Lasso estimation in multiple linear regression

Gregoire Gasparini, Aurora Hofman, Sarah Musiol, Beatriu Tort

03 de marzo de 2020

1. Lasso for the Boston Housing data

The Boston House-price dataset concerns housing values in 506 suburbs of Boston corresponding to year 1978. They are available here: https://archive.ics.uci.edu/ml/datasets/Housing

The Boston House-price corrected dataset (available in boston.Rdata) con- tains the same data (with some corrections) and it also includes the UTM coor- dinates of the geographical centers of each neighborhood.

Lasso estimation using the package 'glmnet'

After loading the right package, the response and explanatory variables from the Boston Housing data are set.

```r
#install.packages("glmnet")
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 3.0-2
```

```r
boston <- load("boston.Rdata")

response <- "CMEDV"

explanatory <- c("CRIM", "ZN", "INDUS", "CHAS", "NOX", "RM", "AGE", "DIS", "RAD", "TAX", "PTRATIO", "B"

# glmnet cannot handle factors -> "CHAS" is a factor
boston.c$CHAS <- as.numeric(boston.c$CHAS)
```

2. A regression model with $p >> n$