FYS-STK4155 Project 1

Bendik Steinsvåg Dalen & Gabriel Sigurd Cabrera

October 4, 2019

Abstract

1 Introduction

Things we need to write about:

Background on regression analysis and resampling methods. Mention: OLS Ridge Lasso k-fold cross-validation Bias-Variance trade of?

We will first study how they preform for the two dimensional Franke-function. (A bit about the Franke-function, maybe a tldr for the method).

We will then implement them for some real terrain data for Møsvatn Austfjell in Norway. mm. (biggest lake in Norway)

2 Data

We will use real terrain data for Møsvatn Austfjell in Telemark, Norway, collected from https://earthexplorer.usgs.gov.

3 Method

4 Results

4.1 Franke-Function

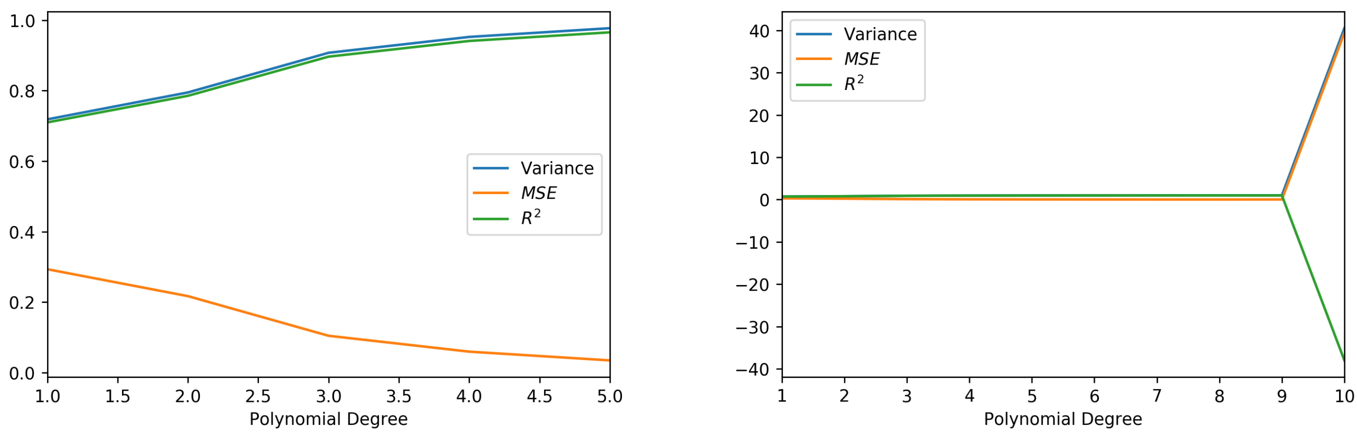

Figure 1: Plots of the mean square error, the R^2-score and the σ variance of the β-values against the polynomial degree after performing a standard least square regression analysis on the Franke-function

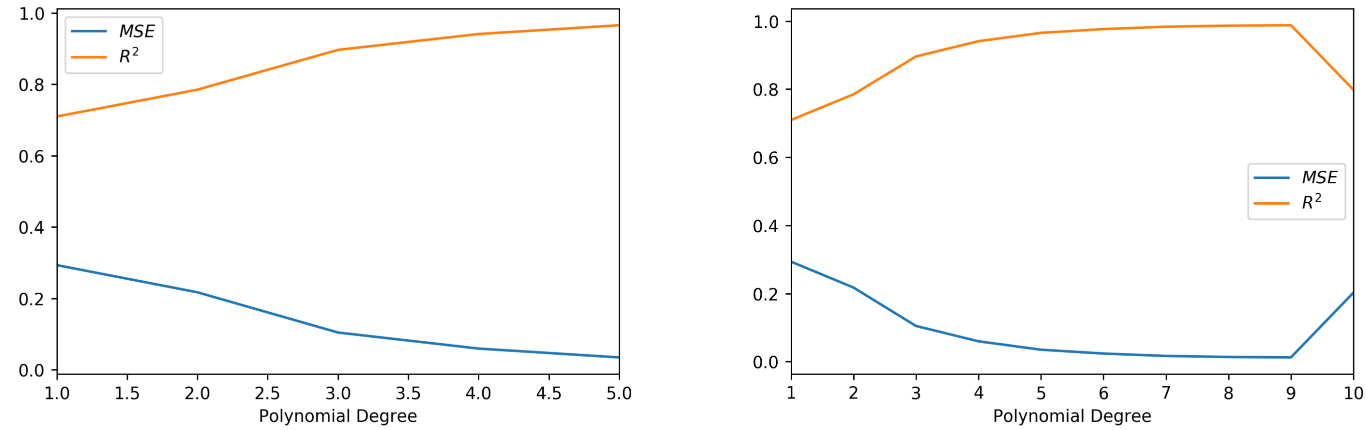

Figure 2: Plots of the mean square error and the R^2-score against the polynomial degree after performing a standard least square regression analysis on the Franke-function and performing a k-fold cross-validation

