FYS-STK4155 Project 1

Bendik Steinsvåg Dalen & Gabriel Sigurd Cabrera

September 26, 2019

Abstract

1 Introduction

yvycjyukioyjfchdxcgfjhbk

2 Data

3 Method

4 Results

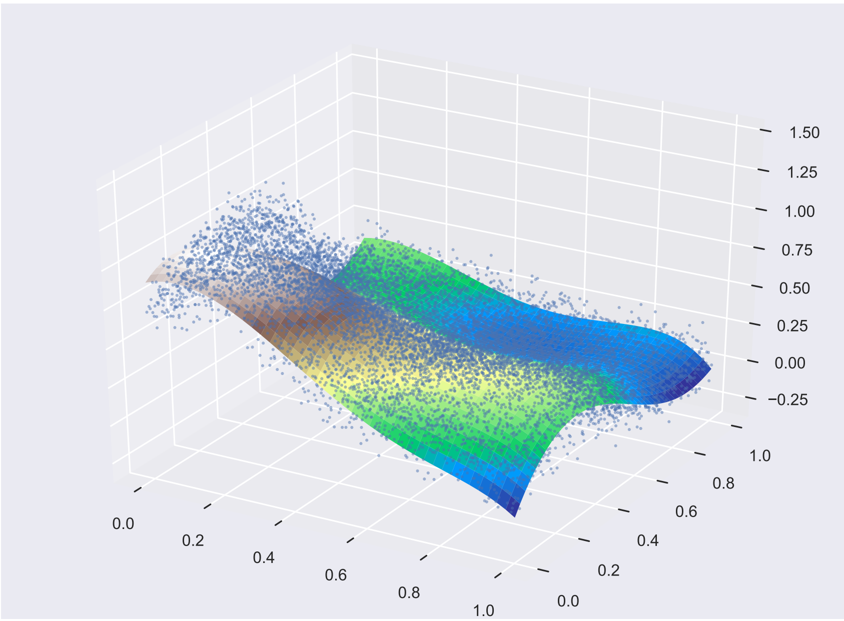

Figure 1: The resulting function after performing a standard least square regression analysis using polynomials in x and y up to fifth order on the Franke function

	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{16}
SLS	0.00789	0.322	2.46	4.81	5.61	1.76	0.322	1.85	4.3	5.97	2.83	2.46	4.3	5.97	4.23	4.81	5.97
k-fold	0.0321	4.03	94	474	517	79.5	4.03	57.3	260	301	60.1	93.9	260	261	57.1	474	301

	β_{17}	β_{18}	β_{19}	β_{20}
SLS	4.23	5.61	2.83	1.76
k-fold	57.1	517	60.1	79.5

Table 1: The variance of β for the standard least square regression and for k-fold validation

