

Project 1
FYS-STK3155

Emilie Fjørner

September 10, 2020

Abstract

a)

Introduction

The main aim of this project is to study in more detail various regression methods, including the Ordinary Least Squares (OLS) method, Ridge regression and finally Lasso regression. Ridge regression will be discussed during the Friday lecture of week 36 while Lasso Regression will be discussed during the lectures of week 37.

The methods are in turn combined with resampling techniques like the bootstrap method and cross validation. These are discussed during weeks 36 and 37.

We will first study how to fit polynomials to a specific two-dimensional function called [Franke's function](#). This is a function which has been widely used when testing various interpolation and fitting algorithms. Furthermore, after having established the model and the method, we will employ resampling techniques such as cross-validation and/or bootstrap in order to perform a proper assessment of our models. We will also study in detail the so-called Bias-Variance trade off.

The Franke function, which is a weighted sum of four exponentials reads as follows

$$f(x, y) = \frac{3}{4} \exp \left(-\frac{(9x-2)^2}{4} - \frac{(9y-2)^2}{4} \right) + \frac{3}{4} \exp \left(-\frac{(9x+1)^2}{49} - \frac{(9y+1)^2}{10} \right) \\ + \frac{1}{2} \exp \left(-\frac{(9x-7)^2}{4} - \frac{(9y-3)^2}{4} \right) - \frac{1}{5} \exp \left(-(9x-4)^2 - (9y-7)^2 \right).$$

The function will be defined for $x, y \in [0, 1]$. Our first step will be to perform an OLS regression analysis of this function, trying out a polynomial fit with an x and y dependence of the form $[x, y, x^2, y^2, xy, \dots]$. We will also include bootstrap first as a resampling technique. After that we will include the cross-validation technique. As in homeworks 1 and 2, we can use a uniform distribution to set up the arrays of values for x and y , or as in the example below just a set of fixed values for x and y with a given step size. We will fit a function (for example a polynomial) of x and y . Thereafter we will repeat much of the same procedure using the Ridge and Lasso regression methods, introducing thus a dependence on the bias (penalty) λ .

Finally we are going to use (real) digital terrain data and try to reproduce these data using the same methods. We will also try to go beyond the second-order polynomials mentioned above and explore which polynomial fits the data best.

Methods

Results

Discussion

Conclusions and Perspectives

Appendix

Bibliography