Regression analysis and resampling methods

FYS-STK4155: Project 1

Morten Ledum & Hkon Kristiansen

September 24, 2018

Abstract

We parameterize digital terrain data using linear regression analysis algorithms: Ordinary least squares (OLS), Ridge regression, and Lasso regression. The bootstrap resampling technique is used to gauge the bias and variance of the models. We use basis sets of homogeneous monomials and polynomials in two variables, up to and including total degree 5. We find that xxxxxx.

For initial validation of our models, we employ the test function of R. Franke.[1]

1 Introduction

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

2 Theory

In the following we briefly introduce the theory underlying the technical aspects of the present work. We begin by considering linear regression in general, and the ordinary least squares (OLS) method.

2.1 Ordinary least squares

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

2.2 Ridge regression

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

2.3 Lasso regression

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

2.4 Resampling and the *Bootstrap* method

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

3 Data sets: The Franke function and U.S. Geological Survey terrain data

We are chiefly interested in parametrizing digital terrain data. However, in order to test and validate our implementation of the regression model and the resampling technique, we employ the Franke function[1] as a test case before considering real data.

3.1 The Franke function

The test function of Franke—originally developed to test and rate different surface interpolation techniques—is "a surface with a variety of behaviour" which consists of "two Gaussian peaks and a sharper Gaussian dip superimposed on a surface sloping towards the first quadrant."[1] It is noted by Franke in the his original paper that the slope was introduced mainly as a visual aid and presumably had little impact on the actual interpolations performed.

3.2 Terrain data

The terrain data used is taken from the U.S. Department of the Interior U.S. Geological Survey's (USGS) EarthExplorer[1] website. The USGS stores data from the Shuttle Radar Topography Mission (SRTM) which maps the earth's land surface topology with a resolution of 1 arc-second (about 30 m). We will use SRTM data taken from the EarthExplorer website as the basis for our terrain parametrization.

More specifically, the Franke function $f_{\mathrm{F}}(x, y)$ takes the full form

$$\begin{aligned} f_{\mathrm{F}}(x, y) = {} & \frac{3}{4} \exp\left\{\frac{-1}{4}\left[(9x-2)^2 + (9y-2)^2\right]\right\} \\ & + \frac{3}{4} \exp\left\{\frac{-1}{49}(9x+1)^2 + \frac{1}{10}(9y+1)^2\right\} \\ & + \frac{1}{2} \exp\left\{\frac{-1}{4}\left[(9x-7)^2 + (9y-3)^2\right]\right\} \\ & - \frac{1}{5} \exp\left\{\frac{-1}{4}\left[(9x+4)^2 + (9y-7)^2\right]\right\}. \end{aligned} \tag{1}$$

A plot of the $f_{\mathrm{F}}(x, y)$ surface can be seen in Fig. 1.

[1]EarthExplorer website: https://earthexplorer.usgs.gov/.

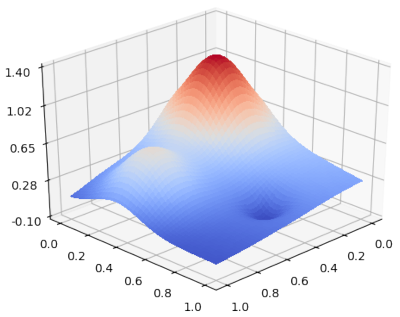

Figure 1: The Franke test function plotted for $0 \leq x, y \leq 1$.

4 Results and discussion

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

5 Conclusion

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

References

[1] Richard Franke. *A critical comparison of some methods for interpolation of scattered data.* Tech. rep. Monterey, California: Naval Postgraduate School., 1979.