Report Template
FYS-STK3155 - Project X

Anton Nicolay Torgersen
University of Oslo
(Dated: September 27, 2025)

A study of various regression methods, including OLS, Ridge and LASSO and how they fit Runge's function. We review the theory and implementation of these methods and see how they compare on this function. At the end, we discuss some resampling techniques on the simpler OLS-method to understand the bias-variance trade-off.

I. INTRODUCTION

The fundamental problem in machine learning and statistics is to find meaningful patterns in data using models that make accurate predictions on unseen data. This is a difficult task, as a complex function might fit the training data perfectly, but fail spectacularly when facing new instances, a phenomenon known as overfitting. On the other hand a model that is too simple may fail to extract the patterns in the data and thus underfit the data. Therefore understanding of the bias-variance tradeoff on simpler models before the more complex ones is paramount to develop robust and generalizable models.

Therefore, this paper will study the bias-variance tradeoff through the lens of regression models, that increase in complexity. We start with the simplest model, Ordinary Least Squares (OLS), where we have a fully analytical solution for the best fit parameters. Then we move on to Ridge regression, which is a regularized version of OLS, where we first introduce the tuning of a penalization parameter λ to find a good balance between bias and variance. Finally we end with LASSO regression, which is another regularized version of OLS, but with a different penalization term.

This analysis will closely follow the lecture notes from FYS-STK3155 [1] and the book "The Elements of Statistical Learning" by Hastie, Tibshirani and Friedman [2]. We will review the theory of the three regression methods and apply the methods to a one-dimensional problem, fitting a polynomial model to Runge's function, $f(x) = 1/(1+25x^2)$. This function is notoriously difficult for high-degree polynomial interpolation and will therefore be a good demonstration of the pitfalls of overfitting and the benefits of regularized techniques like Ridge and LASSO. Together these three models will create a better understanding of the dynamics between bias and variance, how the dynamic shifts with model complexity and methods. The result is a better intuition for tuning models, when is the model inching towards the true function y and when is it stuck in a well.

II. METHODS

We will be assuming that the reader has some familiarity with linear algebra and multivariable calculus. For a more in-depth review of the methods and algorithms, including the motivation for using them and their applicability to the problem, please refer to the lecture notes [1] and the book "The Elements of Statistical Learning" by Hastie, Tibshirani and Friedman [2]. For the scoring measures Mean Squared Error (MSE) and R2, please refer to [**? ? ?**] for information or see the implementation in the code.

A. rungesfunction

Description of the function, why it is interesting to study.

B. Regression Methods

1. *Ordinary Least Squares (OLS)*

algorithm, cost function

Analytical solution

Gradient solution

2. *Ridge Regression*

algorithm, cost function

3. *LASSO Regression*

part f, algorithm, cost function

C. Gradient Descent

1. *Stochastic Gradient Descent*

part f, write about the theory?

D. Sampling methods

part g, write about the implementation of the bootstrap method as a resampling technique on the OLS method. part h, write about the implementation of the k-fold cross-validation algorithm as a resampling technique on the OLS, Ridge and LASSO methods.

E. Implementation

Implementation of the analytical solution of OLS regression using the numpy function pinv citation???.

1. *Regression methods*

perhaps simply a reference to the code for the implementation using numpy. And a test with the scikit library to see that the implementation is correct for these two.

Maybe some extra for the Lasso implementation?

2. *Gradient Descent*

part c, write about how we implementation of the gradient descent algorithm for OLS and Ridge regression. Again squeez it tohether we the above section and reference the code and functions.

Changing learning rate

part,d implementation of the changing learning rate algorithms: momentum, ADAgrad, RMSprop and Adam. Follows the implementations from the lecture notes from week 37 [1] and is explained in the code. Function bla bla bla

Stochastic Gradient Descent

part f, write about the implementation of the stochastic gradient descent algorithm for OLS and Ridge regression.

3. *Resampling techniques*

part g, write about the implementation of the bootstrap method as a resampling technique on the OLS method.

PROBABLY NOT.part h, write about the implementation of the k-fold cross-validation algorithm as a resampling technique on the OLS, Ridge and LASSO methods.

F. Use of AI tools

AI has been used to format and proofread the report and to spot errors in the code.

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

III. RESULTS AND DISCUSSION

Generation of the data set and why and how we scaled it. (part a) We had a set with and without noise in a normal distribution to see how the methods performed on both.

Analysis using the methods described in section II.

A. Regression methods

1. Ordinary Least Squares (OLS)

part a, study the analytical solution

2. Ridge Regression

part b, important to study the dependence on λ the regularization parameter

B. Using Gradient Descent

1. OLS and Ridge with gradient descent

part c,Study the results from the gradient descent algorithm for both OLS and Ridge regression and especially the dependence on the learning rate η and the number of iterations.

2. LASSO Regression

part f, study the results from the stochastic gradient descent algorithm for both OLS and Ridge regression and especially the dependence on the learning rate η and the number of iterations.

C. Changing learning rate algorithms

1. Changing learning rate

part d, study the effect of changing the learning rate η as a function of the number of iterations. THe methods used for this are momentum, ADAgrad, RMSprop and Adam. ON all these three regression methods.

2. Stochastic Gradient Descent

Now drag them into the stochastic gradient descent algorithm too see how they develop here.

D. Resampling Techniques

1. Bootstrap

part g, study the bias-variance tradeoff using the bootstrap method as a resampling technique on the OLS method.

2. Cross Validation

part h, k-fold cross-validation algorithm as a resampling technique on the OLS method. Compare the MSE you get from your cross-validation code with the one you got from your bootstrap code. Comment and interpret your results.

IV. CONCLUSION

* State your main findings and interpretations

* Try as far as possible to present perspectives for future work

* Try to discuss the pros and cons of the methods and possible improvements

[1] 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`.

[2] 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`.