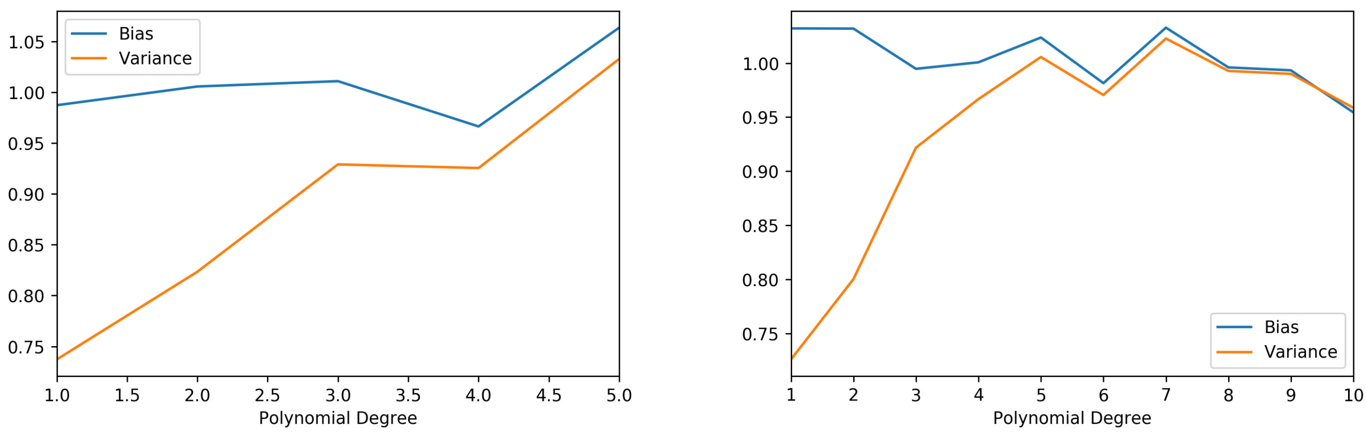

Figure 3: Plots of the bias and the varince against the polynomial degree after performing a standard least square regression analysis on the Franke-function

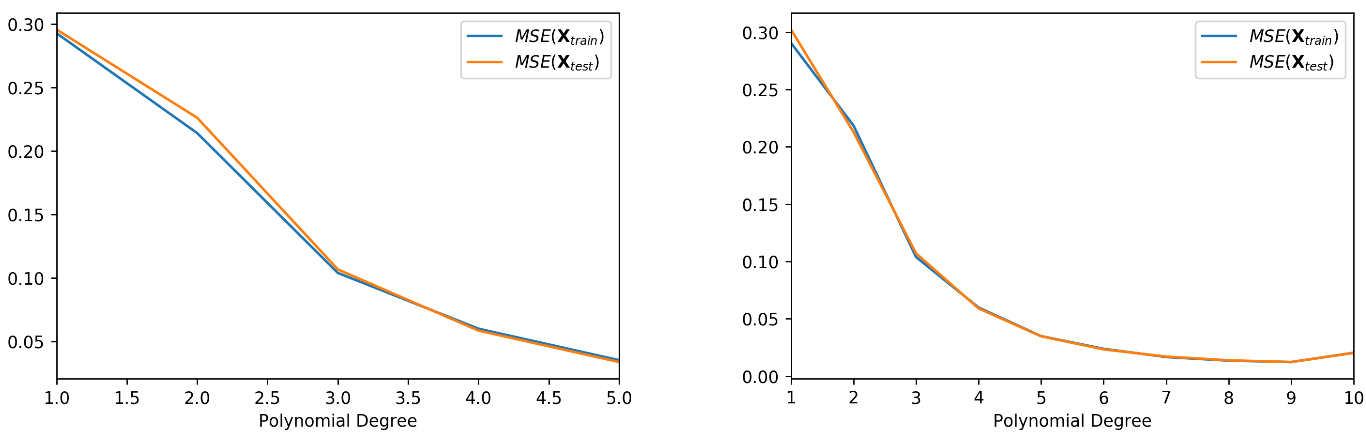

Figure 4: Plots of the mean square error for the training data and the testing data, against the polynomial degree after performing a standard least square regression analysis on the Franke-function

