COMBINING LINEAR REGRESSION MODELS: WHEN AND HOW?

by

Zheng Yuan and Yuhong Yang
University of Michigan and Iowa State University

June 2003

ABSTRACT

Model combining (mixing) methods have been proposed in recent years to deal with uncertainty in model selection. Even though advantages of model combining over model selection have been demonstrated in simulations and data examples, it is still unclear to a large extent when model combining should be preferred. In this work, firstly, an instability measure to capture the uncertainty of model selection in estimation, named PIE, is proposed based on perturbation of the sample. It is demonstrated that estimators from model selection can have large PIE values and model combining substantially reduces the instability for such cases. Secondly, we propose a model combining method, ARMS, and derive a theoretical property. In ARMS, a screening step is taken to narrow down the list of candidate models before combining, which not only saves computing time but also can improve estimation accuracy. Thirdly, we compare ARMS with BMA (Bayesian Model Averaging) and model selection methods in a number of simulations and real data examples. The comparison shows that model combining produces better estimators when the instability of model selection is high and ARMS performs better than BMA in most such cases in our simulations. With respect to the choice between model selection and model combining, we propose a rule of thumb in terms of PIE. The empirical results support that PIE is a sensible indicator of model selection instability in estimation and is useful for understanding whether model combining is a better choice over model selection for the data at hand.