MA8701 Advanced methods in statistical inference and learning

Solutions to Exercise on Ridge Regression

Mette Langaas IMF/NTNU

29 January, 2023

Contents

Ridge regression exercise

This problem is taken, with permission from Wessel van Wieringen, from a course in High-dimensional data analysis at Vrije University, Amsterdam, The Netherlands.

a)

First calculate $\mathbf{X}^T\mathbf{X}$ and $\mathbf{X}^T\mathbf{Y}$. These are given by:

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} 8 & 0 \\ 0 & 16 \end{pmatrix} \qquad \mathbf{X}^T\mathbf{Y} = \begin{pmatrix} 320 \\ 35 \end{pmatrix}.$$

To penalize only the slope parameter add:

$$\lambda = \begin{pmatrix} 0 & 0 \\ 0 & \lambda \end{pmatrix}$$

to $\mathbf{X}^T\mathbf{X}$ in the normal equations. This leads to following ridge estimate:

$$\begin{aligned}
\hat{\beta}_{ridge} &= (\mathbf{X}^T\mathbf{X} + \lambda)^{-1}\mathbf{X}^T\mathbf{Y} \\
&= \begin{pmatrix} 8 & 0 \\ 0 & 16+\lambda \end{pmatrix}^{-1} \begin{pmatrix} 320 \\ 35 \end{pmatrix} \\
&= \begin{pmatrix} 1/8 & 0 \\ 0 & 1/(16+\lambda) \end{pmatrix} \begin{pmatrix} 320 \\ 35 \end{pmatrix} \\
&= \begin{pmatrix} 40 \\ 35/(16+\lambda) \end{pmatrix}.
\end{aligned}$$

Choosing $\lambda = 4$ yields the reported estimates.

b)

A projection matrix $\mathbf{Q}$ would satisfy $\mathbf{Q} = \mathbf{Q}^2$. Verify:

$$
\begin{aligned}
\mathbf{Q}^2 &= \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_p)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_p)^{-1}\mathbf{X}^{\mathrm{T}} \\
&= \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_p)^{-1}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_p)(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_p)^{-1}\mathbf{X}^{\mathrm{T}} \\
&\quad -\lambda\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_p)^{-1}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_p)^{-1}\mathbf{X}^{\mathrm{T}} \\
&= \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_p)^{-1}\mathbf{X}^{\mathrm{T}} - \lambda\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_p)^{-2}\mathbf{X}^{\mathrm{T}} \\
&= \mathbf{Q} - \lambda\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_p)^{-2}\mathbf{X}^{\mathrm{T}} \\
&\neq \mathbf{Q}.
\end{aligned}
$$

Hence, $\mathbf{Q}$ is not a projection matrix.

c)

The ridge fit is given by $\hat{\mathbf{Y}}(\lambda) = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_p)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} = \mathbf{Q}\mathbf{Y}$ and the associated residuals by: $\hat{\varepsilon}(\lambda) = \mathbf{Y} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_p)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} = [\mathbf{I}_p - \mathbf{Q}]\mathbf{Y}$. Would the residual and the fit be orthogonal, their inner product becomes zero: $\langle \hat{\mathbf{Y}}(\lambda), \hat{\varepsilon}(\lambda) \rangle = 0$. Verify:

$$
\begin{aligned}
\langle \hat{\mathbf{Y}}(\lambda), \hat{\varepsilon}(\lambda) \rangle &= [\hat{\mathbf{Y}}(\lambda)]^{\mathrm{T}}\hat{\varepsilon}(\lambda) \\
&= [\mathbf{Q}\mathbf{Y}]^{\mathrm{T}}[\mathbf{I}_p - \mathbf{Q}]\mathbf{Y} \\
&= \mathbf{Y}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}(\mathbf{I}_p - \mathbf{Q})\mathbf{Y} \\
&= \mathbf{Y}^{\mathrm{T}}(\mathbf{Q}^{\mathrm{T}} - \mathbf{Q}^{\mathrm{T}}\mathbf{Q})\mathbf{Y} \\
&= \mathbf{Y}^{\mathrm{T}}(\mathbf{Q} - \mathbf{Q}^2)\mathbf{Y},
\end{aligned}
$$

where we have used the symmetry of $\mathbf{Q}$. Invoke the result of b) to conclude.