Title: A Prediction Divergence Criterion for Model Selection

Abstract:

The problem of model selection is a crucial part of any statistical analysis. In fact, model selection methods become inevitable in an increasingly large number of applications involving partial theoretical knowledge and vast amounts of information, like in medicine, biology or economics. These techniques are intended to determine which variables are “important” to “explain” a phenomenon under investigation. The terms “important” and “explain” can have very different meanings according to the context and, in fact, model selection can be applied to any situation where one tries to balance variability with complexity. For example, these techniques can be applied to select “significant” variables in regression problems, to determine the number of dimensions in principal component analyses or simply to construct histograms. In this respect, we introduce a new class of error measures and of model selection criteria. Moreover, a novel criterion, called the Prediction Divergence Criterion Estimator, is derived from these two classes and we demonstrate that, under some regularity conditions, it is asymptotically loss efficient and can also be consistent. This new criterion is shown to be particularly well suited in “sparse” settings which we believe to be common in many research fields such as Genomics and Proteomics. Our selection procedure is developed for linear regression models.