Ridge, Lasso and Elastic Net

January 24 2024

2024-02-07

## Task

Analyse the Credit and Hitters data sets as done in the Introduction to statistical learning using R and compare the following models: - The multiple linear regression; - The best models under forward selection, backward elimination and stepwise regression; - The ridge regression, the Lasso and elastic net models

Use caret

<http://www.sthda.com/english/articles/37-model-selection-essentials-in-r/153-penalized-regression-essentials-ridge-lasso-elastic-net/>

library(ISLR2)

## Warning: package 'ISLR2' was built under R version 4.3.2

data("Hitters")  
data("Credit")  
  
library(glmnet)

## Loading required package: Matrix

## Loaded glmnet 4.1-8

library(caret)

## Warning: package 'caret' was built under R version 4.3.2

## Loading required package: ggplot2

## Loading required package: lattice

train<-createDataPartition(Credit$Balance,p=.8,list = F)  
train.data<-Credit[train,]  
test.data<-Credit[-train,]  
  
y<-train.data$Balance  
X<-model.matrix(Balance~.,data=train.data)[,-1]  
# X  
  
# Cross-validation  
fit1<-cv.glmnet(X,y,alpha = 0)# Ridge regression  
fit2<-cv.glmnet(X,y,alpha = 1)# Lasso regression  
  
fit1$lambda.min

## [1] 39.77358

fit2$lambda.min

## [1] 0.6482103

fit1<-glmnet(X,y,alpha = 0,lambda = fit1$lambda.min)# Ridge regression  
fit2<-glmnet(X,y,alpha = 1,lambda = fit2$lambda.min)# Lasso regression  
coef(fit1)

## 12 x 1 sparse Matrix of class "dgCMatrix"  
## s0  
## (Intercept) -388.4125372  
## Income -5.2543949  
## Limit 0.1158275  
## Rating 1.6739489  
## Cards 14.8627814  
## Age -0.9427807  
## Education -1.7822835  
## OwnYes -10.6207063  
## StudentYes 385.6792605  
## MarriedYes -10.9722931  
## RegionSouth 14.1683796  
## RegionWest 14.7574225

coef(fit2)

## 12 x 1 sparse Matrix of class "dgCMatrix"  
## s0  
## (Intercept) -454.1646191  
## Income -7.7633367  
## Limit 0.2228866  
## Rating 0.6600982  
## Cards 18.1203454  
## Age -0.5962798  
## Education -2.3461475  
## OwnYes -12.2513449  
## StudentYes 424.3815553  
## MarriedYes -0.7269263  
## RegionSouth 14.7907436  
## RegionWest 14.2820850

<https://writingcenter.gmu.edu/writing-resources/imrad/writing-an-imrad-report> IMRaD # Introdution # Methods # Results # Discussion

**Task**

The classification file contains the analysis of the adequate maternal health care data set.

Using the following variables (or a well selected subset of them), develop a classification model using logistic regression, assess its performance and report your findings:

* province, peduc, educ, wealth, insurance, mediaexpo, mage, ethinicdiv, desirepreg, hospdelivery, blast5yrs, blastyr, childunder5, decisionm, occupation, nchilddead, childeverborn, region,, residence, religion,, sexhhead, status.