Anna Titova, M.Sc.

Data Mining
Home Assignment 2

Consider following random variables

$$u \sim N(0,1) \qquad x_1 \sim N(0,1) \qquad x_2 \sim N(0,1)$$

and a linear regression model

$$y = 1 + 2x_1 + u.$$

Your task is to write an `R` script that contains the following parts. But first, download the script template *HA2_ yourname.R* from OLAT.

1. Generate samples of size $n = 100$ of all random variables, including y.
2. Compare the fits of the following linear models:

$$\begin{aligned} y =& \beta_0 + u \\ y =& \beta_0 + \beta_1 x_1 + u \\ y =& \beta_0 + \beta_2 x_2 + u \\ y =& \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \end{aligned}$$

   using K-fold cross-validation with $K = 5$ and $K = 10$. Do not use any already implemented procedures like `cv.glm`, but rather write the procedures from scratch.
3. Fit the best models according to the results of the cross-validation.
4. For the estimated coefficients calculate the bootstrapped standard errors without using the `boot()` function. Set the number of repetitions to 1000.
5. Comment on your results.

Remarks: Write comments for everything you do. Codes that are not written using the template and/or that return error messages will not be evaluated.
Submission: Submit your scripts via email to *atitova[at]stat-econ.uni-kiel.de* until the end of June 20th (until 00:00:00, June 21st)