Regression analysis and resampling methods

Johan Mylius Kroken[1,2], and Nanna Bryne[1,2]

[1] Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, N-0315 Oslo, Norway
[2] Center for Computational Science, University of Oslo, Norway

September 11, 2022

ABSTRACT

The ...

1. Introduction

2. Theory

* Presisér indeks-notasjon! *

We assume the data $y_i = f(x_i) + \varepsilon_i \in \mathbb{R}^n$ where $f(x_i) \in \mathbb{R}^n$ is a continous function and $\varepsilon_i \in \mathbb{R}^n$ is a normally distributed noise. We let $f(x_i) = X_{ij}\beta_j$ where $X_{ij} \in \mathbb{R}^{n\times p}$ is the design matrix s.t.

$$X_{ij} = (x_i)^{j-1}, \quad i = 1, \dots, n \quad j = 1, \dots, p,$$

and $\beta_j \in \mathbb{R}^p$ are the unknown parameters to be determined. The integers n and p represent the number of data points and features, respectively.

$$\begin{aligned} y_i &= f(x_i) + \varepsilon_i = X_{ij}\beta_j + \varepsilon_i \\ \tilde{y}_i &= X_{ij}\hat{\beta}_j \end{aligned}$$

$$\mathbb{E}\big[X_{ij}\beta_j\big] \overset{\text{non-stochastic}}{=} X_{ij}\beta_j$$

$$\mathbb{E}\big[\varepsilon_i\big] \overset{\text{per def.}}{=} 0$$

$$\begin{aligned} \mathbb{E}\big[y_i\big] &= \mathbb{E}\big[X_{ij}\beta_j + \varepsilon_i\big] \\ &= \mathbb{E}\big[X_{ij}\beta_j\big] + \mathbb{E}\big[\varepsilon_i\big] \\ &= X_{ij}\beta_j \end{aligned}$$

$$\begin{aligned} \mathrm{Var}\big[y_i\big] &= \mathbb{E}\big[y_i y_i^\intercal\big] - \mathbb{E}\big[y_i\big]\mathbb{E}\big[y_i^\intercal\big] \\ &= \mathbb{E}\big[X_{ij}\beta_j\beta_j^\intercal X_{ij}^\intercal + \ X_{ij}\beta_j\varepsilon_i^\intercal + \varepsilon_i\beta_j^\intercal X_{ij}^\intercal + \varepsilon_i\varepsilon_i^\intercal\big] \\ &\quad - X_{ij}\beta_j\beta_j^\intercal X_{ij}^\intercal \\ &= X_{ij}\beta_j\beta_j^\intercal X_{ij}^\intercal + \mathbb{E}\big[\varepsilon_i\varepsilon_i^\intercal\big] - X_{ij}\beta_j\beta_j^\intercal X_{ij}^\intercal \\ &= \sigma^2 \end{aligned}$$

...

$$\begin{aligned} \mathbb{E}\big[\hat{\beta}\big] &= \mathbb{E}\big[(X^\intercal X)^{-1}X^\intercal y\big] \\ &= (X^\intercal X)^{-1}X^\intercal\mathbb{E}\big[y\big] \\ &= (X^\intercal X)^{-1}X^\intercal X\beta \\ &= \beta \end{aligned}$$

$$\begin{aligned} \mathrm{Var}\big[\hat{\beta}\big] &= \mathbb{E}\big[\hat{\beta}\hat{\beta}^\intercal\big] - \mathbb{E}\big[\hat{\beta}\big]\mathbb{E}\big[\hat{\beta}^\intercal\big] \\ &= \mathbb{E}\big[(X^\intercal X)^{-1}X^\intercal yy^\intercal X((X^\intercal X)^{-1})^\intercal\big] - \beta\beta^\intercal \\ &= (X^\intercal X)^{-1}X^\intercal\mathbb{E}\big[yy^\intercal\big]X(X^\intercal X)^{-1} - \beta\beta^\intercal \\ &\overset{*}{=} (X^\intercal X)^{-1}X^\intercal(X\beta\beta^\intercal + \mathbb{I}\sigma^2)X(X^\intercal X)^{-1} \\ &= \beta\beta^\intercal + (X^\intercal X)^{-1}X^\intercal\sigma^2X(X^\intercal X)^{-1} - \beta\beta^\intercal \\ &= \sigma^2(X^\intercal X)^{-1} \end{aligned}$$

$$\begin{aligned} \mathbb{E}\big[yy^\intercal\big] &= \mathbb{E}\big[(X\beta + \varepsilon)(X\beta + \varepsilon)^\intercal\big] \\ &= \mathbb{E}\big[X\beta\beta^\intercal X^\intercal + X\beta\varepsilon^\intercal + \varepsilon\beta^\intercal X^\intercal + \varepsilon\varepsilon^\intercal\big] \\ &= X\beta\beta^\intercal X^\intercal + \mathbb{I}\sigma^2 \end{aligned}$$

3. Analysis

4. Conclusion