FACULTY OF SCIENCE, TECHNOLOGY AND MEDICINE

University of Luxembourg

Thesis for the Bachelor of Mathematics

High Dimensional Regression Models

by Joris LIMONIER

Supervisor: Mark PODOLSKIJ
Submitted: June 4, 2021

Contents

Chapter 1

Introduction

1.1 The first section

1.1.1 The first subsection

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\varepsilon} \tag{1.1}$$

We define $\hat{\beta}$ as follows

$$\hat{\beta} := \arg\min_{\beta} \left\{ \frac{\|\mathbf{Y} - \mathbf{X}\beta\|_2^2}{n} + \lambda\|\beta\|_1 \right\} \tag{1.2}$$

Lemma 1.1 (Basic Inequality)**.**

$$\frac{\|\mathbf{X}(\hat{\beta} - \beta^0)\|_2^2}{n} + \lambda\|\hat{\beta}\|_1 \leq 2\frac{\varepsilon^T\mathbf{X}(\hat{\beta} - \beta^0)}{n} + \lambda\|\beta^0\|_1$$

Proof. By definition of $\hat{\beta}$, we have that

$$\forall\beta \quad \frac{\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|_2^2}{n} + \lambda\|\hat{\beta}\|_1 \leq \frac{\|\mathbf{Y} - \mathbf{X}\beta\|_2^2}{n} + \lambda\|\beta\|_1$$

In particular for $\beta = \beta^0$ we have

$$\frac{\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|_2^2}{n} + \lambda\|\hat{\beta}\|_1 \leq \frac{\|\mathbf{Y} - \mathbf{X}\beta^0\|_2^2}{n} + \lambda\|\beta^0\|_1$$

We now replace $\mathbf{Y}$ using equation (1.1).

$$
\begin{aligned}
&\frac{\|\mathbf{Y}-\mathbf{X}\hat{\beta}\|_2^2}{n}+\lambda\|\hat{\beta}\|_1 \leq \frac{\|\mathbf{Y}-\mathbf{X}\beta^0\|_2^2}{n}+\lambda\|\beta^0\|_1 \\
\implies &\frac{\|(\mathbf{X}\beta^0+\boldsymbol{\varepsilon})-\mathbf{X}\hat{\beta}\|_2^2}{n}+\lambda\|\hat{\beta}\|_1 \leq \frac{\|(\mathbf{X}\beta^0+\boldsymbol{\varepsilon})-\mathbf{X}\beta^0\|_2^2}{n}+\lambda\|\beta^0\|_1 \\
\implies &\frac{\|\mathbf{X}(\beta^0-\hat{\beta})+\boldsymbol{\varepsilon}\|_2^2}{n}+\lambda\|\hat{\beta}\|_1 \leq \frac{\|\mathbf{X}(\beta^0-\beta^0)+\boldsymbol{\varepsilon}\|_2^2}{n}+\lambda\|\beta^0\|_1 \\
\implies &\frac{\langle \mathbf{X}(\beta^0-\hat{\beta})+\boldsymbol{\varepsilon}, \mathbf{X}(\beta^0-\hat{\beta})+\boldsymbol{\varepsilon}\rangle}{n}+\lambda\|\hat{\beta}\|_1 \leq \frac{\|\boldsymbol{\varepsilon}\|_2^2}{n}+\lambda\|\beta^0\|_1 \\
\implies &\frac{\|\mathbf{X}(\beta^0-\hat{\beta})\|_2^2+\|\boldsymbol{\varepsilon}\|_2^2+2\langle \mathbf{X}(\beta^0-\hat{\beta}),\boldsymbol{\varepsilon}\rangle}{n}+\lambda\|\hat{\beta}\|_1 \leq \frac{\|\boldsymbol{\varepsilon}\|_2^2}{n}+\lambda\|\beta^0\|_1 \\
\implies &\frac{\|\mathbf{X}(\hat{\beta}-\beta^0)\|_2^2}{n}+\lambda\|\hat{\beta}\|_1 \leq \frac{2\langle \mathbf{X}(\hat{\beta}-\beta^0),\boldsymbol{\varepsilon}\rangle}{n}+\lambda\|\beta^0\|_1 \\
\implies &\frac{\|\mathbf{X}(\hat{\beta}-\beta^0)\|_2^2}{n}+\lambda\|\hat{\beta}\|_1 \leq \frac{2\boldsymbol{\varepsilon}^T\mathbf{X}(\hat{\beta}-\beta^0)}{n}+\lambda\|\beta^0\|_1
\end{aligned}
$$

□

To be added

- how to get $\hat{b}$ on page 101.
- where the χ^2 distribution comes from in page 101