5 Discussion

	MSE	R^2
SLS	0.015	0.84
k-fold	0.012	0.87

Table 2: MSE and R2 for a and b

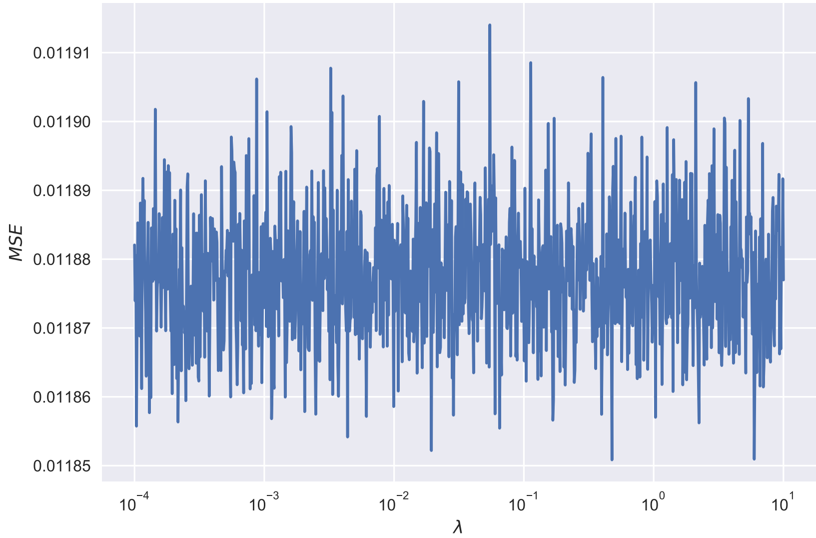

Figure 2: The Mean Squared Error for the Ridge method for different values of λ

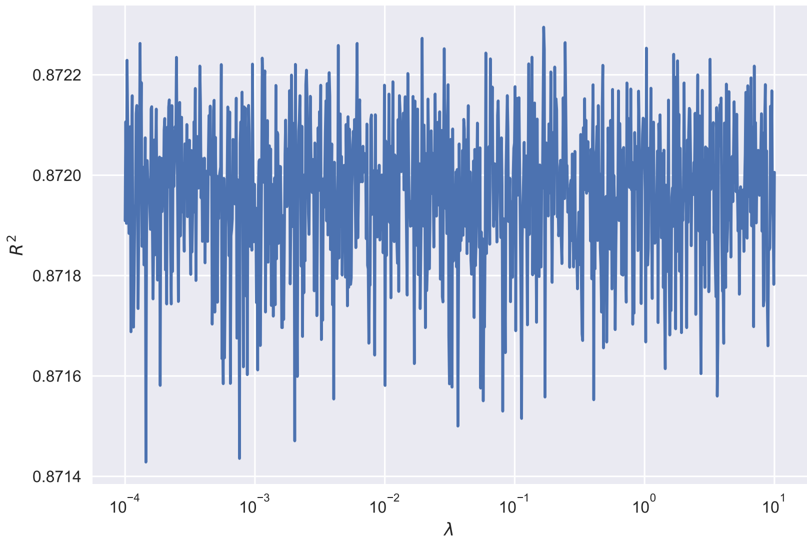

Figure 3: R^2-score for the Ridge method for different values of λ

6 Appendix

6.1 Part c math

We are too show that

$$C(\boldsymbol{X}, \boldsymbol{\beta}) = \frac{1}{n}\sum_{i=0}^{n-1}(y_i - \tilde{y}_i)^2 = \mathbf{E}\left[(\boldsymbol{y} - \tilde{\boldsymbol{y}})^2\right] = \frac{1}{n}\sum_i (f_i - \mathbf{E}[\tilde{\boldsymbol{y}}])^2 + \frac{1}{n}\sum_i (\tilde{y}_i - \mathbf{E}[\tilde{\boldsymbol{y}}])^2 + \sigma^2 \tag{1}$$

