

A Prediction Interval for the Misclassification Rate

*Susan Murphy**
University of Michigan

When small data sets are used for classification it is crucial that we provide some measure of confidence for the estimated misclassification rate. However the misclassification rate is a non-smooth function of the classifier. Furthermore the estimated rate suffers from bias due to overfitting and the is classification rate is a "minimized" quantity. For all of these reasons the construction of measures of confidence such as estimates of variance and confidence/prediction intervals are challenging. We discuss this problem and propose a method based on the use of a smooth upper bound combined with the bootstrap. This upper bound utilizes the surrogate loss that is used in the construction of the classifier.

*Co-author Eric Laber