Project 1 FYS-STK4155 Autumn 2017

Jon Audun Baar & Mikael Ravndal

28. september 2018

All of the plots of the models are done with degree equals 5.

Introduction

In this project we were to analyse different sets of data; test-data from the Franke-function and real data, using different linear regression models. The main goal of the project was to evaulate the different models, and come to a conclusion in terms of which models would fit best with which data.
Most of our code is in the GitHub repository. We have chosen to not use the code in our report.
We have, however, made a python file, project01.py, that provides most of our code in sequence, so that it is possible to follow the code that produce our tests and plots in order.

Ordinary least square on test data

First we have generated some test data which we have plotted to get a more intuitive feel for how it looks. We have done this for all the models.
Ordinary Least square of the Frankefunction(the same as Ridge with $\lambda = 0$):

a-OLS
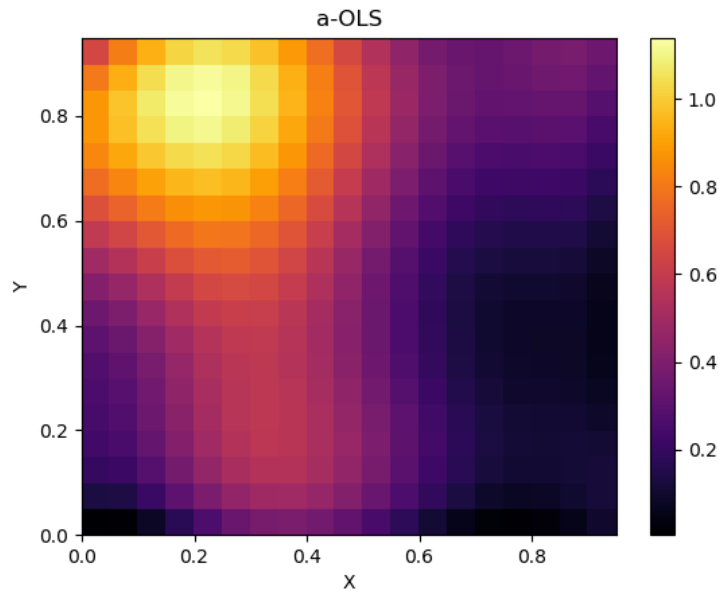
As you can tell from our code, we haven't made an own function for ordinary least square, since it is the same as Ridge, just with the λ set to 0. This will also show later that when the λ gets low it is very similar to OLS.
Something something variance etc

Resampling

Using our bootstrapping algorithm with a resampling of 10, we get these values:
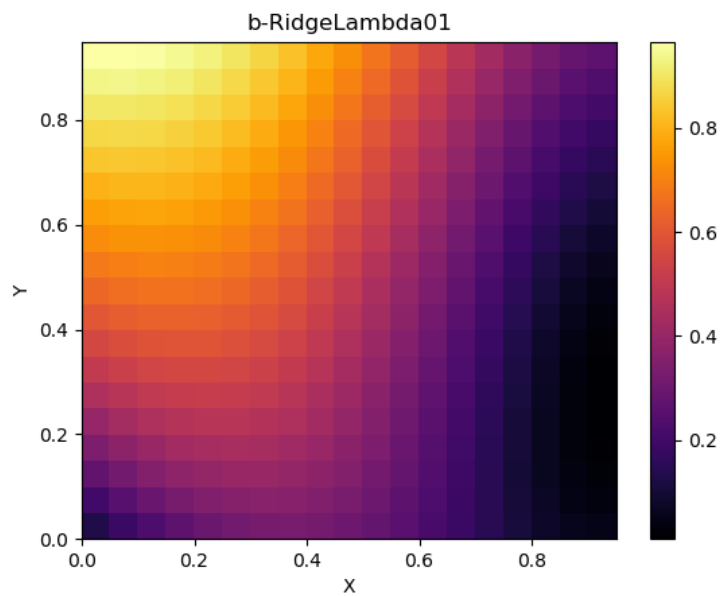
VAR: 6.920492.2.
BIAS: 945.386100.2.
Bootstrap mean of MSE: 0.001121.2.
Bootstrap mean of r2Score: 0.985516.2.
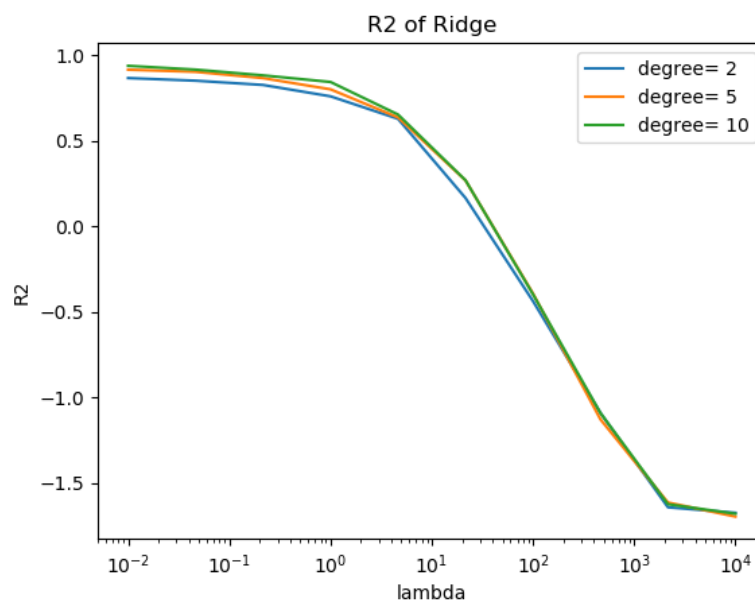
As we can tell