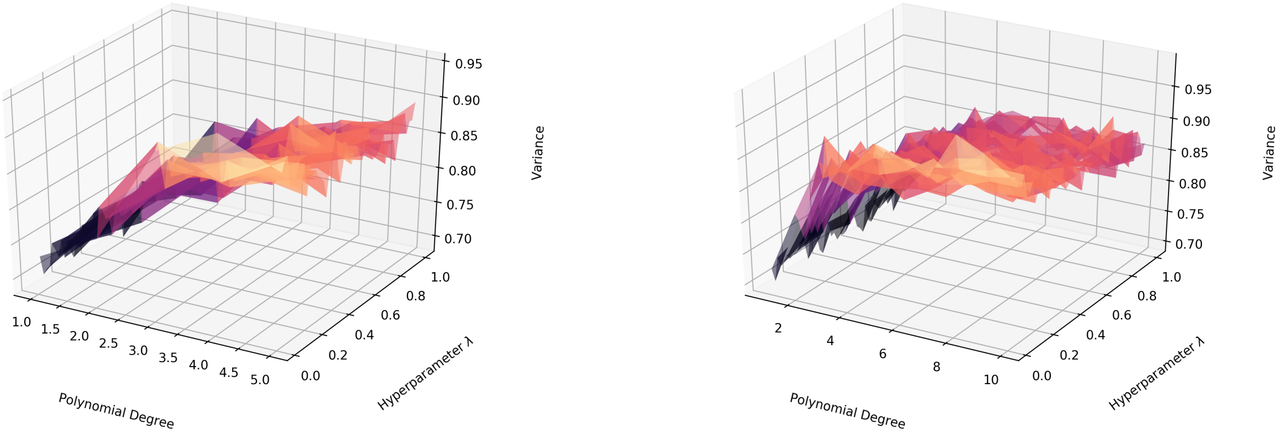

Figure 5: Plots of the varinace against the polyonomial degree and the hyperparameter λ after performing Ridge regression on the Franke-function

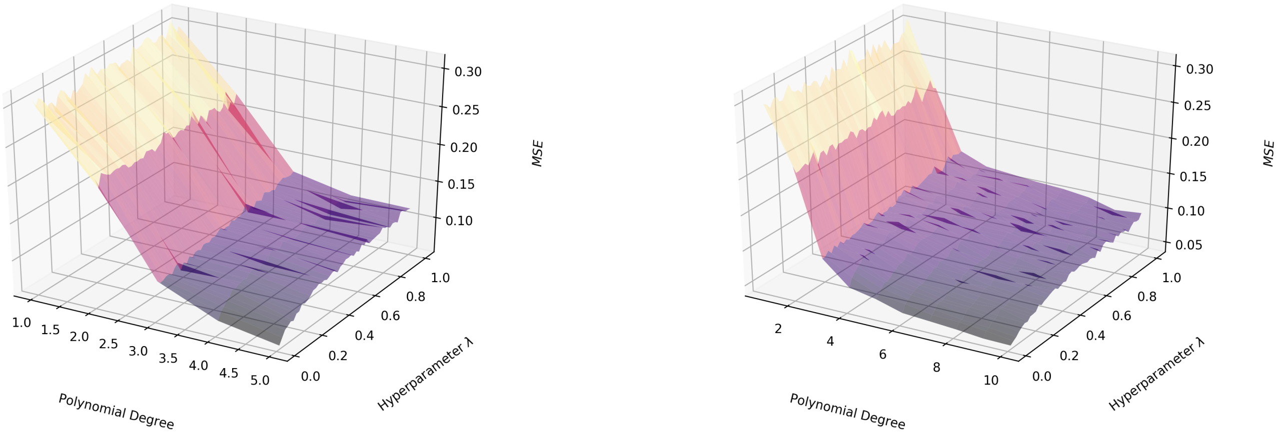

Figure 6: Plots of the mean square error against the polyonomial degree and the hyperparameter λ after performing Ridge regression on the Franke-function

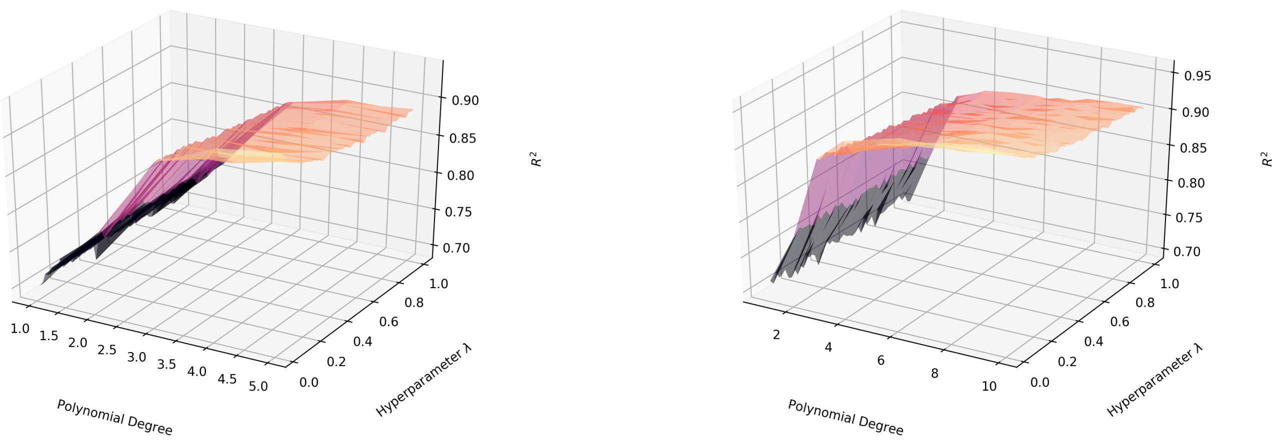

Figure 7: Plots of the R^2-score against the polyonomial degree and the hyperparameter λ after performing Ridge regression on the Franke-function

