

Solution proposals for STK2100

Vera Kvisgaard

25-02-25

ESL Ex. 3.29

The first order conditions for the ridge solutions are given by

$$(X^\top X + \lambda I)\hat{\beta}^{\text{ridge}} = X^\top y$$

such that

$$\hat{\beta}^{\text{ridge}} = (X^\top X + \lambda I)^{-1} X^\top y$$

Note that the first equation defines a system of p equations and p unknowns (the p coefficients $\hat{\beta}^{\text{ridge}}$).

In the univariate case, where $p = 1$ and $X = X_1$ is a $n \times 1$ matrix, $X^\top X$ is a scalar.

Now, if we add an identical variable $X_2 = X_1$ to the model, and set $X = [X_1 \ X_1]$, then

$$X^\top X = \begin{pmatrix} X_1^\top X_1 & X_1^\top X_1 \\ X_1^\top X_1 & X_1^\top X_1 \end{pmatrix}$$

The first order conditions then becomes

$$\begin{pmatrix} X_1^\top X_1 - \lambda & X_1^\top X_1 \\ X_1^\top X_1 & X_1^\top X_1 - \lambda \end{pmatrix} \begin{pmatrix} \hat{\beta}_1^{\text{ridge}} \\ \hat{\beta}_2^{\text{ridge}} \end{pmatrix} = \begin{pmatrix} X_1^\top y \\ X_1^\top y \end{pmatrix}.$$

i.e. a system of two equations with two unknowns. Note that the j th equation for $j = 1, 2$ can be written as

$$X_1^\top X_1 \hat{\beta}_1^{\text{ridge}} + X_1^\top X_1 \hat{\beta}_2^{\text{ridge}} - \lambda \hat{\beta}_j^{\text{ridge}} = X_1^\top y$$

Analogously, if X consist of p columns all identical to X_1 , the first order conditions are p equations on the form

$$\sum_{m=1}^p X_1^\top X_1 \hat{\beta}_m^{\text{ridge}} - \lambda \hat{\beta}_j^{\text{ridge}} = X_1^\top y$$

for each $j = 1, \dots, p$.

Subtract equation $j + 1$ from equation j ,

$$\sum_{m=1}^p X_1^\top X_1 \hat{\beta}_m^{\text{ridge}} - \lambda \hat{\beta}_j^{\text{ridge}} - \left(\sum_{m=1}^p X_1^\top X_1 \hat{\beta}_m^{\text{ridge}} - \lambda \hat{\beta}_{j+1}^{\text{ridge}} \right) = 0,$$

to see that

$$\lambda \hat{\beta}_j^{\text{ridge}} = \lambda \hat{\beta}_{j+1}^{\text{ridge}}.$$

Hence, even without the explicit solution, we can state that the coefficients have the same value. To find the value of β_j^{ridge} , note that, since all coefficients are equal,

$$\sum_{m=1}^p X_1^\top X_1 \hat{\beta}_m^{\text{ridge}} = p X_1^\top X_1 \hat{\beta}_j^{\text{ridge}}$$

and

$$(p X_1^\top X_1 - \lambda) \hat{\beta}_j^{\text{ridge}} = X_1^\top y$$

which gives

$$\hat{\beta}_j^{\text{ridge}} = \frac{X_1^\top y}{(p X_1^\top X_1 - \lambda)}.$$