Ridge regression

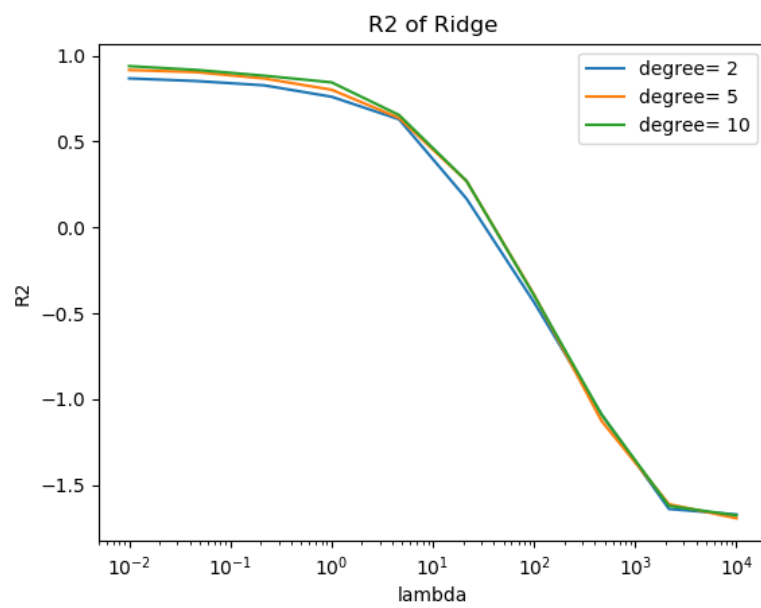Ridge Regression with $\lambda = 0.1$

Resampling

We can take a look at how different lambdas and different degrees of the polynomial makes a change in the R2-score and the MSE.
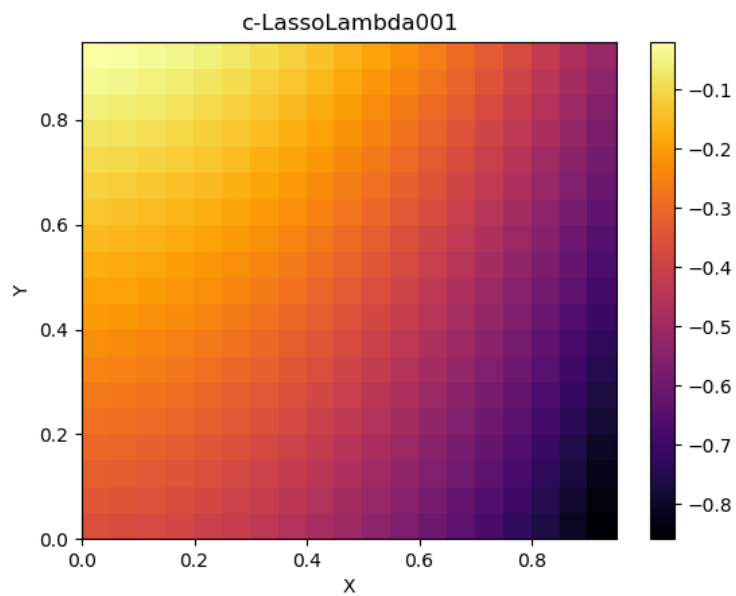Here is a plot to show how they develop as a function of lambda.