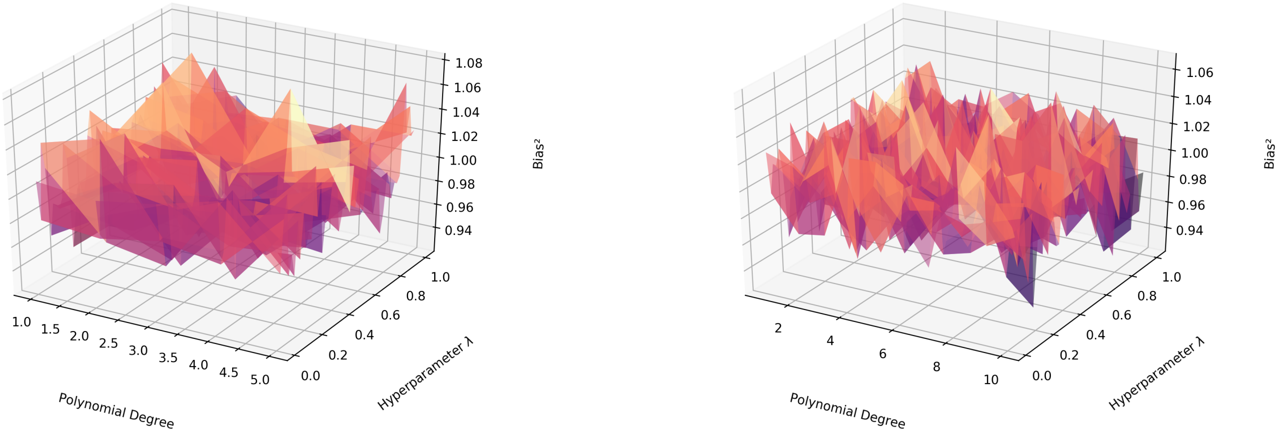

Figure 8: Plots of the bias squared against the polyonomial degree and the hyperparameter λ after performing Ridge regression on the Franke-function

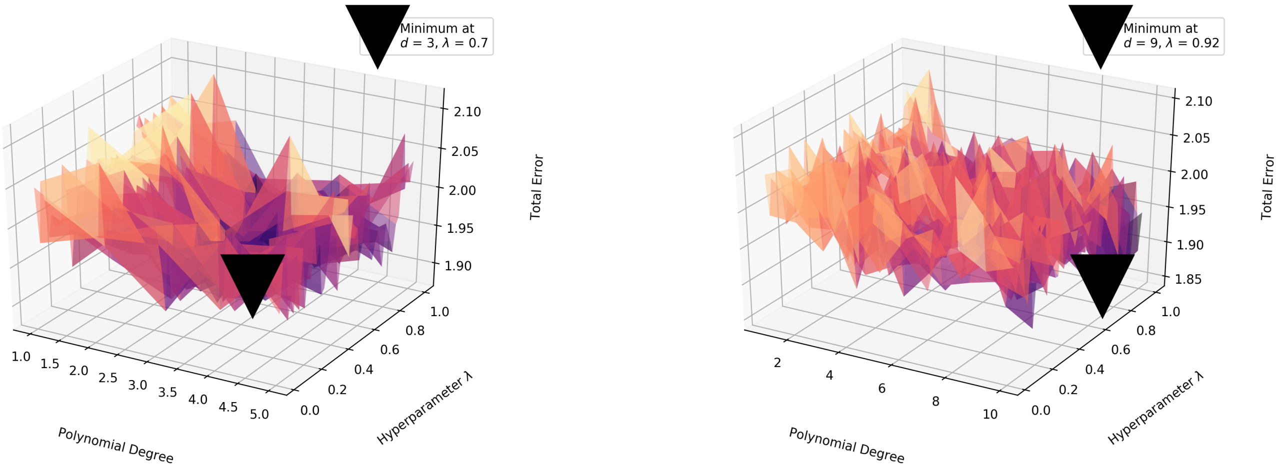

Figure 9: Plots of the total error against the polyonomial degree and the hyperparameter λ after performing Ridge regression on the Franke-function

