

Analysis of Regression and Resampling Methods

Håkon Olav Torvik, Vette Vikenes and Sigurd Sørle Rustad



University of Oslo
Norway
September 29, 2021

CONTENTS

I. Introduction	1
II. Ordinary Least Square	1
III. Bias-variance trade-off and Bootstrapping	1
IV. Cross-validation	1
V. Ridge Regression	1
VI. Lasso regression	1
VII. Analysis of real data	1
References	1

I. INTRODUCTION

Regression analysis is a statistical method for fitting a function to data. It is useful for building mathematical models to explain observations. There are several regression methods to achieve this, all with their strengths and weaknesses. We will in this paper study 3 different methods; ordinary least squares, Ridge and Lasso regression.

Larger datasets contain more information, giving more accurate models. However, real-world datasets usually have a fixed size, and getting more is practically impossible. For smaller datasets it is then useful to have tools mitigating the effects of little data. Resampling methods does this by running the same data through the regression, without over-fitting the model to the sample data. In addition to the regression methods, we will also study the effect of bootstrapping and cross-validating the data.

In order to study this, we need data to an-

alyze. Franke's function is a two-dimensional function resembling a topographical data. We will first do a polynomial fit with x and y dependence of this function. Because it is an analytical function, we can generate as much data as we need. We will therefore also do the same on topographical data from the real world, where the amount of data is limited.

II. ORDINARY LEAST SQUARE

III. BIAS-VARIANCE TRADE-OFF AND BOOTSTRAPPING

IV. CROSS-VALIDATION

V. RIDGE REGRESSION

VI. LASSO REGRESSION

VII. ANALYSIS OF REAL DATA

[1] Ref.