Model Selection Criteria
Based on Computationally Intensive Estimators of the Expected Optimism

Joseph E. Cavanaugh*
Department of Biostatistics, University of Iowa, U.S.A.
joe-cavanaugh@uiowa.edu

A model selection criterion is often formulated by constructing an approximately unbiased estimator of an expected discrepancy, a measure that gauges the separation between the true model and a fitted candidate model. The expected discrepancy reflects how well, on average, the fitted candidate model predicts "new" data generated under the true model. Its natural estimator, the estimated discrepancy, reflects how well the fitted candidate model predicts the data at hand. The expectation of the difference between these measures is known as the expected optimism. The selection criterion arises by adding an approximation of the expected optimism to the estimated discrepancy to correct for its negative bias. Classical approaches to obtaining this approximation often lead to simplistic penalty terms primarily based on the dimension of the candidate model. However, such approaches generally involve large-sample arguments, restrictive assumptions on the form of the candidate model, or both. The resulting penalty terms may fail to perform adequately in small-sample applications or in settings where the requisite assumptions do not hold. In this talk, we propose computationally intensive approaches to approximating the expected optimism. We discuss the properties of some of the resulting criteria and illustrate their use in identifying biosignatures for treatment response.

Key Words: Bootstrap, cross validation, Monte Carlo simulation, variable selection