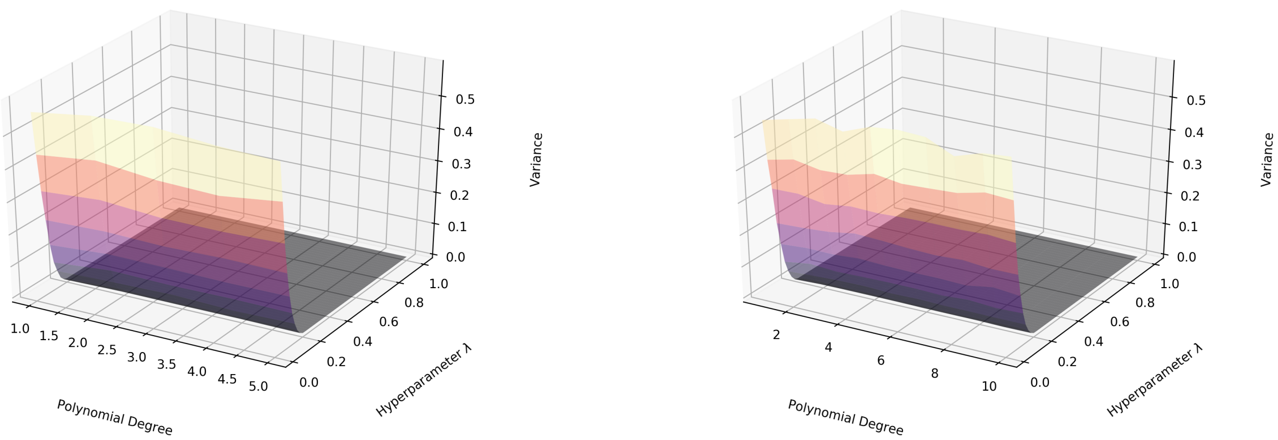

Figure 10: Plots of the varinace against the polyonomial degree and the hyperparameter λ after performing Lasso regression on the Franke-function

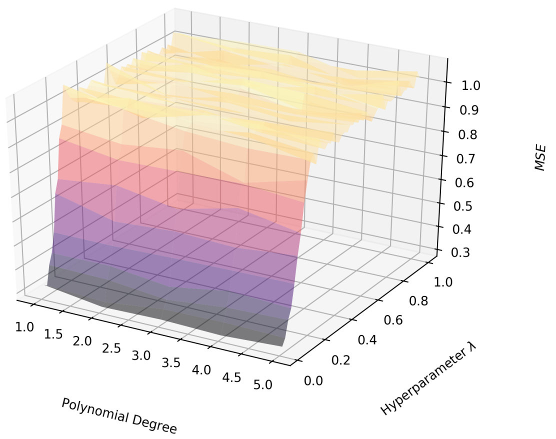

Figure 11: Plots of the mean square error against the polyonomial degree and the hyperparameter λ after performing Lasso regression on the Franke-function

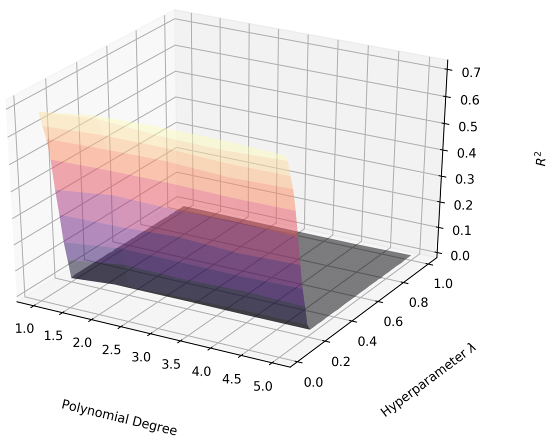

Figure 12: Plots of the R^2-score against the polyonomial degree and the hyperparameter λ after performing Lasso regression on the Franke-function

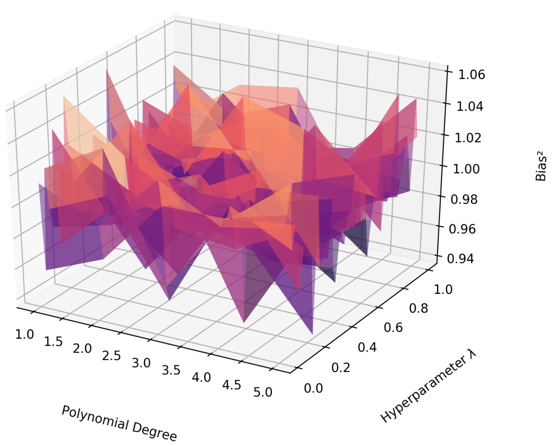

Figure 13: Plots of the bias squared against the polyonomial degree and the hyperparameter λ after performing Lasso regression on the Franke-function

