Selective Inference with Application to L_2-Boosting

David Rügamer
(Institut für Statistik, LMU München)

November 30, 2016, 16:30 s.t.
Department of Statistics, Ludwigstr. 33, Room 144, Munich

Although it is commonly known that model selection invalidates classical statistical inference conducted post selection, many researchers perform (data driven) model selection and infer about the parameter as if the model was chosen in advance. Initialized by the proposal for valid statistical inference after arbitrary selection procedures by Berk et al. (2013), many new findings for inference after model selection have been published in the last years. In this context, important work was done by Tibshirani et al. (2014). Constructing a framework for any sequential regression technique that results in a limitation to the space for inference, which can be characterized by a polyhedral set, Tibshirani et al. paved the way for more general approaches.

In this talk we explain the problem with conducting classical inference after model selection and present fundamentals of *selective inference*, which corrects the inference after the selection procedure. We apply the polyhedron approach by Tibshirani et al. (2014) to L_2-Boosting in order to perform inference after the L_2-Boosting selection procedure. However, this approach conditions on most of the information in the data, thereby leaving no information for a powerful inference post selection. We therefore motivate the use of an adaption of the approach by Yang et al. (2016), which is a powerful and more coherent way to perform selective inference for L_2-Boosting. By circumventing an explicit mathematical definition of the selective inference space, our algorithm is a flexible tool, which is usually less expensive than Bootstrap and may also be used to incorporate resampling schemes or stability selection.