

Bayesian Lasso

Anonylise Wagner

Department of Statistics, University of Washington Seattle, WA, 98195, USA

Abstract

This is an abstraction of what I am doing. I should probably write this. For consideration, I would say, abstractly, that I need to fix my min commands.

1 Introduction

Linear regression is a broad problem with a myriad of proposed techniques for solving. At its heart, linear regression assumes that we have some vector of responses, \mathbf{y} , which depend linearly on some covariates, \mathbf{X} . Formally, accounting for an error term, we wish to fit the model

$$\mathbf{y} = \mu \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where \mathbf{y} is an $n \times 1$ vector, μ is the mean of \mathbf{y} , \mathbf{X} is a matrix of regressors (typically standardized), and $\boldsymbol{\epsilon}$ is a vector of independent and identically distributed zero mean normal variables with an unknown variance. For the sake of this discussion we will assume that the \mathbf{y} has zero mean ($\mu = 0$), noting that in practice this can be easily achieved by subtracting the sample mean estimate.

MAYBE PUT THE OLS SECTION IN BACKGROUND/MOTIVATION RATHER THAN INTRODUCTION What is really of interest is the regression coefficients, $\boldsymbol{\beta}$. These can be used for prediction when new values of regressors are given, or they may be of interest for their interpretation on the effects of certain regressor values on the dependent \mathbf{y} . Irrespective of the use, there are a number of approaches to finding the regressor coefficients. The first, ordinary least squares, seeks to minimize least squares error

$$\min \boldsymbol{\beta}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

While OLS estimates are a straight-forward and simple approach, they perform poorly in a variety of situations.

Penalized regression takes a similar approach, trying to minimize the square error of predicted values, but with further constraints on regression coefficients. Of particular interest is least absolute shrinkage and selection operator (Or Lasso), which penalizes the absolute value of the coefficients. For some penalty weight $\lambda \geq 0$, the Lasso has the form

$$\min \boldsymbol{\beta}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda \sum_{j=1}^p |\beta_j|.$$

2 Background and Motivation

3 Methods

4 Results

5 Discussion

References