FYS-STK4155 project 1
Regression analysis and resampling methods

Mikael B. Kiste, Erik F. Skaar

September 26, 2018

Contents

Abstract

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

1 introduction

A typical problem within the natural sciences is how to interpret the trends and behavior of the results and data from an experiment. As a first approximation this will often be done qualitatively, e.g. "These values appear to increase linearly with time", but a more rigorous approach through regression analysis and resampling methods is eventually more preferable. Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

2 Theory

2.1 Ordinary least squares

We want to get a specific solution of the equation

$$\hat{y} = \hat{X}\hat{\beta} + \hat{\epsilon}$$

Where $\hat{y}$ is a vector of our measured values, $\hat{X}$ is a matrix containing variables and determines how we want to fit our data, $\hat{\beta}$ is a vector of the parameters for our fit and $\hat{\epsilon}$ is a vector representing the error in our datapoints. The variables $\hat{y}$ and $\hat{X}$ are fixed and we want to choose parameters $\hat{\beta}$ in such a way that the errors $\hat{\epsilon}$ are minimized. An example might help clarify the situation
Lets say we have conducted an experiment where we have measured the position of a ball launched straight up into the air from a cannon. Neglecting air resistance we know that the analytical solution is on the form of a second order polynomial $x(t) = x_0 + v_0 t + at^2$, where $x_0 = x(t = 0)$ is the position above ground of the ball before launch and $v_0 = v(t = 0)$ is the velocity of the ball immidiatly at launch. However, our measured results does not necessarily match this prefectly due to errors. In any case, if we measured the position n times our linear algebra problem could be

stated like this:

$$\begin{bmatrix} x(t_0) \\ x(t_1) \\ \vdots \\ x(t_{n-1}) \\ x(t_n) \end{bmatrix}$$

2.2 Ridge

2.3 Lasso

2.4 K-fold and and bootstrap