Problem Set 8

Ridge Regression

Consider the following data generating process with $n = 500$ observations and p covariates. Initially set the number of predictors $p = 10$ and $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$. $\mathbf{\Sigma}$ is diagonal and contains values ranging from 1 to 10 in some sequence. The true coefficients range from $\boldsymbol{\beta} = 0 - 1$ (you can sample values from that range or use equi-spaced values on that interval) and the errors are drawn from a normal distribution $\boldsymbol{\epsilon} \sim \mathcal{N}(0, 1)$.

The aim of this exercise: compare the predictions generated by OLS and Ridge regression.

- Implement the ridge regression estimator using glmnet for a grid of different penalty parameters.
- Draw a test data set and plot the test error for the different values of λ.

- Choose the optimal λ using the build in cross-validation function from glmnet and calculate the mean squared test error drawing a new test data set and compare Ridge regression and OLS.

```
###For calculating the ridge regression coefficients you need to install
###the glmnet package and call the library###########################
library(glmnet)
#####For calculating the prediction error you can use the predict() function##
```