The Annals of Statistics
2006, Vol. 34, No. 6, 2677–2680
DOI: 10.1214/009053606000001055
Main article DOI: 10.1214/009053606000001019

DISCUSSION: LOCAL RADEMACHER COMPLEXITIES AND ORACLE INEQUALITIES IN RISK MINIMIZATION

BY XIAOTONG SHEN AND LIFENG WANG

University of Minnesota

Koltchinskii is to be congratulated for developing a unified framework. This elegant framework is general and allows a user to apply it directly instead of deriving bounds in each risk minimization problem. In the past decade, the problem of risk minimization has been extensively studied in function estimation and classification. In function estimation, it has been investigated using the empirical process technique under the name of minimum contrast or sieve estimation in, for instance, [2, 3, 9, 12, 13, 15]. In classification, it has been studied in a similar fashion; cf. [1, 6, 10, 11]. The general framework derived in this article yields an upper bound of the excess risk through local Rademacher complexities. When applying such a framework to a specific problem, attention is necessary with regard to the specific problem structure that may matter greatly.

Our discussion will be focused on in two aspects: (1) the role that the variance and mean play, particularly in classification, and (2) practicability of an empirical complexity.

1. The role of variance and mean.

1.1. *Variance–mean relationship and the margin condition.* As noted in the paper, one key idea to recover the optimal rate of convergence is to bound the local complexity $E\sup_{f\in\mathcal{F}:P(f-\bar{f})\le\delta}|(P_n-P)(f-\bar{f})|$ instead of the global one. This is achieved by bounding $Var(f-\bar{f})$, or sufficiently the second moment $P(f-\bar{f})^2$, by the mean $P(f-\bar{f})$. Such a variance–mean relationship was essentially used in [9] in a slightly more general form of

$$Var\big(f(X)-\bar{f}(X)\big)\le a\big[E\big(f(X)-\bar{f}(X)\big)\big]^{2\beta} \tag{1}$$

for some constants $a>0$ and $\beta>0$, where an iterative improvement approach is employed to derive fast rates of convergence by exploring $\sup_A|(P_n-P)(f-\bar{f})|$ over local sets A. This is analogous to the fixed point approach used in the present paper.

In what follows, we argue that (1) is more fundamental than the popular margin (low noise) assumption (cf. [10]) commonly used in classification, resulting in fast rates of convergence. In classification, (1) summarizes not only the

Received March 2006.