

# Model selection criterion and its consistency in boosting for high-dimensional sparse linear models

Ching-Kang ING

*Institute of Statistical Science, Academia Sinica*

A fundamental difficulty in model selection for high-dimensional regressions is that the number of observations is much less than the number of candidate variables. Recently, Zhao and Yu (2006) and Meinshausen and Bühlmann (2006) showed that LASSO is consistent for variable selection when the number of candidate variables grows exponentially fast with the number of observations. However, to show consistency, these authors needed to impose the so called "irrepresentable" condition, which seems to be a rather stringent restriction. To avoid assuming this condition yet still achieve consistency in high-dimensional regressions, we propose a two-stage model selection criterion combining  $L_2$ Boosting and a modified information criterion.  $L_2$ Boosting plays a key role in the first stage to screen variables, whereas the modified information criterion is used in the second stage to get a further refined set. A theoretical justification of this two-stage criterion is given. In addition, simulations in image sparse representations are used to illustrate the advantages of the proposed approach. Finally the relative performance of LASSO and the proposed criterion is presented through several numerical examples.