R2 of Ridge

We can tell pretty easily that the degree of the predictions doesn't matter much, but it varies heavily with lambda.
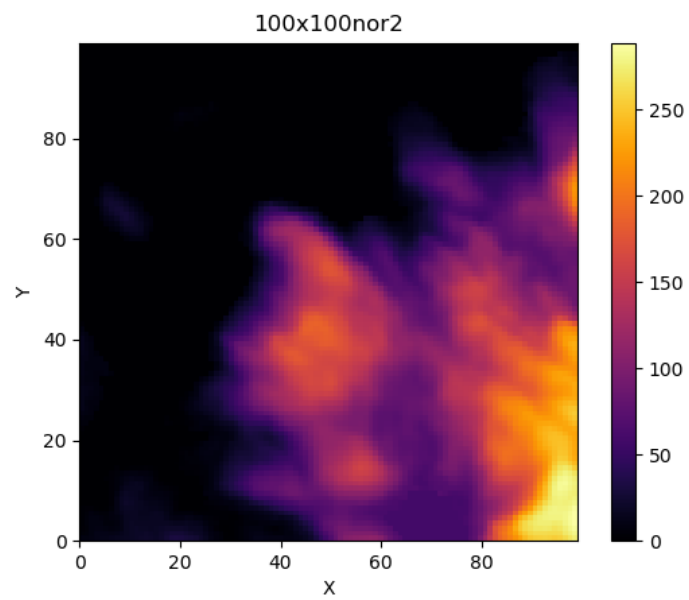
Part c)
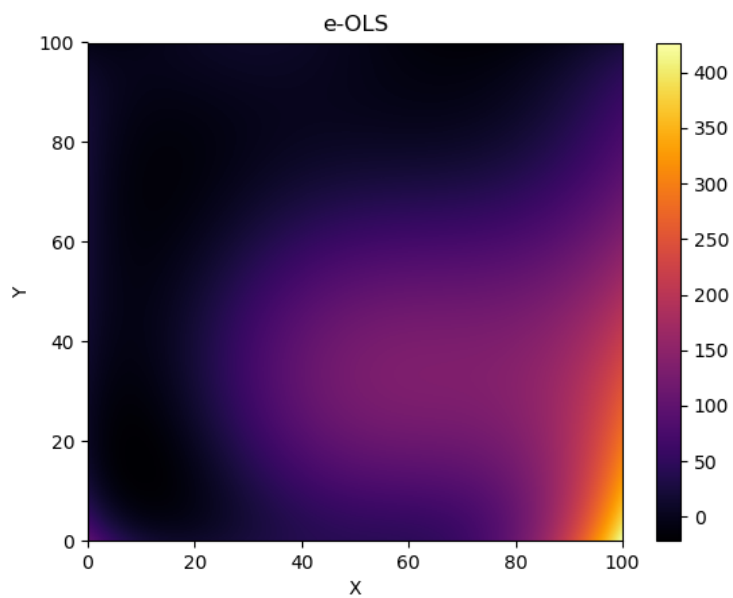
Lasso Regression with $\lambda = 0.01$

Part d)

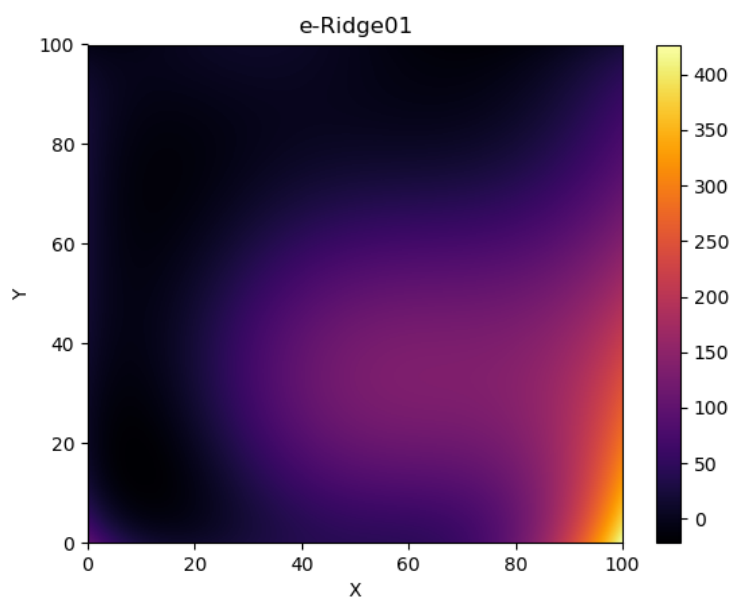Imports 100x100 chunk of real data from top left corner of dataset nr.1.
Plot of real data

Part e)

Repeat of a-c, but with real data from d)
OLS:

e-OLS

Ridge:

e-Ridge01

Lasso:

e-Lassolamda001