

# Exercises week 37

## Contents

- [Overarching aims of the exercises this week](#)
- [Exercise 1: Expectation values for ordinary least squares expressions](#)
- [Exercise 2: Expectation values for Ridge regression](#)

September 11-15, 2023

Date: **Deadline is Sunday September 17 at midnight**

## Overarching aims of the exercises this week

This exercise deals with various mean values and variances in linear regression method (here it may be useful to look up chapter 3, equation (3.8) of [Trevor Hastie, Robert Tibshirani, Jerome H. Friedman, The Elements of Statistical Learning, Springer](#)). The exercise is also a part of project 1 and can be reused in the theory part of the project.

For more discussions on Ridge regression and calculation of expectation values, [Wessel van Wieringen's](#) article is highly recommended.

The assumption we have made is that there exists a continuous function  $f(\mathbf{x})$  and a normal distributed error  $\varepsilon \sim N(0, \sigma^2)$  which describes our data

$$\mathbf{y} = f(\mathbf{x}) + \varepsilon$$

We then approximate this function  $f(\mathbf{x})$  with our model  $\tilde{\mathbf{y}}$  from the solution of the linear regression equations (ordinary least squares OLS), that is our function  $f$  is approximated by  $\tilde{\mathbf{y}}$  where we minimized  $(\mathbf{y} - \tilde{\mathbf{y}})^2$ , with

$$\tilde{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta}.$$

The matrix  $\mathbf{X}$  is the so-called design or feature matrix.

## Exercise 1: Expectation values for ordinary least squares expressions

Show that the expectation value of  $\mathbf{y}$  for a given element  $i$

$$\textcircled{1} \mathbb{E}(y_i) = \sum_j x_{ij}\beta_j = \mathbf{X}_{i,*}\boldsymbol{\beta},$$

and that its variance is

$$\textcircled{2} \text{Var}(y_i) = \sigma^2.$$

Hence,  $y_i \sim N(\mathbf{X}_{i,*}\boldsymbol{\beta}, \sigma^2)$ , that is  $\mathbf{y}$  follows a normal distribution with mean value  $\mathbf{X}\boldsymbol{\beta}$  and variance  $\sigma^2$ .

# EXERCISE 1

$$\begin{aligned} \textcircled{1} E[y_i] &= E\left[\sum_{j=0}^{n-1} x_{ij} \beta_j + \varepsilon_i\right] = E\left[\sum_{j=0}^{n-1} x_{ij} \beta_j\right] + E[\varepsilon_i] = E\left[\sum_{j=0}^{n-1} x_{ij} \beta_j\right] + 0 = \\ &= E[X_i * \beta] = X_i * \beta \end{aligned}$$

$$\begin{aligned} \textcircled{2} \text{Var}(y_i) &= E[y_i^2] - E[y_i]^2 \quad \text{where } y_i = X_i * \beta + \varepsilon_i \\ E[y_i^2] &= E[(X_i * \beta + \varepsilon_i)^2] = E[(X_i * \beta)^2] + E[\varepsilon_i^2] + 2E[X_i * \beta \varepsilon_i] \\ &\quad \begin{array}{l} \nearrow E[\varepsilon_i^2] = \text{Var}(\varepsilon_i) + E[\varepsilon_i] = \sigma^2 + 0 = \sigma^2 \\ \phantom{\nearrow} = (X_i * \beta)^2 + \sigma^2 + 0 = (X_i * \beta)^2 + \sigma^2 \end{array} \\ \text{Var}(y_i) &= (X_i * \beta)^2 + \sigma^2 - (X_i * \beta)^2 = \sigma^2 \end{aligned}$$

$$\textcircled{3} E[\hat{\beta}] = E[(X^T X)^{-1} X^T y] = (X^T X)^{-1} X^T \underbrace{E[y]}_{= X\beta} = \underbrace{(X^T X)^{-1} X^T X}_{= I} \beta = \beta$$

$$\begin{aligned} \textcircled{4} \text{Var}(\hat{\beta}) &= \text{Var}[(X^T X)^{-1} X^T y] = (X^T X)^{-1} X^T \underbrace{\text{Var}(y)}_{= \sigma^2 I} (X^T X)^{-1} X^T = \sigma^2 \underbrace{(X^T X)^{-1} X^T X}_{= I} (X^T X)^{-1} = \\ &= \sigma^2 (X^T X)^{-1} \end{aligned}$$

With the OLS expressions for the optimal parameters  $\hat{\beta}$  show that

$$\textcircled{3} \mathbb{E}(\hat{\beta}) = \beta.$$

Show finally that the variance of  $\beta$  is

$$\textcircled{4} \text{Var}(\hat{\beta}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}.$$

We can use the last expression when we define a [so-called confidence interval](#) for the parameters  $\beta$ . A given parameter  $\beta_j$  is given by the diagonal matrix element of the above matrix.

## Exercise 2: Expectation values for Ridge regression

Show that

$$\textcircled{1} \mathbb{E}[\beta^{\text{Ridge}}] = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}_{pp})^{-1} (\mathbf{X}^T \mathbf{X}) \beta^{\text{OLS}}.$$

We see clearly that  $\mathbb{E}[\beta^{\text{Ridge}}] \neq \beta^{\text{OLS}}$  for any  $\lambda > 0$ .

Show also that the variance is

$$\textcircled{2} \text{Var}[\beta^{\text{Ridge}}] = \sigma^2 [\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}]^{-1} \mathbf{X}^T \mathbf{X} \{[\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}]^{-1}\}^T,$$

and it is easy to see that if the parameter  $\lambda$  goes to infinity then the variance of Ridge parameters  $\beta$  goes to zero.

---

By Morten Hjorth-Jensen

© Copyright 2021.

## EXERCISE 2

$$\textcircled{1} \mathbb{E}[\beta^{\text{ridge}}] = \mathbb{E}[(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}] = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbb{E}[\mathbf{y}] = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{X} \beta^{\text{OLS}}$$

$$\begin{aligned} \textcircled{2} \text{Var}(\beta^{\text{ridge}}) &= \text{Var}((\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}) = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \underbrace{\text{Var}(\mathbf{y})}_{=\sigma^2 \mathbf{I}} \cdot [(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T]^T \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{X} [(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1}]^T \end{aligned}$$