# 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 $\boldsymbol{y}$ is

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

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

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

and

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

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

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

Similarly the variance is given by

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

The optimal parameters $\beta$ for OLS are given by

$$\hat{\boldsymbol{\beta}}_{\mathrm{OSL}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

which yields the expectation value

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

and the variance

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

## Exercise 2

Show that

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

**Solution:**
We have that

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

thus

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

and

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