Cross Validation and Localized Model Selection

Yuhong Yang, University of Minnesota

Abstract

Cross validation (CV) is commonly used for models selection. However, there is wide spread confusion on choice of a CV method (e.g., data splitting ratio) for application. We will examine the consistency property of CV for comparing two general regression/classification procedures and share our understanding on proper use of the CV approach.

So far, research on model/procedure selection in statistical learning has focused on selecting a single model globally. In many applications, especially for high-dimensional or complex data, however, the relative performance of the candidate procedures typically depends on the location, and the globally best procedure can often be much improved when selection of a model/procedure is allowed to depend on location. We develop methods for localized model/procedure selection and derive their theoretical properties.