*Astronomy & Astrophysics* manuscript no. project1
September 15, 2022


# 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 15, 2022

**ABSTRACT**

The ...

## 1. Introduction

## 2. Theory

* DECIDE ON z vs. y

$$\mathrm{MSE}\big(\mathbf{y},\tilde{\mathbf{y}}\big) = \frac{1}{n}(\mathbf{y}-\tilde{\mathbf{y}})^2 = \frac{1}{n}\sum_{i=0}^{n-1}(y_i-\tilde{y}_i)^2 \tag{1}$$

$$R^2\big(\mathbf{y},\tilde{\mathbf{y}}\big) = 1-\frac{(\mathbf{y}-\tilde{\mathbf{y}})^2}{(\mathbf{y}(1-\bar{y}))^2} = 1-\frac{\sum_{i=0}^{n-1}(y_i-\tilde{y}_i)^2}{\sum_{i=0}^{n-1}(y_i-\bar{y})^2} \tag{2}$$

$$\bar{y} = \frac{1}{n}\sum_{i=0}^{n-1}y_i$$

### *2.1. Regression*

We assume the vector $\mathbf{y}\in\mathbb{R}^n$ consisting of observed values $y_i$ to take the form $\mathbf{y}=f(\mathbf{x})+\boldsymbol{\varepsilon}$ where $f(\mathbf{x})\in\mathbb{R}^n$ is a continous function and $\boldsymbol{\varepsilon}\in\mathbb{R}^n$ is a normally distributed noise. We approximate $f$ by $\tilde{\mathbf{y}}=X\boldsymbol{\beta}$, where $X\in\mathbb{R}^{n\times p}$ is the design matrix of $n$ row vectors $\mathbf{x}_i\in\mathbb{R}^p$, and $\boldsymbol{\beta}\in\mathbb{R}^p$ are the unknown parameters to be determined. The integers $n$ and $p$ represent the number of data points and features, respectively.

For an observed value $y_i$ we have $y_i=\mathbf{x}_i^\intercal\boldsymbol{\beta}+\varepsilon_i=(X\boldsymbol{\beta})_i+\varepsilon_i$[1]. The inner product $\mathbf{x}_i^\intercal\boldsymbol{\beta}$ is non-stochastic, hence

$$\mathbb{E}\big[(X\boldsymbol{\beta})_i\big] = (X\boldsymbol{\beta})_i$$

and since

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

we have the expectation value

$$\begin{aligned}\mathbb{E}\big[y_i\big] &= \mathbb{E}\big[(X\boldsymbol{\beta})_i+\varepsilon_i\big]\\ &= \mathbb{E}\big[(X\boldsymbol{\beta})_i\big]+\mathbb{E}\big[\varepsilon_i\big]\\ &= (X\boldsymbol{\beta})_i.\end{aligned}$$

[1] Can also write $X_{ij}\beta_j$, but we try to be consistent with our notation.

To find the variance, we need the expetation value of the outer product $\mathbf{y}\mathbf{y}^\intercal$,

$$\begin{aligned}\mathbb{E}\big[\mathbf{y}\mathbf{y}^\intercal\big] &= \mathbb{E}\big[(X\boldsymbol{\beta}+\boldsymbol{\varepsilon})(X\boldsymbol{\beta}+\boldsymbol{\varepsilon})^\intercal\big]\\ &= \mathbb{E}\big[X\boldsymbol{\beta}\boldsymbol{\beta}^\intercal X^\intercal + X\boldsymbol{\beta}\boldsymbol{\varepsilon}^\intercal + \boldsymbol{\varepsilon}\boldsymbol{\beta}^\intercal X^\intercal + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\intercal\big]\\ &= X\boldsymbol{\beta}\boldsymbol{\beta}^\intercal X^\intercal + \mathbb{I}\sigma^2.\end{aligned} \tag{3}$$

The variance becomes

$$\begin{aligned}\mathrm{Var}\big[y_i\big] &= \mathbb{E}\big[(\mathbf{y}\mathbf{y}^\intercal)_{ii}\big] - \Big(\mathbb{E}\big[y_i\big]\Big)^2\\ &= (X\boldsymbol{\beta})_i(X\boldsymbol{\beta})_i + \sigma^2 - (X\boldsymbol{\beta})_i(X\boldsymbol{\beta})_i\\ &= \sigma^2.\end{aligned}$$

Taking $\hat{\boldsymbol{\beta}}$ as the Ordinary Least Square (OLS) expression for the optimal parameter, i.e.

$$\hat{\boldsymbol{\beta}} = (X^\intercal X)^{-1}X^\intercal\mathbf{y},$$

we get the expected value

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

The variance is then

$$\begin{aligned}\mathrm{Var}\big[\hat{\boldsymbol{\beta}}\big] &= \mathbb{E}\big[\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^\intercal\big] - \mathbb{E}\big[\hat{\boldsymbol{\beta}}\big]\mathbb{E}\big[\hat{\boldsymbol{\beta}}^\intercal\big]\\ &= \mathbb{E}\big[(X^\intercal X)^{-1}X^\intercal\mathbf{y}\mathbf{y}^\intercal X((X^\intercal X)^{-1})^\intercal\big] - \boldsymbol{\beta}\boldsymbol{\beta}^\intercal\\ &= (X^\intercal X)^{-1}X^\intercal\mathbb{E}\big[\mathbf{y}\mathbf{y}^\intercal\big]X(X^\intercal X)^{-1} - \boldsymbol{\beta}\boldsymbol{\beta}^\intercal\\ &\overset{(3)}{=} (X^\intercal X)^{-1}X^\intercal(X\boldsymbol{\beta}\boldsymbol{\beta}^\intercal X^\intercal + \mathbb{I}\sigma^2)X(X^\intercal X)^{-1}\\ &= \boldsymbol{\beta}\boldsymbol{\beta}^\intercal + (X^\intercal X)^{-1}X^\intercal\sigma^2X(X^\intercal X)^{-1} - \boldsymbol{\beta}\boldsymbol{\beta}^\intercal\\ &= \sigma^2(X^\intercal X)^{-1}.\end{aligned}$$

#### 2.1.1. Ordinary Least Squares (OLS)

#### 2.1.2. Ridge regression

#### 2.1.3. Lasso regression

### *2.2. Resampling*

#### 2.2.1. Bootstrap method

#### 2.2.2. Cross-validation

## 3. Analysis

## 4. Conclusion