

FYS-STK4155 Week 36

Edvard B. Rørnes, Isak O. Rukan, Anton A. Brekke

September 16, 2024

Exercise 1

Show that the expectation value and variance of \mathbf{y} is

$$\mathbb{E}(y_i) = \sum_j x_{ij}\beta_j = \mathbf{X}_{i,*}\boldsymbol{\beta}, \quad \text{Var}(y_i) = \sigma^2$$

where \mathbf{y} is defined by $\mathbf{y} = f(\mathbf{x}) + \boldsymbol{\varepsilon}$. Here $\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$ is a normal distributed error and $f(\mathbf{x})$ is the approximated function given our model $\hat{\mathbf{y}}$ obtained by minimizing $(\mathbf{y} - \hat{\mathbf{y}})^2$ with $\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta}$. With the OLS expression for $\hat{\boldsymbol{\beta}}$ also show that

$$\mathbb{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

and

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^T \mathbf{X})^{-1}$$

Solution:

Trivially $\mathbb{E}(\varepsilon_i) = 0$ from its definition. Thus from the definition of \mathbf{y} we have that

$$\mathbb{E}(y_i) = \mathbb{E}(f(x_i)) = \mathbf{X}_{i,*}\boldsymbol{\beta}$$

Similarly the variance is given by

$$\begin{aligned} \text{Var}(y_i) &= \mathbb{E}\{[y_i - \mathbb{E}(y_i)]^2\} = \mathbb{E}\{(\mathbf{X}_{i,*}\boldsymbol{\beta} + \varepsilon_i)^2\} - (\mathbf{X}_{i,*}\boldsymbol{\beta})^2 \\ &= (\mathbf{X}_{i,*}\boldsymbol{\beta})^2 + \mathbb{E}(\varepsilon_i^2) + 2\mathbb{E}(\varepsilon_i)\mathbf{X}_{i,*}\boldsymbol{\beta} - (\mathbf{X}_{i,*}\boldsymbol{\beta})^2 \\ &= \text{Var}(\varepsilon_i^2) = \sigma^2 \end{aligned}$$

The optimal parameters $\boldsymbol{\beta}$ for OLS are given by

$$\hat{\boldsymbol{\beta}}_{\text{OLS}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

which yields the expectation value

$$\mathbb{E}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) = \mathbb{E}[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}] = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbb{E}[\mathbf{y}] = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \boldsymbol{\beta}.$$

and the variance

$$\begin{aligned} \text{Var}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) &= \mathbb{E}\{[\boldsymbol{\beta} - \mathbb{E}(\boldsymbol{\beta})][\boldsymbol{\beta} - \mathbb{E}(\boldsymbol{\beta})]^T\} \\ &= \mathbb{E}\{[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} - \boldsymbol{\beta}][(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} - \boldsymbol{\beta}]^T\} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbb{E}\{\mathbf{y} \mathbf{y}^T\} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} - \boldsymbol{\beta} \boldsymbol{\beta}^T \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T [\mathbf{X} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{X}^T + \sigma^2 \mathbf{I}] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} - \boldsymbol{\beta} \boldsymbol{\beta}^T \\ &= \boldsymbol{\beta} \boldsymbol{\beta}^T + \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} - \boldsymbol{\beta} \boldsymbol{\beta}^T = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \end{aligned}$$

Exercise 2

Show that

$$\begin{aligned}\mathbb{E}(\hat{\beta}_{\text{Ridge}}) &= (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{X}) \boldsymbol{\beta} \\ \text{Var}(\hat{\beta}_{\text{Ridge}}) &= \sigma^2 (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{X} \{(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1}\}^T\end{aligned}$$

Solution:

We have that

$$\hat{\beta}_{\text{Ridge}} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}$$

thus

$$\begin{aligned}\mathbb{E}(\hat{\beta}_{\text{Ridge}}) &= \mathbb{E}[(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}] \\ &= (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbb{E}[\mathbf{y}] \\ &= (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}\end{aligned}$$

and

$$\begin{aligned}\text{Var}(\hat{\beta}_{\text{Ridge}}) &= \mathbb{E}\{[\boldsymbol{\beta}_{\text{Ridge}} - \mathbb{E}(\boldsymbol{\beta})][\boldsymbol{\beta}_{\text{Ridge}} - \mathbb{E}(\boldsymbol{\beta})]^T\} \\ &= \mathbb{E}\{[(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y} - \boldsymbol{\beta}][(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y} - \boldsymbol{\beta}]^T\} \\ &= (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbb{E}\{\mathbf{y} \mathbf{y}^T\} \mathbf{X} [(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1}]^T - \boldsymbol{\beta} \boldsymbol{\beta}^T \\ &= (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T [\mathbf{X} \boldsymbol{\beta} \boldsymbol{\beta}^T \mathbf{X}^T + \sigma^2 \mathbf{I}] \mathbf{X} [(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1}]^T - \boldsymbol{\beta} \boldsymbol{\beta}^T \\ &= \boldsymbol{\beta} \boldsymbol{\beta}^T + \sigma^2 (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{X} [(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1}]^T - \boldsymbol{\beta} \boldsymbol{\beta}^T \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{X} [(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1}]^T.\end{aligned}$$