Using R for regression model selection with adaptive penalties procedures based on the False Discovery Rate (FDR) criteria

Tal Galili*, Yulia Gavrilov and Yoav Benjamini
Department of Statistics and Operations Research · Tel Aviv University
***galilita@tau.ac.il**

Keywords: model selection, Linear regression models, Forward model Selection, FDR adaptive penalties, Multiple testing

Background: Multiple hypothesis testing has become an integral component in numerous modern age statistical challenges, ranging from microarray genomic analyses to fMRI brain scans. The control the False Discovery Rate (FDR) has proven to be an effective tool in addressing the multiple testing in large problems. Similarly, recent research has introduced the FDR approach to the problem of model selection, when the number of potential variables is very large**.**

In our presentation we discuss an implementation in R of a set of methods applying forward variable selection to linear regression models, based on penalized FDR controlling procedures such as the procedure in Benjamini-Hochberg(1995) [**] and the adaptive one in Benjamini and Gavrilov (2009) [**]**.**

We first review the theoretical and empirical merits of the proposed FDR based procedures. Theoretically [*] - It is asymptotically minimax for l^2 loss simultaneously throughout a range of sparsity classes for an orthogonal design matrix. Empirically [***] - it showed good performance over other penalized methods in a recent comprehensive simulation study. The study was conducted with a wide range of realistic settings, including non-orthogonal explanatory variables and it compared the methods using empirical minimaxity relative to a 'random oracle' (the oracle model selection performance on data dependent forward selected family of potential models.)

We proceed with results from our ongoing research on implementing these methods to general linear model, emphasizing their implementing within other R packages (such as MASS, leaps, biglm, ff and more) with the possibility of creating a dedicated package for their use**.**

References

[*] Abramovich, F., Benjamini, Y., Donoho, D. and Johnstone, I. (2000) 'Adapting to Unknown Sparsity by Controlling the False Discovery Rate', Technical Report No. 2000-19, May 2000**.**
[**] Benjamini, Y., Hochberg, Y., (1995) " Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing ", Journal of the Royal Statistical Society B, 57 289-300.
[***] Benjamini, Y., Gavrilov, Y., (2009) 'A SIMPLE FORWARD SELECTION PROCEDURE BASED ON FALSE DISCOVERY RATE CONTROL', Annals of Applied Statistics (in press).