

Evaluating and selecting models for prediction out-of-sample

Hannes Leeb
Yale University

In regression with random design, we study the problem of selecting a model that performs well for out-of-sample prediction. We focus on a statistically challenging scenario where the number of potentially important explanatory variables can be infinite, where no regularity conditions are imposed on unknown parameters, where the number of explanatory variables in a 'good' model can be of the same order as sample size, and where the number of candidate models can be of larger order than sample size.