Linear Regression Models

Segment 5 – Model Selection

Topic 3 – Regularization Approaches Using `glmnet`

Sudarsan N.S. Acharya (sudarsan.acharya@manipal.edu)

Topics

1. Regularization Overview

2. Regularization: Ridge, Lasso, & Elastic-net

3. The `glmnet` Package

4. Some Remarks on Ridge and Lasso Regularization

Regularization Overview

Regularization Overview

- The stepwise methods for model selection involved building several models using subsets of predictors.

Regularization Overview

- The stepwise methods for model selection involved building several models using subsets of predictors.
- An alternative approach for model selection is to build a model using all predictors while *shrinking* the coefficient estimates towards zero.

Regularization Overview

- The stepwise methods for model selection involved building several models using subsets of predictors.
- An alternative approach for model selection is to build a model using all predictors while *shrinking* the coefficient estimates towards zero.
- Shrinking the coefficient estimates can significantly reduce their variance.

Regularization Overview

- The stepwise methods for model selection involved building several models using subsets of predictors.
- An alternative approach for model selection is to build a model using all predictors while *shrinking* the coefficient estimates towards zero.
- Shrinking the coefficient estimates can significantly reduce their variance.
- Such shrinkage approaches are referred to as *regularization methods* which can also be used for feature selection.

Regularization: Ridge, Lasso, & Elastic-net

Regularization: Ridge, Lasso, & Elastic-net

Regularization is achieved by adding to the RSS constraints on the coefficient estimates:

Regularization: Ridge, Lasso, & Elastic-net

Regularization is achieved by adding to the RSS constraints on the coefficient estimates:

$$\underbrace{\sum_{i=1}^{n}\left(y^{(i)}-\left(\beta_0+\beta_1 x_1^{(i)}+\cdots+\beta_p x_p^{(i)}\right)\right)^2}_{RSS}$$

Regularization: Ridge, Lasso, & Elastic-net

Regularization is achieved by adding to the RSS constraints on the coefficient estimates:

$$\underbrace{\sum_{i=1}^{n}\left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_p x_p^{(i)}\right)\right)^2}_{RSS} + \left\{\vphantom{\int_0^1}\right.$$

Regularization: Ridge, Lasso, & Elastic-net

Regularization is achieved by adding to the RSS constraints on the coefficient estimates:

$$\underbrace{\sum_{i=1}^{n}\left(y^{(i)}-\left(\beta_0+\beta_1 x_1^{(i)}+\cdots+\beta_p x_p^{(i)}\right)\right)^2}_{RSS}+\begin{cases}\lambda\sum_{j=1}^{p}|\beta_j|^2 \\ \text{(ridge)} \\ \\ \\ \end{cases}$$

Regularization: Ridge, Lasso, & Elastic-net

MANIPAL ACADEMY of HIGHER EDUCATION (Institution of Eminence Deemed to be University)

Regularization is achieved by adding to the RSS constraints on the coefficient estimates:

$$\underbrace{\sum_{i=1}^{n}\left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_p x_p^{(i)}\right)\right)^2}_{RSS} + \begin{cases} \lambda \sum_{j=1}^{p} |\beta_j|^2 \\ \text{(ridge)} \\ \quad\text{or} \\ \\ \\ \end{cases}$$

Regularization: Ridge, Lasso, & Elastic-net

Regularization is achieved by adding to the RSS constraints on the coefficient estimates:

$$\underbrace{\sum_{i=1}^{n}\left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_p x_p^{(i)}\right)\right)^2}_{RSS} + \begin{cases} \lambda \sum_{j=1}^{p} |\beta_j|^2 \\ (\text{ridge}) \\ \qquad \text{or} \\ \lambda \sum_{j=1}^{p} |\beta_j| \\ (\text{lasso}) \end{cases}$$

Regularization: Ridge, Lasso, & Elastic-net

Regularization is achieved by adding to the RSS constraints on the coefficient estimates:

$$\underbrace{\sum_{i=1}^{n}\left(y^{(i)}-\left(\beta_0+\beta_1 x_1^{(i)}+\cdots+\beta_p x_p^{(i)}\right)\right)^2}_{RSS}+\begin{cases}\lambda\sum_{j=1}^{p}|\beta_j|^2\\ \text{(ridge)}\\ \quad\text{or}\\ \lambda\sum_{j=1}^{p}|\beta_j|\\ \text{(lasso)}\\ \quad\text{or}\end{cases}$$

Regularization: Ridge, Lasso, & Elastic-net

Regularization is achieved by adding to the RSS constraints on the coefficient estimates:

$$\underbrace{\sum_{i=1}^{n}\left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_p x_p^{(i)}\right)\right)^2}_{RSS} + \begin{cases} \lambda \sum_{j=1}^{p} |\beta_j|^2 \\ \text{(ridge)} \\ \quad\quad \text{or} \\ \lambda \sum_{j=1}^{p} |\beta_j| \\ \text{(lasso)} \\ \quad\quad \text{or} \\ \lambda \left[\alpha \sum_{j=1}^{p} |\beta_j| + (1-\alpha) \sum_{j=1}^{p} |\beta_j|^2\right] \\ \text{(elastic-net)}, \end{cases}$$

Regularization: Ridge, Lasso, & Elastic-net

Regularization is achieved by adding to the RSS constraints on the coefficient estimates:

$$\underbrace{\sum_{i=1}^{n}\left(y^{(i)}-\left(\beta_0+\beta_1 x_1^{(i)}+\cdots+\beta_p x_p^{(i)}\right)\right)^2}_{RSS}+\begin{cases}\lambda\sum_{j=1}^{p}|\beta_j|^2\\ \text{(ridge)}\\ \qquad\text{or}\\ \lambda\sum_{j=1}^{p}|\beta_j|\\ \text{(lasso)}\\ \qquad\text{or}\\ \lambda\left[\alpha\sum_{j=1}^{p}|\beta_j|+(1-\alpha)\sum_{j=1}^{p}|\beta_j|^2\right]\\ \text{(elastic-net)},\end{cases}$$

where $\lambda > 0$ is the strength of regularization.

