STATISTICS COLLOQUIUM

MONDAY, NOVEMBER 26, 2007
TALK: 4:00 PM — SCIENCE CENTER RM. 309
RECEPTION: 5:15 PM — SCIENCE CENTER, 7TH FLOOR

"Local minimax estimation, fusion in machine learning and overcoming the curse of dimensionality"

Lee K. Jones
Department of Mathematical Sciences
University of Massachusetts Lowell

ABSTRACT

A local finite sample minimax approach to regression is described and compared to the global minimax counterpart in the mathematics of learning of Stone, Cucker –Smale and DeVore-Kerkyacharian-Picard-Temlyakov. While the global approach stresses the importance of approximation theory in statistical inference, the local paradigm poses challenges in both approximation theory and computational algebra (in particular in the implementation of the Tarski quantifier elimination principle). It provides optimal accuracy bounds (for the given finite sample), based on assumptions about the behavior of the regression function f(x), for several new and old algorithms for learning f(xo) for a fixed query xo.

Empirical evidence indicates that averaging the learned estimate of f(x) from many different algorithms (estimates) often leads to increased accuracy - as with random subspace methods, random forests and varying kernel covariances. Furthermore, when statistically analyzed, different algorithms are often estimating different conditional expectations of the response. The method of fusion in local minimax estimation optimally combines the estimates from a given set (ensemble) of such algorithms and obtains an accuracy bound on such combination. It is presently unclear (to the speaker) how to modify the concept of fusion for the global minimax method. A simple lower dimensional example of the advantages of fusion is given.

The local minimax approach allows the fusion procedure to account for the possibility of spurious features causing (with a known high probability) only a small number of algorithms to overfit. When a large number of the algorithms are highly predictive the fusion estimate remains accurate and the curse of dimensionality may be avoided. An example using random forests is presented.