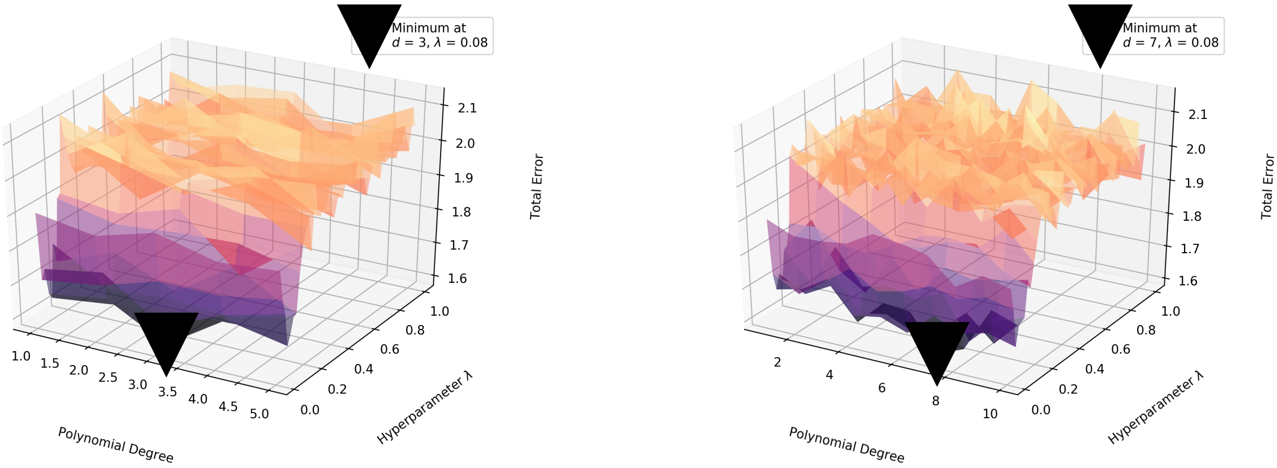

Figure 14: Plots of the total error against the polyonomial degree and the hyperparameter λ after performing Lasso regression on the Franke-function

4.2 Real Data

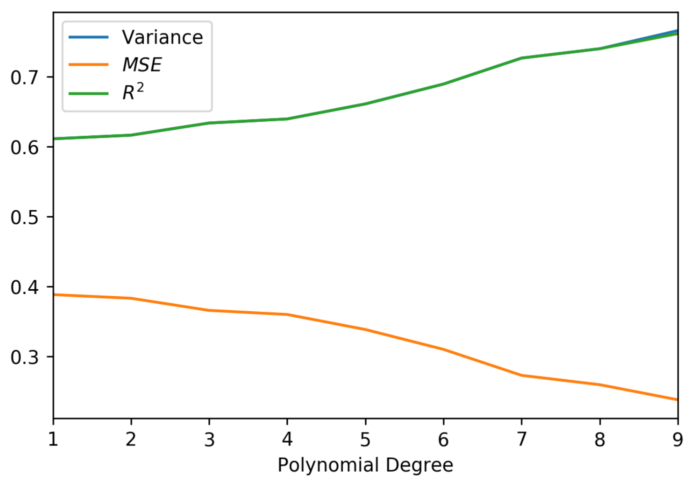

Figure 15: A plot of the mean square error, the R^2-score and the σ variance of the β-values against the polynomial degree after performing a standard least square regression analysis on the real terrain data of Møsvatn Austfjell

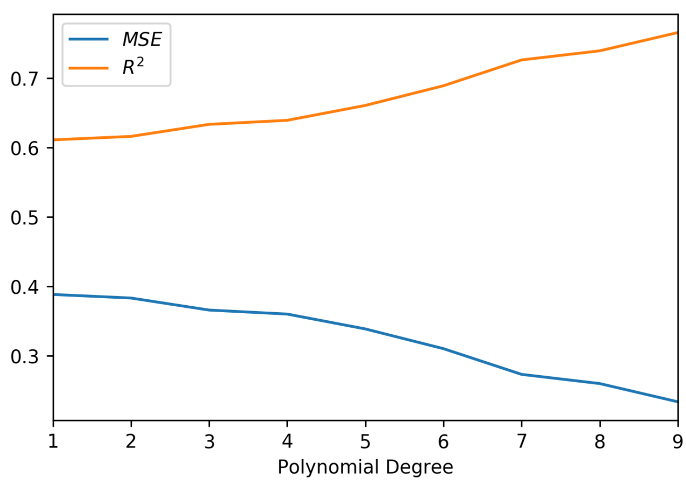

Figure 16: A plot of the mean square error and the R^2-score against the polynomial degree after performing a standard least square regression analysis on the real terrain data of Møsvatn Austfjell and performing a k-fold cross-validation

