Exercise 3

Advanced Methods for Regression and Classification

November 8, 2018

Consider the data `Hitters` from the package `ISLR`, see last exercise. Again, our goal is to predict the variable `Salary`. Remove all observations which contain missing values.

For the following tasks, split the data randomly into training and test data (about equal halves), build the model with the training data, and evaluate for the test data (using the MSE as a criterion).

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 do you obtain? What is the meaning of GCV?
   (b) Use the optimal ridge parameter for the model estimation. Which values do you obtain as regression coefficients?
   (c) Predict the Salary for the test data set, and compare the predictions with the reported values graphically and with the MSE. Compare with the results 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 Salary for the test data. Compare the predictions with the reported values graphically and with the MSE. Compare with the results from Ridge regression.

Save your (successful) R code together with short documentations and interpretations of results in a text file (= R script file), named as *Matrikelnummer_3.R* (no word document, no plots). Submit this file to Exercise 3 of our tuwel course (deadline November 7).