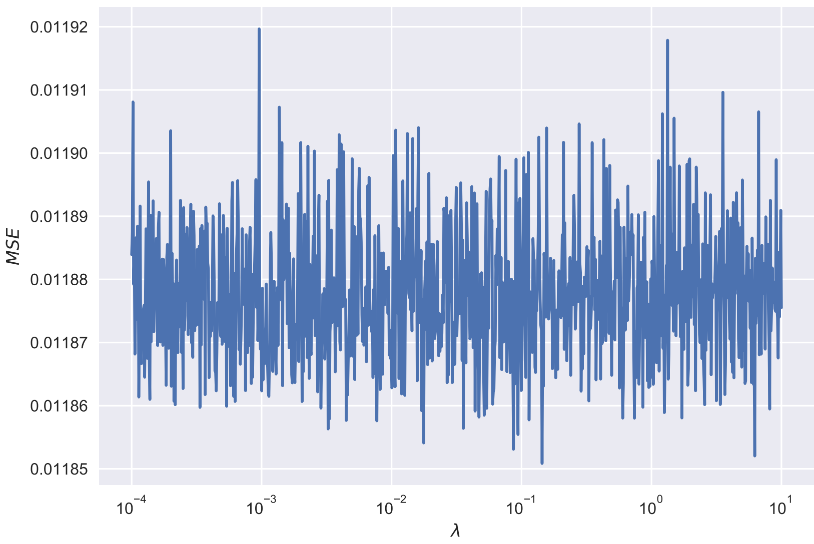

Figure 4: The Mean Squared Error for the Lasso method for different values of λ

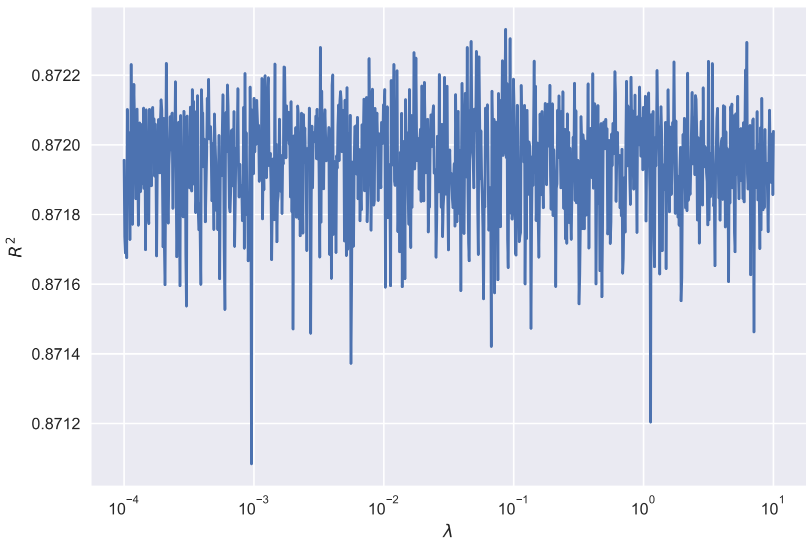

Figure 5: R^2-score for the Lasso method for different values of λ

$$
\begin{aligned}
\mathbf{E}\left[(\boldsymbol{y}-\tilde{\boldsymbol{y}})^2\right] =& \frac{1}{n}\sum_i (y_i-\tilde{y}_i)^2 = \frac{1}{n}\sum_i (f_i+\varepsilon-\tilde{y}_i)^2 && (2)\\
=& \frac{1}{n}\sum_i (f_i+\varepsilon-\tilde{y}_i+\mathbf{E}[\tilde{\boldsymbol{y}}]-\mathbf{E}[\tilde{\boldsymbol{y}}])^2 \qquad \mid \text{introduce } a=f_i-\mathbf{E}[\tilde{\boldsymbol{y}}] \text{ and } b=\tilde{y}_i-\mathbf{E}[\tilde{\boldsymbol{y}}] && (3)\\
=& \frac{1}{n}\sum_i (a-b+\varepsilon)^2 = \frac{1}{n}\sum_i \left(a^2-2ab+b^2-2b\varepsilon+\varepsilon^2+2a\varepsilon\right) && (4)\\
=& \frac{1}{n}\sum_i (f_i-\mathbf{E}[\tilde{\boldsymbol{y}}])^2 + \frac{1}{n}\sum_i \left(\varepsilon^2\right) + \frac{1}{n}\sum_i (\tilde{y}_i-\mathbf{E}[\tilde{\boldsymbol{y}}])^2 - 2\frac{1}{n}\sum_i \varepsilon\,(\tilde{y}_i-\mathbf{E}[\tilde{\boldsymbol{y}}]) + 2\frac{1}{n}\sum_i \varepsilon\,(f_i-\mathbf{E}[\tilde{\boldsymbol{y}}]) && (5)\\
& -2\frac{1}{n}\sum_i (f_i-\mathbf{E}[\tilde{\boldsymbol{y}}])\,(\tilde{y}_i-\mathbf{E}[\tilde{\boldsymbol{y}}]) && (6)\\
=& \frac{1}{n}\sum_i (f_i-\mathbf{E}[\tilde{\boldsymbol{y}}])^2 + \frac{1}{n}\sum_i (\tilde{y}_i-\mathbf{E}[\tilde{\boldsymbol{y}}])^2 + \sigma^2 - 2\mathbf{E}[\varepsilon]\frac{1}{n}\sum_i (\tilde{y}_i-\mathbf{E}[\tilde{\boldsymbol{y}}]) + 2\mathbf{E}[\varepsilon]\frac{1}{n}\sum_i (f_i-\mathbf{E}[\tilde{\boldsymbol{y}}]) && (7)\\
& -2\frac{1}{n}\sum_i (f_i-\mathbf{E}[\tilde{\boldsymbol{y}}])\,(\tilde{y}_i-\mathbf{E}[\tilde{\boldsymbol{y}}]) && (8)\\
=& \frac{1}{n}\sum_i (f_i-\mathbf{E}[\tilde{\boldsymbol{y}}])^2 + \frac{1}{n}\sum_i (\tilde{y}_i-\mathbf{E}[\tilde{\boldsymbol{y}}])^2 + \sigma^2 \quad \blacksquare && (9)
\end{aligned}
$$

