UNIVERSITY OF OSLO

Project 1

FYS-STK4155 – Applied Data Analysis and Machine Learning

Hishem Kløvnes

Gerlind Deschner

Juan Manuel Scarpetta

7 October 2024

Contents

Abstract

This project analyzes multiple regression techniques, including Ordinary Least Squares (OLS), Ridge, and Lasso, using Franke's function as a test case. The objective is to fit polynomial models to this two-dimensional function, evaluate model performance, and study the Bias-Variance trade-off. In addition to standard regression methods, model accuracy is assessed using resampling techniques such as bootstrap and cross-validation, focusing on the Mean Squared Error (MSE). Finally, these methods are applied to real terrain data, enabling comparison of results across various datasets and regression models. **[Add some results]**

1 Introduction

In the era of big data, predictive models are at the heart of decision-making across domains like finance, healthcare, and climate science. One of the key techniques driving these predictions is regression analysis, which identifies relationships between variables and enables us to predict how changes in one factor affect others. In this project, we explore multiple regression techniques—Ordinary Least Squares (OLS), Ridge, and Lasso—using Franke's function, a well-known test case for developing and evaluating regression methods.

Our goal is to understand how these techniques handle model complexity, noise, and the need for reliable metrics. In the training phase, resampling methods such as bootstrap and cross-validation are employed to ensure our models generalize well, avoiding overfitting while maintaining stability and robustness. These methods are indispensable in modern machine learning, where ensuring the reliability of models is critical. The Mean Squared Error (MSE) is used as the primary metric to assess model performance and compare the results of different regression techniques.

After rigorously testing these methods on the Franke function, we extend our analysis to real-world terrain data. This shift allows us to evaluate how well theoretical models perform on actual datasets, providing insight into the practical challenges of regression modeling. Although the Franke function is synthetic, it offers a rich testing ground for understanding polynomial fitting, and comparing it with real data further highlights the strengths and limitations of each regression approach.

In a world where predictive accuracy can have far-reaching consequences, this project underscores the critical role of foundational regression methods and resampling techniques in building reliable and robust models.