# 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

In today's modern world machine learning has emerged as a revolutionary technology. Even if just superficially, it has become widely known around the world. By allowing computers to learn from supplied data we create powerful tools with incredible real world applications for analyzation and prediction.

In this project we will explore and study low-level machine learning methods. Our focus will be on understanding and optimizing the potential of data regression methods such as ordinary least squares (OLS), ridge regression, and lasso regression. We will look into the factors that influence their accuracy and predictive capabilities, employing a variety of strategies to minimize errors and improve performance. Finally, we will apply our finely-tuned algorithms to analyze real-world topographic data.

## 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 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

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

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 $Var(x) = \mathbb{E}(x^2) - \mathbb{E}(x)^2$:

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

*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 $\beta_{OLS}$:

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

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].