The Annals of Statistics
2006, Vol. 34, No. 6, 2664–2671
DOI: 10.1214/009053606000001037
Main article DOI: 10.1214/009053606000001019

DISCUSSION: LOCAL RADEMACHER COMPLEXITIES AND ORACLE INEQUALITIES IN RISK MINIMIZATION

BY GILLES BLANCHARD AND PASCAL MASSART

Fraunhofer FIRST, Berlin and Université de Paris-Sud

These last years, much attention has been paid to the construction of model selection criteria via penalization. Vladimir Koltchinskii has to be congratulated for providing a theory reaching a level of generality that is sufficiently high to recover most of the recent results obtained on this topic in the context of statistical learning. Thanks to concentration inequalities and empirical process theory, we are now at a point where the problem of understanding what is the order of the excess risk for the empirical minimizer on a given model is elucidated. Koltchinskii's paper provides several ways of expressing that this excess risk can be sharply bounded by quantities depending on the complexity of the model in various senses. The most prominent relies on Rademacher processes, which Vladimir Koltchinskii himself pioneered in introducing in statistics. We even know that these upper bounds on the excess risk are often unimprovable (see the lower bounds in [6], e.g.).

The same machinery used to analyze the excess risk can be applied to produce penalized criteria and to establish oracle-type risk bounds for the so-defined penalized empirical risk minimizer. The problem of defining properly penalized criteria is particularly challenging in the classification context, since it is connected to the question of defining optimal classifiers *without* knowing in advance the "noise condition" of the underlying distribution [(8.2) of the discussed paper]. This condition determines the attainable rates of convergence and is a topic attracting much attention in the statistical learning community at this moment (see the numerous references in the discussed paper).

What we would like to discuss is the gap between the theory and the practice of model selection. Of course, the existence of a gap between the methods which are analyzed in theory, and those which are used in practice, is in some sense unavoidable. Our purpose here is to express our perception of the current situation regarding this gap, and to propose some ideas which could contribute to reducing it.

As a starting point for our discussion, we would like to briefly analyze the behavior of the so-called *hold-out* selection procedure. This procedure should be seen as some primitive version of the V-fold cross-validation method, which is probably the most commonly used model selection method in practice, in the context of statistical learning. One advantage of hold-out is that it is very easy to study from

Received March 2006.