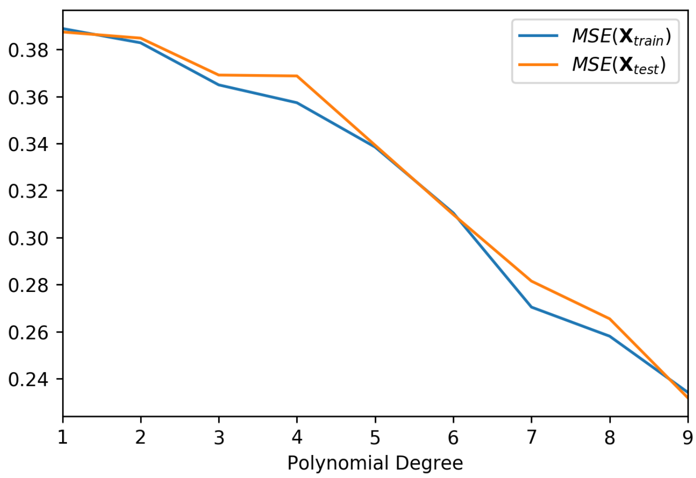

Figure 17: A plot of the mean square error for the training data and the testing data, against the polynomial degree after performing a standard least square regression analysis on the real terrain data of Møsvatn Austfjell

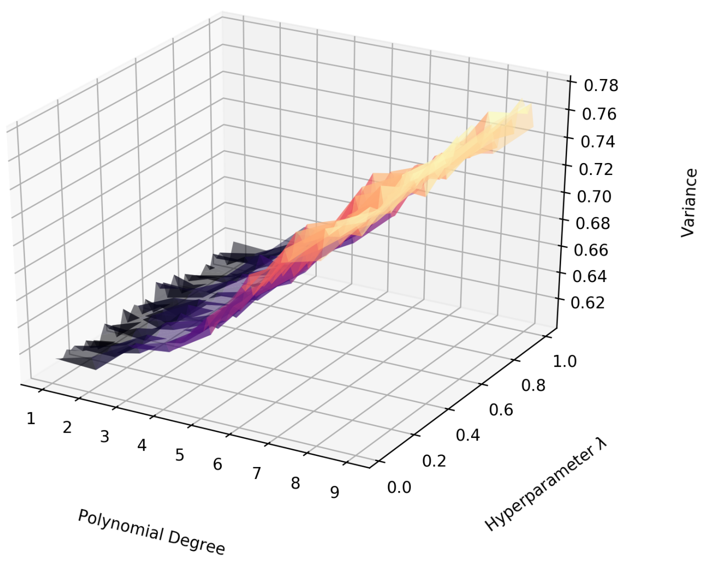

Figure 18: Plots of the varinace against the polyonomial degree and the hyperparameter λ after performing Ridge regression on the real terrain data of Møsvatn Austfjell

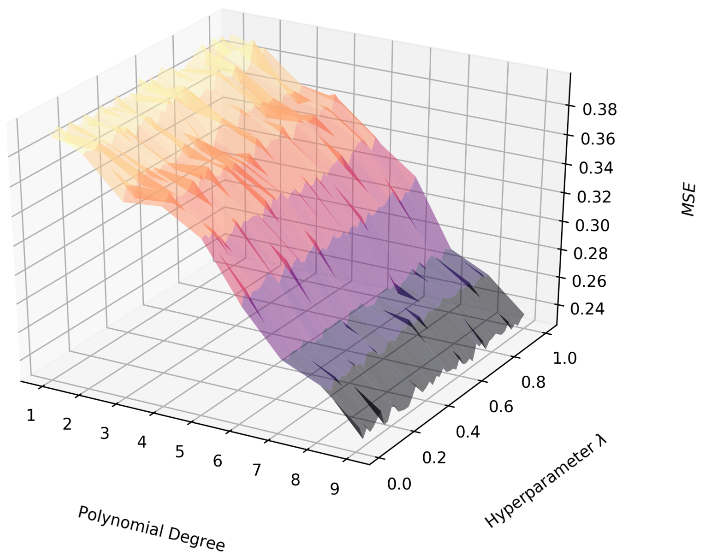

Figure 19: Plots of the mean square error against the polyonomial degree and the hyperparameter λ after performing Ridge regression on the real terrain data of Møsvatn Austfjell

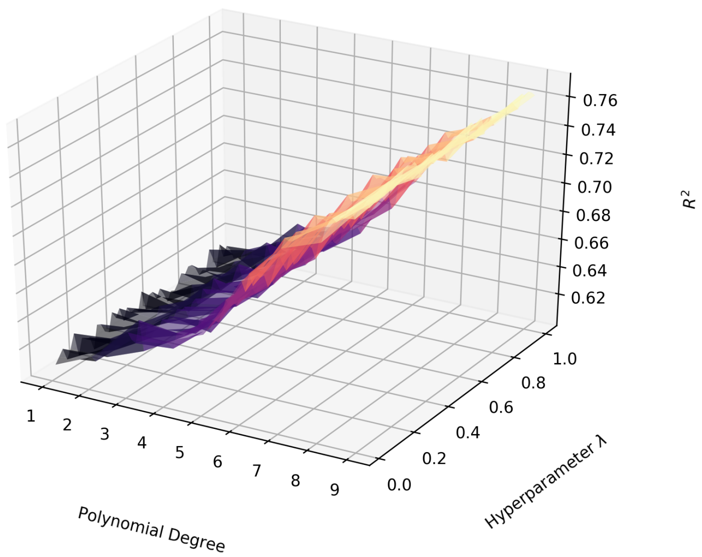

Figure 20: Plots of the R^2-score against the polyonomial degree and the hyperparameter λ after performing Ridge regression on the real terrain data of Møsvatn Austfjell

