MTH6101 Introduction to Machine Learning

Laboratory week twelve

The intention of this laboratory is to first have another look at the lasso methodology. To this end, we will use a small synthetic data set whose Lasso analysis exhibits unusual features. We then later concentrate on elastic nets penalization, using the same example.

When you **start** your session, open `RStudio` and install/load the following libraries: `lars` and `glmnet`. The dataset to be used is located in the qmplus page.

1. Begin by loading the dataset from the file `smalldata.txt`. Examine this data set and note that the last column will be used as response variable.

   Lasso analysis

2. Carry out a lasso fit using the function `lars` from the same package. Do not forget to select `type="lasso"` and set the parameters `intercept` and `normalize` to `FALSE`. Once this model has been fit, plot the lasso path. Describe the path starting from the least squares estimate $\hat{\boldsymbol{\beta}}$.

3. Examine the table of coefficients and link the contents of it to your description. Add the values of λ and the shrinkage achieved $||\boldsymbol{\beta}(\lambda)||_1/\max_\lambda||\boldsymbol{\beta}(\lambda)||_1$ to this table.

4. Use the function `coef.lars` to retrieve coefficients between the breakpoints 5 and 6 of the Lasso fit. Use 11 uniformly spaced values between these breakpoints.

5. Add these interpolated coefficients to your lasso plot.

   Glmnet

6. This second part of the lab examines the net penalty $g(\alpha, \boldsymbol{\beta}) = \big(\alpha||\boldsymbol{\beta}||_1 + (1-\alpha)||\boldsymbol{\beta}||_2^2\big)$. To this end, load the library `glmnet` and do a simple `glmnet` fit to the same data. Plot the path and compare with the previous result.

7. Using the same function `glmnet` and a range of values for the parameter `alpha`, plot the paths for elastic nets.