

Weekly Exercises - FYS-STK3155

Oscar Atle Brovold

Week 39

Part *i* - Write an abstract for your project

Abstract

Linear regression is an essential tool for modeling the relationship between a dependent variable and one or more independent variables. This paper explores the predictive performance of different linear regression models in their ability to capture the features of a given landscape or terrain. In particular, we examine Ordinary Least Square, Ridge and Lasso regression. By employing resampling methods, such as bootstrap and cross-validation, in combination with the bias-variance tradeoff analysis, we provide a qualitative analysis of the strengths and limitations of these regression methods.

Part *ii* - Write an introduction

Linear regression is a fundamental statistical method used to predict the value of a dependent variable given one or more independent variables [1]. The method tracks back to 1805, when Legendre introduced the Ordinary Least Squares (OLS) approach [2], which remains one of the most used regression techniques today. Since then, additional linear regression techniques have been introduced, including Ridge regression and Lasso regression, which expand upon the principal of OLS.

All these regression methods aim to minimize a so-called cost-function, which quantifies how well the model's predictions align with the actual outcomes. The choice of cost-function is critical in determining model performance and will be examined in this paper.

Ridge regression and Lasso regression differ from OLS by incorporating a penalty to the cost-function. This is done by adding a regularization parameter, commonly denoted as λ . This becomes what we call a hyperparameter, its tuning significantly influences the model. Other key factors such as the polynomial degree of the model and the number of samples significantly affect the model's performance. An important part of this paper is determining the optimal values for λ and the polynomial degree that yield the best performance.

Moreover, data-preprocessing-particularly feature scaling- plays an essential role for the successful application of these methods. This paper will focus on evaluating model performance by tuning key parameters and examining the effects of preprocessing techniques like scaling.

References

- [1] IBM, "What is Linear Regression?" IBM, <https://www.ibm.com/topics/linear-regression>, accessed on September 26, 2024.
- [2] A.-M. Legendre, *Nouvelles méthodes pour la détermination des orbites des comètes*, Paris: F. Didot, 1805.