Programming (ERIM)

2. Exercise

Deadline for submission: 2013-11-03 23:59 CET

Instructions / ordinary least squares

In this exercise we will implement a linear regression model estimation using matrix operations (`http://en.wikipedia.org/wiki/Linear_regression`). Let us consider a multiple linear regression model with n observed responses on p regressors. Then the model can be stated as $n \times 1$ vector of responses $\mathbf{y}$, $p \times 1$ vector of unknown variables $\boldsymbol{\beta}$, $n \times p$ matrix matrix of regressors $\mathbf{X}$, and $n \times 1$ vector of error variables $\boldsymbol{\epsilon}$ with:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

The constant term is included in the set of regressors $\mathbf{X}$ by taking $x_{i1} = 1 \; \forall i \in \{1, \ldots, n\}$. The coefficient β_1 corresponding to this regressor is the intercept. Minimizing the sum of squared residuals of this model provides us the OLS estimator (`http://en.wikipedia.org/wiki/Ordinary_least_squares`) for $\boldsymbol{\beta}$ with the following formula:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y} \tag{2}$$

Exercise

Download the data set from `http://smaa.fi/static/prog2/2011/data/food.dat`. The dataset consists of 40 observations of household weekly incomes (in 100$'s, first column) and food expenditures (in $'s, second column), and two other variables that are not of interest to us. Make a program that

1. Implements Ordinal Least Squares estimator (Eq. 2) as a function taking as parameters:
   (a) matrix of observations
   (b) column index of the variable to be predicted
   (c) set of column indices of the regressor variables to use

   and returning the $\hat{\beta}$.
2. Loads the data set into R (read.table) / Matlab (load)
3. Calculates parameters of the regression model using the OLS estimator
4. Plots the data points as a scatter diagram together with the regression line. The x-axis should be the weekly income and y-axis the food expenditures. Name the axes appropriately and color the data points plotted in black and the regression line in red. Include the functional form of the regression line next to it in the diagram.

For more information about matrix computations in Matlab, see `http://www.mathworks.nl/help/techdoc/learn_matlab/f2-8955.html`.

Instructions / submission

Include in each source file your names and student numbers as a comment in the beginning of the file. Submit the exercise as a zip file containing only the source files (.m or .R) in root of the zip. Submit via Blackboard. Note that incorrectly submitted or non-running exercises are automatically awarded 0 points. Remember to document your code and use descriptive variable names.