

Practical 5

Statistical Learning, 2011

Niels Richard Hansen
September 21, 2011

Ridge and Lasso regression

Download the LA ozone data set from the book homepage. Below, you are asked to regress the *cube root* of the ozone concentration on the other variables.

1. Divide the data set into two groups at random. One group, which we call the training data, containing 2/3 of the observations and one group, which we call the test data, with 1/3 of the observations. Use only the training data below for the estimation.
2. Compute the best model for each dimension (best subset selection), estimate test error and compute the training error. Make a plot.
3. Compute the ridge and lasso paths. Estimate the test error, compute the training error and plot them as a function of the penalization parameter.
4. Compute the C_p statistic for best subset selection and ridge regression and compare with the estimated test errors.