Exercise week 39
FYS-STK4155

Helene Lane and Manuela Leal Nader
University of Oslo
(Dated: September 26, 2025)

Machine Learning methods have increased in use, both in and outside academia, over the last years. However, the immense number of methods available might present a challenge to the user, particularly in selecting appropriate models and evaluating their performance. In this report, we have investigated in more detail three different regression methods: Ordinary Least Squares (OLS), Ridge and Lasso. Based on analysis performed with the one-dimensional Runge's function, we have concluded that the OLS method provides the best fit. Although it is the least complex of the three, it shows the best performance for the simple case used here.

I. INTRODUCTION

Machine learning has spread to all areas of daily life, bringing different challenges depending on how it is used. Within science, these methods are vastly used across research areas. Not only that, they are constantly being improved and new methods developed, posing a challenge in regard to how to choose among them. When it comes to the type of training data available, there are two types of learning models, supervised and unsupervised ([1]). When the training data has already the property you want to predict, we say that the learning process is supervised. That is the type of problem we will focus here. Models can also be divided into two other categories, regression and classification [1]. Regression models are the ones that deals with quantitative labels, while classification models aims to predict qualitative ones. In this report, we will explore three different regression methods: Ordniary Least Squares (OLS), Ridge and Lasso. We will use these three regression methods to approximate Runge's function:

$$f(x) = \frac{1}{1 + 25x^2} \tag{1}$$

This function was chosen because it is a function which Runge discovered that approximating it with higher order polynomials can be tricky [2]. The methods were chosen because they are easy to implement by people who have just started learing about machine learning.

In section II we will cover the three regression methods used and the variations of them used in our analysis. First, we will present OLS for the Runge function before Ridgre regression for the same function. Next, we will present our own gradient descent code before adding momentum and updating the learning rate. Next, we will present Lasso regression and stochastic gradient descent. Lastly, we will present cross-validation before describing our use of AI tools.

In section III we will present our results from using these various regression models for Runge's function. We will present and compare the methods used and consider the bias-variance trade off in our analysis.

Section IV concludes our report with a summary.

II. METHODS

A. Scaling

Scaling the data is important in many machine learning algorithms. If the features in our data are measured on propotionally very different scales, scaling can prevent the algorithm from only optimising on the feature with largest error [3]. Standardising is a form of scaling the data where each feature has mean 0 and standard deviation 1. Standardising data is often a good choice of scaling because it does not change the shape of the distribution [3]. We chose to scale our data because ...

B. Ordinary Least Squares

Ordinary Least Squares is a form of linear regression with the aim of minimizing the squared distance from each data point to our suggested function [1].

The cost function for Ordinary Least Squares is given by the squared distance

$$C(\theta) = \frac{1}{n}\left\{ (\boldsymbol{y} - X\theta)^T (\boldsymbol{y} - X\theta) \right\} \tag{2}$$

Minimizing the cost function means taking the derivative and setting it equal to zero:

$$\frac{\partial C(\theta)}{\partial \theta} = 0 = X^T (\boldsymbol{y} - X\theta) \tag{3}$$

Solving for θ gives the following expression which gives the optimal parameters.

$$\hat{\theta}_{OLS} = \left(X^T X\right)^{-1} X^T \boldsymbol{y} \tag{4}$$

C. Method 1/X

D. Implementation

E. Use of AI tools

We used GPT UiO for information on how to write an abstract. The exported conversation is in the Github repository of the project (m-nader/FYS-STK4155-group14), inside a folder called LLM.

III. RESULTS AND DISCUSSION

For exercise 2 of week 39, we have added here a plot illustrating the Bias-Variance trade-off (Figure 1) and a heatmap showing the MSE of a Ridge regression model for various polynomial degrees and lambda values (Figure 2).

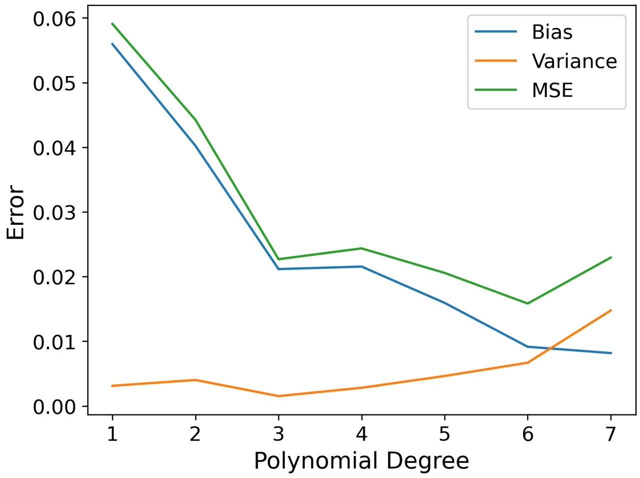

Figure 1: Illustrative example of the Bias-Variance trade-off. At a low model complexity (small polynomial degrees), the error of the model is dominated by the bias of it. As the complexity increases, the bias is reduced and the variance of the model becomes the dominating term of the error.

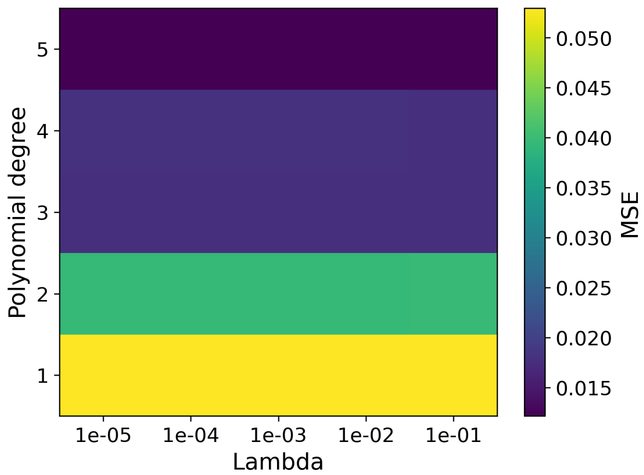

Figure 2: MSE of a Ridge regression model for different polynomial degrees and lambda values.

As requested in exercise 5 of week 39, we added citations to Hastie et al. [1] and sklearn [4].

IV. CONCLUSION

[1] T. Hastie, R.Tibshirani, and J.Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer Series in Statistics* (Springer, New York, 2009), URL `https://link.springer.com/book/10.1007%2F978-0-387-84858-7`.

[2] Runge's phenomenon, *Runge's phenomenon — Wikipedia, the free encyclopedia* (2025), [Online; accessed 25-September-2025], URL `http://en.wikipedia.org/w/index.php?title=Estimation_lemma&oldid=375747928`.

[3] S. Raschka, *Machine learning with pytorch and scikit-learn : develop machine learning and deep learning models with python* (Packt Publishing, Birmingham, England, 2022), ISBN 9781801816380.

[4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Journal of Machine Learning Research **12**, 2825 (2011).