Report Template
FYS-STK3155 - Project X

Boon Kiat Khaw
University of Oslo
(Dated: September 26, 2025)

This project investigates regression analysis methods with emphasis on the trade-offs between model complexity, bias, and variance. Using Runge's function as a test case, we applied Ordinary Least Squares (OLS), Ridge, and Lasso regression to study how polynomial fitting behaves under different levels of regularization. To evaluate the models, we computed mean squared error (MSE) and R^2 scores, and implemented resampling techniques including bootstrap and k-fold cross-validation. The results demonstrate how OLS suffers from overfitting at higher polynomial degrees, while Ridge and Lasso provide improved stability through regularization. We further explored gradient descent optimization, including adaptive learning rate methods, and discussed their convergence properties. The main findings highlight the importance of balancing bias and variance when selecting model complexity and regularization strength. These results provide insights into how resampling and penalization can improve predictive accuracy in regression tasks.

I. INTRODUCTION

Regression analysis plays a central role in statistical learning and data-driven modeling, providing tools to approximate underlying functional relationships from observed data. In this project, we focus on polynomial regression as a framework for exploring fundamental machine learning methods, including Ordinary Least Squares (OLS), Ridge regression, and Lasso regression. The motivation is to understand how different approaches handle the challenges of overfitting, bias, and variance, particularly when approximating functions with high curvature such as Runge's function.

The study addresses the classical problem of the bias–variance trade-off, which describes how increasing model complexity reduces bias but increases variance, and vice versa. Regularization techniques such as Ridge and Lasso regression mitigate overfitting by introducing penalty terms that control model flexibility. Alongside these methods, we implemented gradient descent algorithms and adaptive optimization schemes (momentum, RMSprop, ADAM), which allowed us to study convergence behavior under different learning rates.

These topics are well-established in the literature, particularly in the comprehensive treatment by Hastie, Tibshirani, and Friedman [1]. For the implementation of regression methods, we relied on the widely used `scikit-learn` library [2], which provides efficient and well-tested tools for machine learning in Python.

Structurally, this report is organized as follows. Section II describes the theoretical background and implementation details of the regression models and optimization methods. Section III presents and discusses the results, including figures that illustrate the bias–variance trade-off and the effect of regularization. Finally, Section IV summarizes the main findings and implications for regression analysis in machine learning.

II. METHODS

A. Method 1/X

- Describe the methods and algorithms, including the motivation for using them and their applicability to the problem
- Derive central equations when appropriate, the text is the most important part, not the equations.

B. Implementation

- Explain how you implemented the methods and also say something about the structure of your algorithm and present very central parts of your code, not more than 10 lines
- You should plug in some calculations to demonstrate your code, such as selected runs used to validate and verify your results. A reader needs to understand that your code reproduces selected benchmarks and reproduces previous results, either numerical and/or well-known closed form expressions.

C. Use of AI tools

In the course of this project, we made use of a large language model (LLM), ChatGPT, to support both the implementation and the reporting process. The LLM was used to debug LaTeX compilation errors for exercise 39, as well as in the debugging of code implementations in exercise 2.

III. RESULTS AND DISCUSSION

- Present your results

- Give a critical discussion of your work and place it in the correct context.
- Relate your work to other calculations/studies
- An eventual reader should be able to reproduce your calculations if she/he wants to do so. All input variables should be properly explained.
- Make sure that figures1, 2 and tables contain enough information in their captions, axis labels etc. so that an eventual reader can gain a good impression of your work by studying figures and tables only.

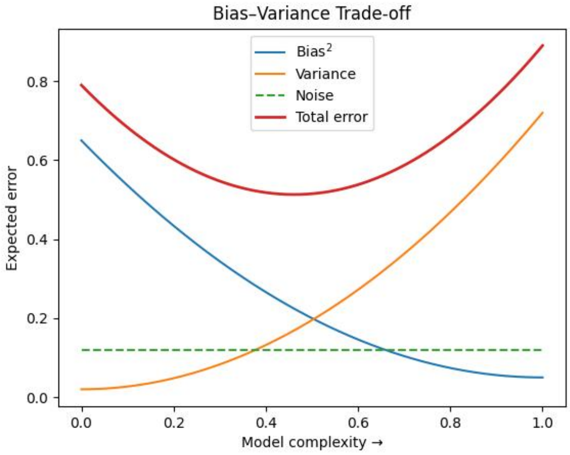

Figure 1: Bias-Variance Trade-off illustrating how the error component changes with model complexity.

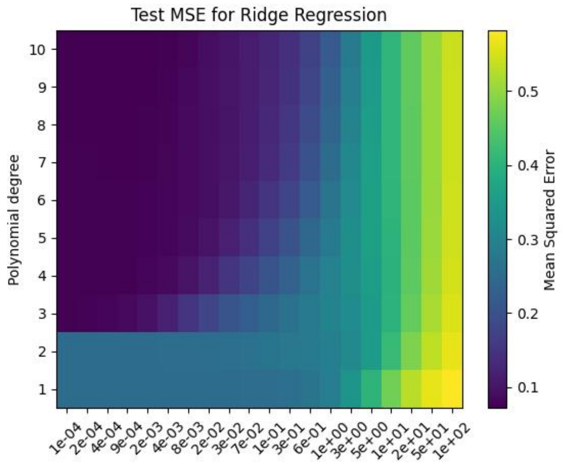

Figure 2: Heatmap of MSE for Ridge Regression across polynomial degrees and regularization strengths λ.

IV. CONCLUSION

- State your main findings and interpretations
- Try to discuss the pros and cons of the methods and possible improvements
- State limitations of the study
- Try as far as possible to present perspectives for future work

[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] 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).