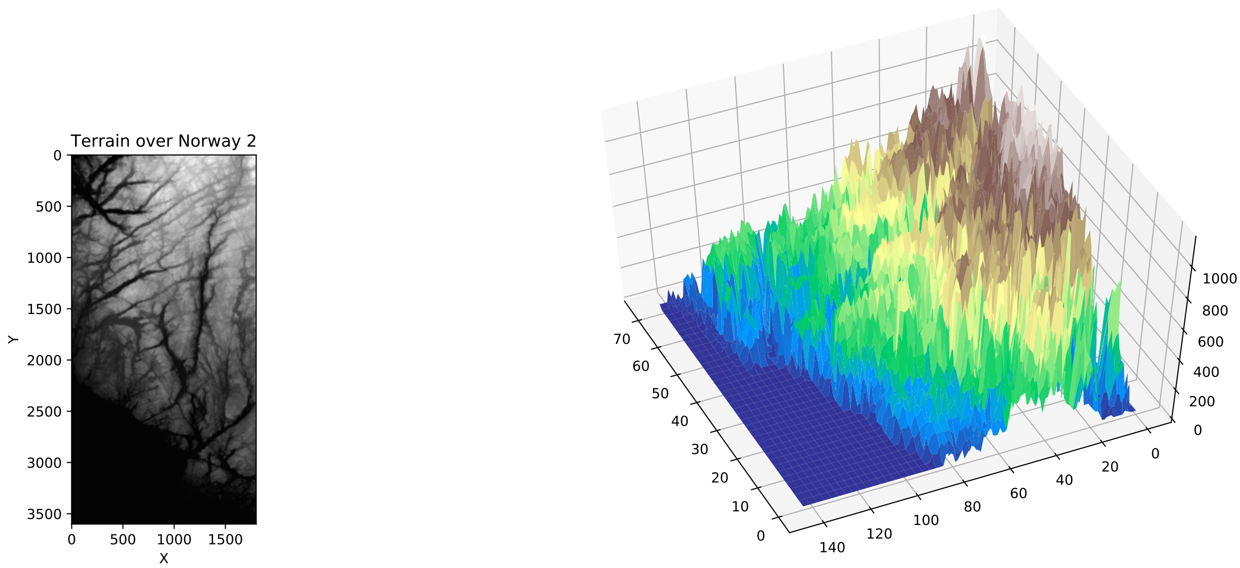

Figure 6: The terrain-data we are studying, from Møsvatn Austfjell in Norway

Where $\frac{1}{n}\sum_i (f_i - \mathbf{E}[\tilde{\boldsymbol{y}}])$ is the bias and $\frac{1}{n}\sum_i (\tilde{y}_i - \mathbf{E}[\tilde{\boldsymbol{y}}])^2$ is the variance.
(skal vi gjøre noe annet og?)