# FYS-STK3155 Project 1

Lasse Pladsen, Parham Qanbari, & Sander V. Vattøy
(Dated: September 28, 2023)

By comparing predictions of a two-dimensional function by three different methods – Ordinary Least Squares, Ridge Regression and Lasso – we will confirm their degree of accuracy up against theory. Thereafter we go on to do a bias-variance trade-off and cross validation. Here we find that ...

## I. INTRODUCTION

Many have seen the amazing Norwegian nature. The fjords are remenants of the iceage shaping the rock, and the mountains are a memory of the forces of tectonic activity that shaped the land as we know it.

In this report we want to explore the Norwegian topography using basic machine learning methods. Essentially, we are interested in predicting the topography using real data. For the specific methods we are going to use Ordinary Least Squares (OLS), Ridge and Lasso regression. We will exploring these regression methods and compare them. For this, we will use metrics such as bias-variance tradeoff, cross-validation and in addition we are going to explore the resulting Mean Sqaured Errors (MSE). The point of this will be to see what methods produce the best fit to the actual topography.

Before actually using the topography data we will test the code on the Franke's function (See *Franke's function* 1).

## II. METHODS

We begin by defining the function we are going to do regressions on. This is called a Franke function, and is a two dimentional weighted function over four exponentials:

$$\begin{aligned} f(x,y) = &\frac{3}{4}\exp\left(-\frac{(9x-2)^2}{4} - \frac{(9y-2)^2}{4}\right) \\ &+ \frac{3}{4}\exp\left(-\frac{(9x+1)^2}{49} - \frac{(9y+1)^2}{10}\right) \\ &+ \frac{1}{2}\exp\left(-\frac{(9x-7)^2}{4} - \frac{(9y-3)^2}{4}\right) \\ &- \frac{1}{5}\exp\left(-(9x-4)^2 - (9y-7)^2\right) \end{aligned} \tag{1}$$

Firstly we are going to use our own code doing the Ordinary Least Squares (OLS) for a linear regression of the Franke function up to it's fifth order (link to code is in the appendix section A). Linear regression using OLS is given from

$$\tilde{\mathbf{z}} = X\beta \tag{2}$$

and

$$\beta_{OLS} = (X^TX)^{-1}X^T\mathbf{z} \tag{3}$$

where $\beta_{OLS}$ is the optimal predictor for our function with function-output $\mathbf{z}$, and $X$ is the design matrix for our two-dimensional function.

After defining this, we are able to calculate the Mean Square Error (MSE) defined by

$$MSE(\mathbf{z}, \tilde{\mathbf{z}}_i) = \frac{1}{n}\sum_{i=0}^{n-1}(z_i - \tilde{z}_i)^2, \tag{4}$$

where the mean value of $\mathbf{z}$

$$\bar{z} = \frac{1}{n}\sum_{i=0}^{n-1} z_i \tag{5}$$

Can also calculate the $R^2$-score, given by

$$R^2(\mathbf{z}, \tilde{\mathbf{z}}_i) = 1 - \frac{\sum_{i=0}^{n-1}(z_i - \tilde{z}_i)^2}{\sum_{i=0}^{n-1}(z_i - \bar{z}_i)^2} \tag{6}$$

To better understand the concept of the linear regression, $\beta_{OLS}$, MSE and the $R^2$-score should be plotted with respect to the polynomial clomplexity.

## III. RESULTS

test [1]

## IV. DISCUSSION

## V. CONCLUSION

### Appendix A: Github repository

`https://github.com/LassePladsen/FYS-STK3155-projects/tree/main/project1`

### Appendix B: Part d derivations

#### a. *Expectation value of* $\mathbf{y}$

We will show that $\mathbb{E}(y_i) = \mathbf{X}_{i,*}\beta$ by using $\mathbf{y} = \mathbf{f} + \epsilon \simeq \mathbf{X}\beta + \epsilon$ and separating the expectation value of a sum. Here we approximated $\mathbf{f}$ with $\mathbf{X}\beta$ using OLS. Then taking a value of $\mathbf{y}$ with index $i$ we get $y_i = \sum_j X_{ij}\beta_j + \epsilon_i$:

$$
\begin{aligned}
\mathbb{E}(y_i) &= \mathbb{E}(\Sigma_j X_{ij}\beta_j + \epsilon_i) \\
&= \mathbb{E}(\Sigma_j X_{ij}\beta_j) + \mathbb{E}(\epsilon_i) \\
&= \mathbb{E}(\Sigma_j X_{ij}\beta_j) \\
&= \Sigma_j X_{ij}\beta_j \\
&= \mathbf{X}_{i,*}\beta
\end{aligned}
$$

### b. *Variance of* $\mathbf{y}$

Using the same method as above we will now show $Var(y_i) = \sigma^2$ where $\sigma^2$ is the variance of our data's stochastic noise $\epsilon$. Here we use the definition of variance being $Var(x) = \mathbb{E}(x^2) - \mathbb{E}(x)^2$:

$$
\begin{aligned}
Var(y_i) &= \mathbb{E}(y_i^2) - \mathbb{E}(y_i)^2 \\
&= \mathbb{E}[(\mathbf{X}_{\mathbf{i},*}\beta) + \epsilon_i)^2] - (\mathbf{X}_{\mathbf{i},*}\beta)^2 \\
&= \mathbb{E}[(\mathbf{X}_{\mathbf{i},*}\beta)^2 + \epsilon_i^2 + 2\mathbf{X}_{\mathbf{i},*}\beta\epsilon_i] - (\mathbf{X}_{\mathbf{i},*}\beta)^2 \\
&= \mathbb{E}[(\mathbf{X}_{\mathbf{i},*}\beta)^2] + \mathbb{E}[\epsilon_i^2] + \mathbb{E}[2\mathbf{X}_{\mathbf{i},*}\beta\epsilon_i] - (\mathbf{X}_{\mathbf{i},*}\beta)^2 \\
&= (\mathbf{X}_{\mathbf{i},*}\beta)^2 + \mathbb{E}[\epsilon_i^2] + 2\mathbf{X}_{\mathbf{i},*}\beta\mathbb{E}[\epsilon_i] - (\mathbf{X}_{\mathbf{i},*}\beta)^2 \\
&= \mathbb{E}[\epsilon_i^2] \\
&= Var(\epsilon_i) + \mathbb{E}(\epsilon)^2 \\
&= Var(\epsilon_i) \\
&= \sigma^2
\end{aligned}
$$

### c. *OLS expectation value of optimal* $\beta$

Here we will show that the expectation value for the optimal $\beta$ for OLS, $\hat{\beta}_{OLS}$, equals $\beta_{OLS}$:

$$
\begin{aligned}
\mathbb{E}(\hat{\beta}_{OLS}) &= \mathbb{E}[(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y}] \\
&= (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbb{E}[\mathbf{y}] \\
&= (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{X}\beta_{OLS} \\
&= \beta_{OLS}
\end{aligned}
$$

Here we used that the expectation value of the non-stochastic matrix is just the matrix itself ($\mathbb{E}(\mathbf{X}) = X$) since it has zero variance (non-stochastic).

[1] Hjorth-Jensen, M. (2023). Project 1 on machine learning. `https://compphysics.github.io/MachineLearning/doc/LectureNotes/_build/html/project1.html`. [Online; accessed 26-September-2023].