Machine learning methods and their implications for the linear regression problem
FYS-STK3155 - Project 1

Simen L. Wærstad*
John-Magnus Johnsen
University of Oslo
(Dated: September 26, 2025)

We investigate and implement various solutions to the linear regression problem, namely ordinary least squares (OLS), Ridge regression, and Lasso regression, and compare them by modeling the Runge function. We also study the effects on the regression models of using stochastic and non-stochastic gradient descent and different algorithms for updating the learning rate, including momentum, ADAgrad, RMSprop and ADAM. In addition, we derive the bias-variance trade-off and discuss resampling techniques and cross-validation. We find that OLS performs better than ridge when using the analytical solutions. In contrast, when using gradient descent we find that ridge out-performs OLS for the right choices of the shrinkage parameter λ. LASSO, however, outperforms both models in the high complexity cases, having both lower mean squared error and higher R^2. In our case of the Runge function, the choice of gradient descent method had no effect in the model accuracy and a minuscule effect on the computation time, where ADAM performed slightly better than the other methods.

I. INTRODUCTION

The linear regression problem shows up in a myriad of fields where we want to understand the underlying relationship between data. In particular, we are often interested in using recorded data to create a model so that we can predict the result of future observations. There exists many different types of models but we will only consider ordinary least squares (OLS), ridge, and LASSO.

In the case of both OLS and ridge regression there exists analytical solutions to these problems, allowing us to directly derive the optimal model parameters for the given data. This is, however, not the case for LASSO and many other methods[1]. For these models we have to use non-analytical methods. One such class of methods is gradient descent methods.

Gradient descent methods treat the parameter space as a multidimensional surface where we, as the name suggests, use the gradient of this surface to choose parameters that locally minimizes (or nearly locally minimizes) the error of the model. These methods work iteratively by computing the gradient at the current position and moving a specified distance in a direction determined by the gradient. This gives us a numerical approach that does not require the model to have analytical solutions.

Various modifications to gradient descent exists. We will consider momentum, ADAgrad, RMSprop and ADAM. Each variation makes modifications to how the position is updated, either by tweaking the direction or the step size (learnig rate).

Stochastic gradient descent is another variation of gradient descent. This method chooses a random subset of the training data used to compute the gradient. This decreases the computational cost of calculating the gradient at each step and is particularly useful if you have large data sets or a high dimensional parameter space.

We will study the different regression models and how the various gradient descent methods affect them.

We also discuss the relation between the bias, variance and the mean squared error of a model. This relationships is helpful in evaluating wether the model is over- or under-fitted.

Finally, a discussion of resampling techniques and cross validation is included.

Utilizing the open source Scikit-learn library for machine learning in Python [2].

Following the derivation presented in Chapter 3.2, page 44. [1]

When you write the introduction you should focus on the following aspects:

- Motivate the reader, the first part of the introduction gives always a motivation and tries to give the overarching ideas. Citing some central ideas or problems in the literature is a good idea here. [3][4][1, 5]

- What you have done, with a focus on choice of problem and method, and why these were chosen.

- The structure of the report, how it is organized. List the sections, and very briefly describe what is in them and how they fit together.

---

*https://github.com/Waerstad/4155-Project-1

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

- Describe how AI tools like ChatGPT were used in the production of the code and report.

The use of LLM, primarely Microsoft Copilot, was helpfull in finding kode for heatmaps and making plots latex friendly.

III. RESULTS AND DISCUSSION

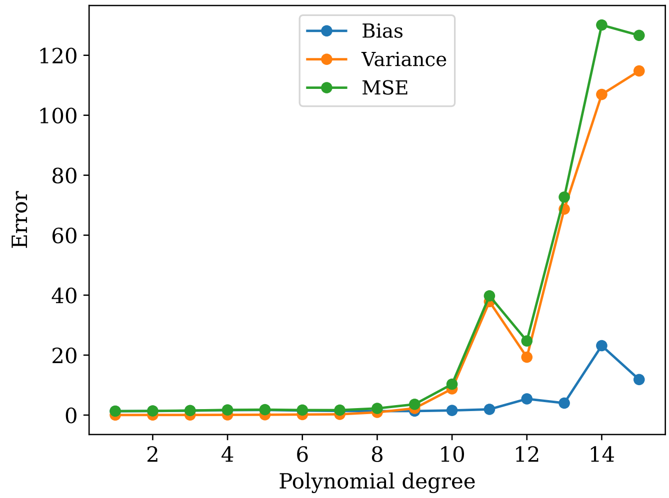

Figure 1: Visualization of the bias-variance tradeoff for linear regression with OLS for a variety of fitting polynomials. Here for a dataset of 100 points, resampled 100 times through bootstrapping. We see a clear point where increased model complexity does not benefit the bias noteworthy, but instead comes at the cost of an increased variance and error to the predicted values.

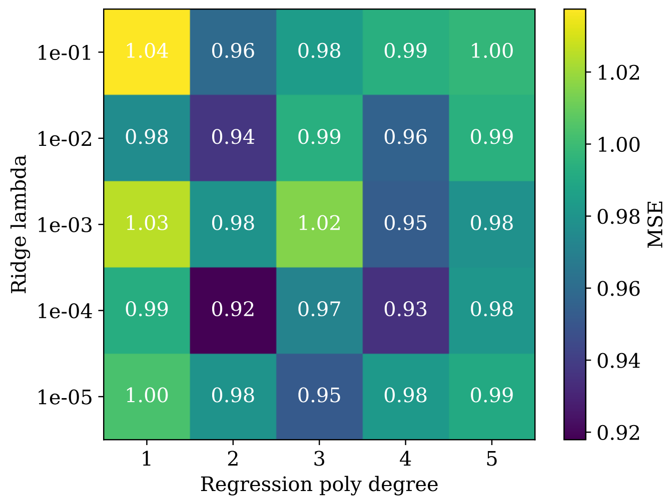

Figure 2: MSE for Ridge regression of the Runge function for polynomials of degree 1 to 5 and ridge parameter λ, averaged over 100 runs.

- 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 figures?? 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.

As we see in Figure 1.

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), URL `https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html`.

[3] M. Hjorth-Jensen, *Computational Physics Lecture Notes 2015* (Department of Physics, University of Oslo, Norway, 2015), URL `https://github.com/CompPhysics/ComputationalPhysics/blob/master/doc/Lectures/lectures2015.pdf`.

[4] H. J. T. Zhang Yi, Yan Fu, Computers and Mathematics with Applications **47**, 1155 (2004), URL `https://doi.org/10.1016/S0898-1221(04)90110-1`.

[5] K. B. Hein, *Data Analysis and Machine Learning: Using Neural networks to solve ODEs and PDEs* (Department of Informatics, University of Oslo, Norway, 2018), URL `https://compphysics.github.io/MachineLearning/doc/pub/odenn/html/._odenn-bs000.html`.