n data points & predict the ys:

$$\hat{y} = \begin{pmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{pmatrix} = X\hat{\beta}_{ridge}$$

(2) For $k = 1, 2, \dots, n$, remove y_k from data, fit model on the remaining $(n-1)$ points, & predict y_k: $\hat{y}_{-k}$, yields $\hat{y}_{-1}, \dots, \hat{y}_{-n}$

fit based on these, & predict y_k

$$\hat{y}_{-k} = \underline{x}_k^T \hat{\beta}_{ridge}$$

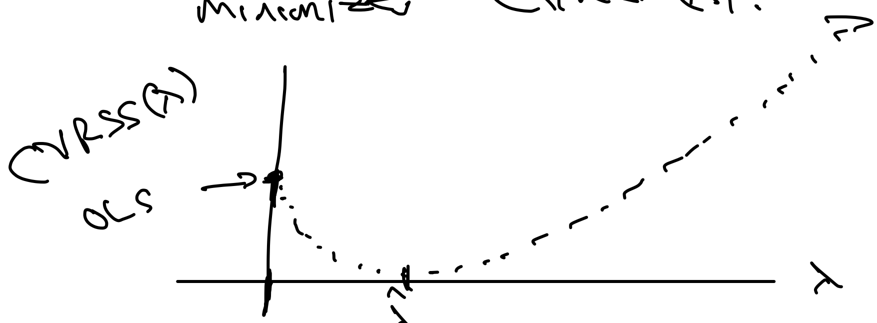
(3) Store <u>cross-validated residuals:</u>

$$Y_k - \hat{Y}_{-k}, \quad k = 1, 2, \ldots, n$$

(4) Calculate CV RSS

$$\sum_{k=1}^{n} \left(Y_k - \hat{Y}_{-k} \right)^2 = CVRSS(\lambda)$$

(5) Do 1-4 for many different λs, find λ that minimizes $CVRSS(\lambda)$.



$\boxed{\text{Problem}}$ Each $\hat{Y}_{-k}$ requires refitting the model, + have to do it for many λs

$\boxed{\text{Solution}}$ The hat matrix for ridge is

$$X(X^TX+\lambda I)^{-1}X^T = H(\lambda)$$

It turns out

$$Y_k - \hat{Y}_{-k} = \frac{Y_k - \hat{Y}_k}{1-H(\lambda)_{kk}}$$

$\rightarrow$ fit based on all data

Thus, we choose λ to minimize

$$\sum_{k=1}^{n}\left(\frac{Y_k-\hat{Y}_k}{1-H(\lambda)_{kk}}\right)^2$$

$\Rightarrow$ don't have to refit model for each k.

Alternative: generalized cross-validation (GCV)

minimize

$$\sum_{k=1}^{n} \left(\frac{y_k - \hat{y}_k}{1 - \frac{\operatorname{tr} H(\lambda)}{n}} \right)^2$$

GCV uses average leverage rather than individual leverage.