Lasso Regression

Jan-Philipp Kolb

03 Mai, 2019

Regularization

elitedatascience.com definition

Regularization is a technique used to prevent overfitting by artificially penalizing model coefficients.

- It can discourage large coefficients (by dampening them).
- It can also remove features entirely (by setting their coefficients to 0).
- The "strength" of the penalty is tunable. (More on this tomorrow. . .)

Wikipedia definition of Regularization

Regularization is the process of adding information in order to solve an ill-posed problem or to prevent overfitting.

Strenghts and weaknesses of regularization

Regularization is a technique for penalizing large coefficients in order to avoid overfitting, and the strength of the penalty should be tuned.

Strengths:

Linear regression is straightforward to understand and explain, and can be regularized to avoid overfitting. In addition, linear models can be updated easily with new data

Weaknesses:

Linear regression performs poorly when there are non-linear relationships. They are not naturally flexible enough to capture more complex patterns, and adding the right interaction terms or polynomials can be tricky and time-consuming.

Regularized Regression Algos

Lasso Regression   Ridge Regression   Elastic-Net

Lasso regression

LASSO, stands for least absolute shrinkage and selection operator

- Lasso regression penalizes the absolute size of coefficients.
- Practically, this leads to coefficients that can be exactly 0.
- Thus, Lasso offers automatic feature selection because it can completely remove some features.
- Remember, the "strength" of the penalty should be tuned.
- A stronger penalty leads to more coefficients pushed to zero.

Ridge regression

- Ridge regression penalizes the squared size of coefficients.
- Practically, this leads to smaller coefficients, but it doesn't force them to 0.
- In other words, Ridge offers feature shrinkage.
- Again, the "strength" of the penalty should be tuned.
- A stronger penalty leads to coefficients pushed closer to zero.

Elastic net

- Elastic-Net is a compromise between Lasso and Ridge.

-Elastic-Net penalizes a mix of both absolute and squared size. - The ratio of the two penalty types should be tuned. - The overall strength should also be tuned.

- There's no "best" type of penalty. It depends on the dataset and the problem.
- We recommend trying different algorithms that use a range of penalty strengths as part of the tuning process

Lasso regression overview

- lasso is a regression analysis method that performs variable selection and regularization (reduce overfitting)
- We want to enhance prediction accuracy and interpretability of the statistical model.

- We could remove less important variables, after checking that they are not important.
- We can do that manually by examining p-values of coefficients and discarding those variables whose coefficients are not significant.
- This can become tedious for classification problems with many independent variables

History of lasso

- Originally introduced in geophysics literature in 1986
- Independently rediscovered and popularized in 1996 by Robert Tibshirani, who coined the term and provided further insights into the observed performance.

Lasso was originally formulated for least squares models and this simple case reveals a substantial amount about the behavior of the estimator, including its relationship to ridge regression and best subset selection and the connections between lasso coefficient estimates and so-called soft thresholding. It also reveals that (like standard linear regression) the coefficient estimates need not be unique if covariates are collinear.

Lasso for other models than least squares

Though originally defined for least squares, lasso regularization is easily extended to a wide variety of statistical models including generalized linear models, generalized estimating equations, proportional hazards models, and M-estimators, in a straightforward fashion.

- Lasso's ability to perform subset selection relies on the form of the constraint and has a variety of interpretations including in terms of geometry, Bayesian statistics, and convex analysis.
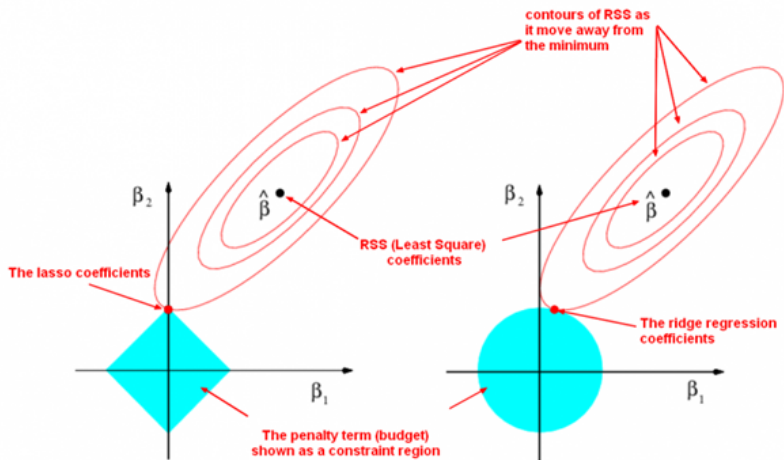
The LASSO is closely related to basis pursuit denoising.

What is lasso regression

- Lasso regression uses shrinkage
- data values are shrunk towards a central point
- Ridge and lasso regularization work by adding a penalty term to the log likelihood function.
- A tuning parameter, λ controls the strength of the L1 penalty.

$$\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| = RSS + \lambda \sum_{j=1}^{p} |\beta_j|.$$

Ridge Regression and the Lasso

Import of build in data

```
swiss <- datasets::swiss
```

First impression of the data

```
library(DT)
DT::datatable(swiss)
```

Preparing the model

```
x <- model.matrix(Fertility~., swiss)[,-1]
y <- swiss$Fertility
lambda <- 10^seq(10, -2, length = 100)
```

Test and train dataset

```r
library(glmnet)
set.seed(489)
train = sample(1:nrow(x), nrow(x)/2)
test = (-train)
ytest = y[test]
```

A first ols model

```
#OLS
swisslm <- lm(Fertility~., data = swiss)
coef(swisslm)
```

```
##      (Intercept)      Agriculture      Examination
##       66.9151817       -0.1721140       -0.2580082
##         Catholic Infant.Mortality
##        0.1041153        1.0770481
```

A ridge model

```
#ridge
ridge.mod <- glmnet(x, y, alpha = 0, lambda = lambda)
predict(ridge.mod, s = 0, type = 'coefficients')[1:6,]

##      (Intercept)      Agriculture      Examination
##        66.8911177       -0.1714307       -0.2603091
##          Catholic Infant.Mortality
##         0.1037196        1.0776950
```

Lasso regression with package `glmnet`

```
install.packages("glmnet")

library(glmnet)

x=matrix(rnorm(100*20),100,20)
g2=sample(1:2,100,replace=TRUE)
fit2=glmnet(x,g2,family="binomial")

caret::varImp(fit2,lambda=0.0007567)

##         Overall
## V1  0.150397416
## V2  0.208434920
## V3  0.633739581
## V4  0.546691236
## V5  0.040870939
## V6  0.131795416
## V7  0.039071452
```

- LASSO is a feature selection method.
- LASSO regression has inbuilt penalization functions to reduce overfitting.

- The logarithmic function is used for the link between probability and logits

- The Logit function is used to linearize sigmoid curves.

- Classification and Regression Training

```
install.packages("caret")
```

```
library("caret")
```

- **Vignette `caret` package** -

```
?caret::train

logit<-train(,data = gp.train.c,
                         method = 'glm',
                         family = 'binomial',
                         trControl = ctrl0)")
```

```
# https://cran.rstudio.com/web/packages/biglasso/biglasso.p
install.packages("biglasso")
```

A comprehensive beginners guide for Linear, Ridge and Lasso Regression

- Course for statistical learning - Youtube - Videos
- pcLasso: a new method for sparse regression
- Youtube - lasso regression - clearly explained
- Glmnet Vignette
- Regularization Methods in R
- A gentle introduction to logistic regression and lasso regularisation using R
- Penalized Regression in R
- Penalized Logistic Regression Essentials in R
- All you need to know about Regularization