The `glmnet` Package

The `glmnet` Package

- `glmnet` is a powerful R package for building generalized linear models with regularization

The `glmnet` Package

- `glmnet` is a powerful R package for building generalized linear models with regularization (linear,

The `glmnet` Package

- `glmnet` is a powerful R package for building generalized linear models with regularization (linear, logistic,

The `glmnet` Package

- `glmnet` is a powerful R package for building generalized linear models with regularization (linear, logistic, multinomial etc.)

The `glmnet` Package

- `glmnet` is a powerful R package for building generalized linear models with regularization (linear, logistic, multinomial etc.)
- The `glmnet` object contains all relevant information of the fitted model for further use.

The `glmnet` Package

- `glmnet` is a powerful R package for building generalized linear models with regularization (linear, logistic, multinomial etc.)
- The `glmnet` object contains all relevant information of the fitted model for further use.
- Additional methods for visualizing coefficients, prediction on new data, cross validation for selecting optimal regularization strength λ are available.

The `glmnet` Package

- `glmnet` is a powerful R package for building generalized linear models with regularization (linear, logistic, multinomial etc.)
- The `glmnet` object contains all relevant information of the fitted model for further use.
- Additional methods for visualizing coefficients, prediction on new data, cross validation for selecting optimal regularization strength λ are available.
- It is important to *scale the predictors* and *response variable* before running regularized regression using `glmnet`.

Some Remarks on Lasso and Ridge Regularization

Some Remarks on Lasso and Ridge Regularization

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Both ridge and lasso approaches for regularization shrink the coefficient estimates towards zero but lasso typically yields a much smaller subset of nonzero coefficient estimates.

Some Remarks on Lasso and Ridge Regularization

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Both ridge and lasso approaches for regularization shrink the coefficient estimates towards zero but lasso typically yields a much smaller subset of nonzero coefficient estimates.
- Thus, lasso regularization can be used for *feature selection*.

Some Remarks on Lasso and Ridge Regularization

- Both ridge and lasso approaches for regularization shrink the coefficient estimates towards zero but lasso typically yields a much smaller subset of nonzero coefficient estimates.
- Thus, lasso regularization can be used for *feature selection*.
- The drop in the magnitude of the coefficients in lasso regularization is not monotone.

Some Remarks on Lasso and Ridge Regularization

- Both ridge and lasso approaches for regularization shrink the coefficient estimates towards zero but lasso typically yields a much smaller subset of nonzero coefficient estimates.
- Thus, lasso regularization can be used for *feature selection*.
- The drop in the magnitude of the coefficients in lasso regularization is not monotone.
- By modifying the values of the regularization strength λ increases, we can trade-off the bias and variance of the model:

Some Remarks on Lasso and Ridge Regularization

- Both ridge and lasso approaches for regularization shrink the coefficient estimates towards zero but lasso typically yields a much smaller subset of nonzero coefficient estimates.
- Thus, lasso regularization can be used for *feature selection*.
- The drop in the magnitude of the coefficients in lasso regularization is not monotone.
- By modifying the values of the regularization strength λ increases, we can trade-off the bias and variance of the model: $\lambda = 0 \Rightarrow$ low bias, high variance, and

Some Remarks on Lasso and Ridge Regularization

- Both ridge and lasso approaches for regularization shrink the coefficient estimates towards zero but lasso typically yields a much smaller subset of nonzero coefficient estimates.
- Thus, lasso regularization can be used for *feature selection*.
- The drop in the magnitude of the coefficients in lasso regularization is not monotone.
- By modifying the values of the regularization strength λ increases, we can trade-off the bias and variance of the model: $\lambda = 0 \Rightarrow$ low bias, high variance, and $\lambda \to \infty \Rightarrow$ high bias, low variance.

Some Remarks on Lasso and Ridge Regularization

- Both ridge and lasso approaches for regularization shrink the coefficient estimates towards zero but lasso typically yields a much smaller subset of nonzero coefficient estimates.
- Thus, lasso regularization can be used for *feature selection*.
- The drop in the magnitude of the coefficients in lasso regularization is not monotone.
- By modifying the values of the regularization strength λ increases, we can trade-off the bias and variance of the model: $\lambda = 0 \Rightarrow$ low bias, high variance, and $\lambda \to \infty \Rightarrow$ high bias, low variance.
- Ridge regression has a unique OLS solution $\hat{\beta} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$.

Some Remarks on Lasso and Ridge Regularization

- Both ridge and lasso approaches for regularization shrink the coefficient estimates towards zero but lasso typically yields a much smaller subset of nonzero coefficient estimates.
- Thus, lasso regularization can be used for *feature selection*.
- The drop in the magnitude of the coefficients in lasso regularization is not monotone.
- By modifying the values of the regularization strength λ increases, we can trade-off the bias and variance of the model: $\lambda = 0 \Rightarrow$ low bias, high variance, and $\lambda \to \infty \Rightarrow$ high bias, low variance.
- Ridge regression has a unique OLS solution $\hat{\beta} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$.
- Optimal λ for ridge and lasso are typically not the same.

Summary

- Importance of regularization in model building
- Different regularization approaches
- `glmnet` package