THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

RYAN TIBSHIRANI

Department of Statistics
Carnegie Mellon University

Excess Optimism: How Biased is the Apparent Error of an Estimator Tuned by SURE?

MONDAY, September 25, 2017 at 4:30 PM

Eckhart 133, 5734 S. University Avenue

Refreshments before the seminar at 4:00PM in Jones 111

ABSTRACT

Nearly all estimators in statistical prediction come with an associated tuning parameter, in one way or another. Common practice, given data, is to choose the tuning parameter value that minimizes a constructed estimate of the prediction error of the estimator; we focus on Stein's unbiased risk estimator, or SURE (Stein, 1981; Efron, 1986), which forms an unbiased estimate of the prediction error by augmenting the observed training error with an estimate of the degrees of freedom of the estimator. Parameter tuning via SURE minimization has been advocated by many authors, in a wide variety of problem settings, and in general, it is natural to ask: what is the prediction error of the SURE-tuned estimator? An obvious strategy would be simply use the apparent error estimate as reported by SURE, i.e., the value of the SURE criterion at its minimum, to estimate the prediction error of the SURE-tuned estimator. But this is no longer unbiased; in fact, we would expect that the minimum of the SURE criterion is systematically biased downwards for the true prediction error. We define the excess optimism to be the amount of this downward bias in the SURE minimum. We argue that the following two properties motivate the study of excess optimism: (i) an unbiased estimate of excess optimism, added to the SURE criterion at its minimum, gives an unbiased estimate of the prediction error of the SURE-tuned estimator; (ii) excess optimism serves as an upper bound on the excess risk, i.e., the difference between the risk of the SURE-tuned estimator and the oracle risk (where the oracle uses the best fixed tuning parameter choice). We study excess optimism analytically in various settings (e.g., the families of shrinkage and subset regression estimators), and demonstrate how bounds on excess optimism lead to oracle inequalities. We also describe a bootstrap method for estimating excess optimism in practice, and outline some extensions of our framework beyond the standard, homoskedastic Gaussian error model considered throughout.

This represents work with Saharon Rosset.

_____