

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
DEPARTMENTAL SEMINAR

4:15 p.m., Tuesday, November 23, 2004
Sequoia Hall Room 200
(Cookies at 3:45 in 1st Floor Lounge)

Peter Bartlett
University of California, Berkeley

Empirical minimization and risk bounds
Stanford-Berkeley Joint Colloquium

We consider empirical risk minimization methods, which choose a function from a class to minimize the sample average of some loss function. This is an important family of methods for pattern classification and other prediction problems. We present bounds on the risk (expected loss) of the empirical risk minimizer under mild assumptions on the function class. These bounds are in terms of various notions of complexity of the function class that are defined in terms of a function ξ that maps from a value r to the supremum of the empirical process indexed by functions with expectation r (that is, to the maximal deviation between sample averages and expectations over these functions). The first complexity measure is the fixed point of this function. We show that it is possible to estimate this fixed point from data. However, the bound on the empirical minimization algorithm can be improved further by a direct analysis. We show that the maximizer of the difference $\xi(r) - r$, which is smaller than the fixed point, is always close to the risk of the empirical minimizer. We give an example showing that the difference between these estimates can be significant. However, another example shows that, in general, the difference between the improved bound and the fixed point cannot be estimated from data. Our proofs are based on Talagrand's concentration inequality for empirical processes.

This talk describes joint work with Olivier Bousquet, Shahar Mendelson and Petra Philips.