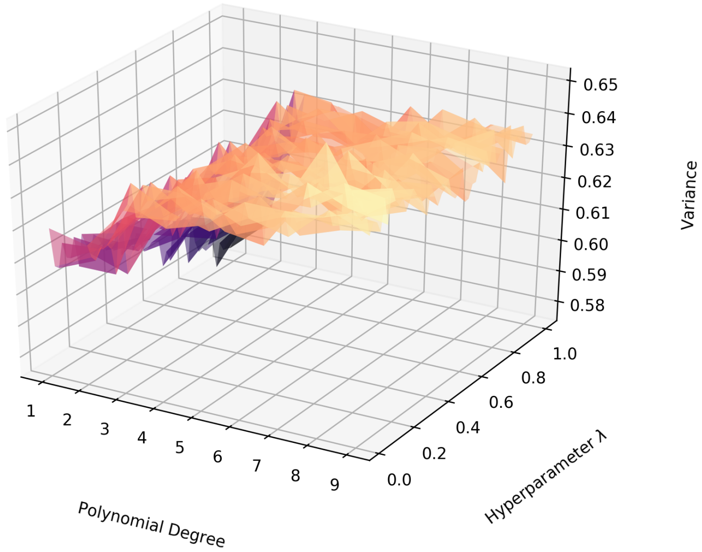

Figure 21: Plots of the varinace against the polyonomial degree and the hyperparameter λ after performing Lasso regression on the real terrain data of Møsvatn Austfjell

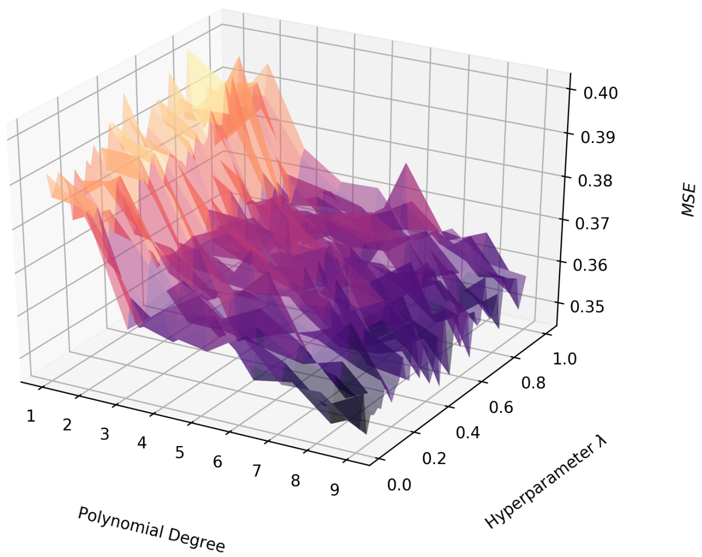

Figure 22: Plots of the mean square error against the polyonomial degree and the hyperparameter λ after performing Lasso regression on the real terrain data of Møsvatn Austfjell

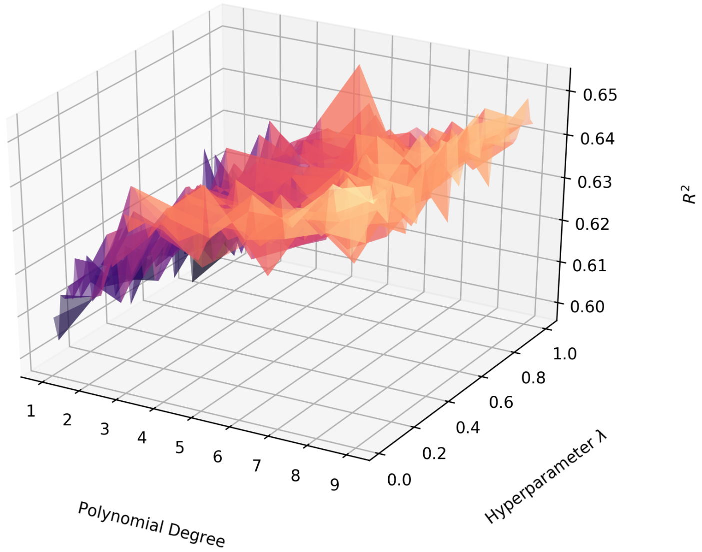

Figure 23: Plots of the R^2-score against the polyonomial degree and the hyperparameter λ after performing Lasso regression on the real terrain data of Møsvatn Austfjell

5 Discussion

6 Appendix

6.1 Part c math

We are too show that

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

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

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?)