

Project 1

Isak O. Rukan* and Edvard B. Rørnes†

*Institute of Physics, University of Oslo,
0371 Oslo, Norway*

(Dated: September 26, 2024)

We investigate and compare three regression models^a: Ordinary Least Squares, Ridge, and Least Absolute Shrinkage and Selection Operator. These were applied and analyzed on a two-dimensional Franke function and later tested on real terrain data from the US Geological Survey Earth Explorer [1]. The models' performances are evaluated using metrics such as the Mean Squared Error and the coefficient of determination R^2 . The bias-variance trade-off for OLS is studied in detail where it was found that ... (TBD). We used resampling techniques such as bootstrapping and cross-validation to probe the quality of the evaluations and determine the predictiveness and generalizability of the models. In our analysis the best performing method was ... (TBD). **I find it weird that we are supposed to be writing the abstract before actually concluding all the results. In general the abstract should be written last...shouldn't it?**

1. INTRODUCTION

Regression models are essential tools in data analysis and prediction, particularly within the realm of physics. They are used to enable understanding of the relationships between different variables and improve our ability to create predictions. In this report we study and compare three common regression techniques: Ordinary Least Squares (OLS) regression, Ridge regression and Least Absolute Shrinkage and Selection Operator (LASSO) regression. Each come with their unique strengths and weaknesses that make them suitable for different data sets.

OLS is the simplest and most foundational method which estimates relationship by minimizing the difference between observed and predicted values. A large downside with OLS is when there are many related variables as this can lead to unstable coefficient estimates [2]. Ridge regression can partially fix this issue by adding a penalty to large coefficients, effectively shrinking them. This in turn creates a more stable model in the event of correlated variables. LASSO regression takes this a step further by once again shrinking coefficients, but also setting some of them to zero. This allows LASSO to effectively choose important variables, which may be helpful when

pursuing a simpler model.

To compare these regression model, we use a two-dimensional Franke function which allows us to test the performance of each under controlled conditions. Later we apply each model to real terrain data which is obtained from [1].

The models are evaluated by considering their Mean Squared Error (MSE) and coefficient of determination R^2 . Further to make our evaluation more reliable, resampling techniques such as bootstrapping and cross-validation are used. These methods provide a better understanding of how well each model performs on different data.

2. THEORY

3. IMPLEMENTATION

4. RESULTS

5. DISCUSSION

6. CONCLUSION

[1] U.S. Geological Survey, "EarthExplorer: Terrain Elevation Data." <https://earthexplorer.usgs.gov>, 2024. Accessed: 20/09/2024. Extracted terrain elevation data.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 1. Springer, 2006.

* Insert Email

† e.b.rornes@fys.uio.no

^a GitHub Repository: <https://github.com/EdvardRornes/>