

Title:

The Estimation of Prediction Error: Covariance Penalties and Cross-Validation

Author(s):

Bradley Bfron

Technical Report number (Dept. of Statistics, Stanford Univ.):

2003-27B/224

Date:

July 2003

Abstract:

Having constructed a data-based estimation rule, perhaps a logistic regression or a classification tree, the statistician would like to know its performance as a predictor of future cases. There are two main theories concerning prediction error: (1) penalty methods such as C_p , AIC, and SURE that depend on the covariance between data points and their corresponding predictions; (2) Cross-validation and related nonparametric bootstrap techniques. This paper concerns the relationship between the two theories. A Rao-Blackwell type of relation is derived, in which nonparametric methods like cross-validation are seen to be randomized versions of their covariance penalty counterparts. The model-based penalty methods offer substantially better accuracy, assuming that the model is believable.