

Small Sample Inference for Generalization Error

Eric Laber

January 12, 2008

Abstract

Many classification problems take place in a data-impooverished setting where an independent test set is unavailable. The absence of a test set typically requires some kind of data resampling (cross-validation, bootstrap, etc.) in order to estimate the generalization error. Given this point estimate, how does one attach measures of confidence? In the described setting, it is unclear how to estimate the distribution of the generalization error. Instead, we propose a method for constructing confidence sets which relies on an upper bound on the difference between the estimated and true generalization error. Confidence sets are constructed by bootstrapping this upper bound. This is in contrast to bootstrapping the generalization error directly, which can lead to unacceptable type I errors. Furthermore, we provide a computationally efficient algorithm for computing these confidence sets when the classifier is approximated by a parametric additive model, which is fit by minimizing a convex surrogate for 0-1 loss. In a preliminary simulation study, our method shows favorable performance to several standard methods on a catalog of 10 test sets.