Variable selection with error control:
another look at Stability Selection

Richard John Samworth

Abstract

Stability Selection was introduced in a recent JRSSB discussion paper by Meinshausen and Buhlmann (2010) as a very general technique designed to improve the performance of a variable selection algorithm. It is based on aggregating the results of applying a selection procedure to subsamples of the data. We introduce a variant, called Complementary Pairs Stability Selection (CPSS), and derive bounds both on the expected number of variables included by CPSS that have low selection probability under the original procedure, and on the expected number of high selection probability variables that are excluded. These results require no (e.g. exchangeability) assumptions on the underlying model or on the quality of the original selection procedure. Under reasonable shape restrictions, the bounds can be further tightened, yielding improved error control, and therefore increasing the applicability of the methodology.