

Exercise 3

Advanced Methods for Regression and Classification

November 17, 2022

Use again the `College` data (preferably transformed), and split it into training and test data. For the following tasks, always consider the RMSE as evaluation measure.

1. *Ridge Regression:*

- (a) Use the function `lm.ridge()` from the `library(MASS)` and apply it to the training data. Consider a range of values for the ridge parameter λ to find an optimal parameter. Plot the resulting GCV against the examined λ . Which λ is optimal, and which GCV error do you obtain? Note that GCV approximates the LOO cross-validation error.
- (b) Use the optimal ridge parameter for the model estimation. Which values do you obtain as regression coefficients?
- (c) Predict the response for the test data set, and compare the predictions with the reported values graphically. Compare with the results of the RMSE with those from the previous exercise.

2. *Lasso Regression:*

- (a) Use the function `glmnet()` from the `library(glmnet)` and apply it to the training data. Plot the result object. How can you interpret the plot? Which default parameters are used for `lambda`? What is the meaning of the parameter `alpha`?
- (b) Use the function `cv.glmnet()` and apply it to the training data. Visualize and interpret the results. How do you obtain the optimal tuning parameter and the regression coefficients?
- (c) Use the optimal model to predict the response for the test data. Compare the predictions with the reported values graphically. Compare the RMSE results with the results from Ridge regression.