The Annals of Statistics
2006, Vol. 34, No. 6, 2688–2696
DOI: 10.1214/009053606000001073
Main article DOI: 10.1214/009053606000001019

DISCUSSION: LOCAL RADEMACHER COMPLEXITIES AND ORACLE INEQUALITIES IN RISK MINIMIZATION

BY SARA VAN DE GEER

ETH Zürich

1. Introduction. This paper unifies and extends important theoretical results on empirical risk minimization and model selection. It makes extensive and efficient use of new probability inequalities for the amount of concentration of the (possibly symmetrized) empirical process around its mean. The results are very subtle and very pleasing indeed, as they show that oracle inequalities exist for very general problems.

There are in my view two aspects which need special attention. First, the paper assumes that the loss functions $f \in \mathcal{F}$ satisfy $|f| \le K$ for some fixed constant K. Let us call this the *uniform bound condition* (Condition B below). Second, it is not clear how the approach used will work in practice: the estimators depend on (unspecified) constants which may be too large for all practical purposes, and moreover, it is difficult to explain the method to nonspecialists. This discussion will address these two problems.

We reformulate some of the results as a starting point for possible extensions or alternative approaches. For transparency, we will invoke simple, and not the most general, assumptions.

Section 2 in this discussion presents a distribution-dependent upper bound for the excess risk, replacing the uniform bound condition by convexity conditions and a bound on the renormalized loss functions (Condition BB).

The background of Section 3 in this discussion is the question whether cross-validation can be a more user-friendly model selection method than applying bounds in terms of Rademacher complexities. We first study why (data-dependent) upper and lower bounds for excess risks are useful when aiming at oracle behavior in model selection. We then show that when the margin behavior of the excess risk in each model is known, cross-validation can lead to oracle behavior.

Let us now first introduce our notation, following mostly that of the paper. Assume the observations $X_1, \dots, X_n$ are i.i.d. copies of a random variable $X \in S$ with distribution P. Let $\mathcal{F}$ be a given class of functions f on S. The empirical risk minimizer is

$$\hat{f} := \arg\min_{f \in \mathcal{F}} P_n f,$$

Received March 2006.