

Homework 1 Predictive modeling

Due on 2021-04-16

2021-03-06

Notice

- All homework should be turned in pdf format.
- Homework can be written in Korean.

1 Ridge regression solution

Let $X \in \mathbb{R}^{n \times d}$ ($n \geq d$) denote a matrix with the singular value decomposition given by $X = U\Sigma V^T$, where $U \in \mathbb{R}^{n \times d}$ and $V \in \mathbb{R}^{d \times d}$ are orthonormal matrices satisfying $U^T U = I_d$ and $V^T V = I_d$, and $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_d)$ is a diagonal matrix with $\sigma_i \geq 0, i = 1, \dots, d$.

- 1) Simplify the following expression into products of three matrices, i.e, U , V and another diagonal matrix depending on σ_i and λ .

$$(\lambda I_d + X^T X)^{-1} X^T,$$

where I_d is an identity matrix of size $d \times d$.

- 2) We have learned that the optimal solution of ridge regression can be written as

$$\hat{\beta} = (\lambda I + X^T X)^{-1} X^T y.$$

Show that the Euclidean norm of the optimal solution $\|\hat{\beta}\|_2$ will decrease as λ increases.

2 Kaggle competition data

In this problem, we will examine the behaviors of Ridge regression and Lasso regression coefficients with respect to the penalty values in `tidymodels`.

- 1) Using the `tune_grid()` function in `tidymodels`, calculates the all the coefficients of ridge and lasso regression which corresponding penalty parameter for the train data set in [the advanced regression competition in Kaggle](#),

The preset penalty values are stored in the folloing code.

```
library(tidymodels)
grid_regular(penalty(), levels = 100)
```

- 2) Plot the coefficients vs. the specified penalty values above for ridge regression and lasso regression. You are expecting to see the following types of plots with respect to the penalty as follows.

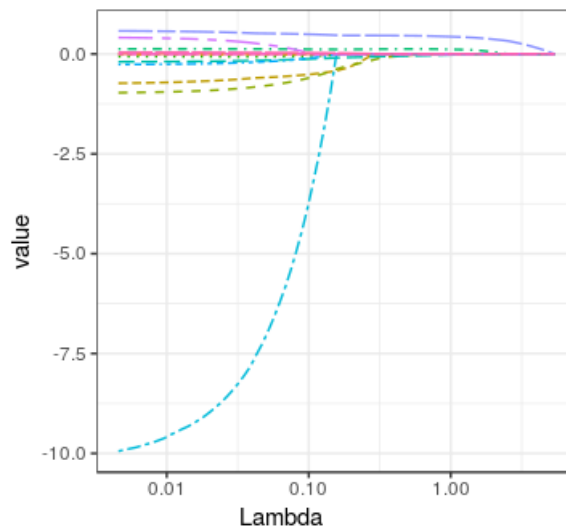


그림 1: coefficients value vs. penalty