The Annals of Statistics
2006, Vol. 34, No. 6, 2681–2687
DOI: 10.1214/009053606000001064
Main article DOI: 10.1214/009053606000001019

DISCUSSION: LOCAL RADEMACHER COMPLEXITIES AND ORACLE INEQUALITIES IN RISK MINIMIZATION[1]

BY A. B. TSYBAKOV

Université Paris 6

The paper of Vladimir Koltchinskii has been circulating around for several years and already has become an important reference in statistical learning theory. One of the main achievements of the paper (further abbreviated as [VK]) is to propose very general techniques of proving oracle inequalities for excess risk under a control of the variance, that is, for example, under conditions (6.1) or (6.2) (often called margin or low noise conditions) or similar assumptions in terms of L_2-diameters $D_P(\mathcal{F}, \delta)$ and other related characteristics. These conditions lead to fast rates for the excess risk, that is, to rates that are faster than $n^{-1/2}$. The setup in [VK] is classical: methods based on empirical risk minimizers (ERM) $\hat{f}_n$ are studied under the bounded loss functions.

My comments and questions will be mainly about optimality of the excess risk bounds. This issue is not at all obvious, even in the case where the underlying class $\mathcal{F}$ is finite. We assume in what follows that either $\mathcal{F} = \{f_1, \dots, f_M\}$, where f_j are some functions on S, or this class is a convex hull $\mathcal{F} = \mathrm{conv}\{f_1, \dots, f_M\}$. Such classes $\mathcal{F}$ are used in aggregation problems where the functions f_j are viewed either as "weak learners" or as some preliminary estimators constructed from a training sample which is considered as frozen in further analysis.

Let $Z_1, \dots, Z_n$ be i.i.d. random variables taking values in a space $\mathcal{Z}$, with common distribution P, and denote by $\mathcal{F}_0$ the space where the f_j live. Consider a loss function $Q: \mathcal{Z} \times \mathcal{F}_0 \to \mathbb{R}$ and the associated risk

$$R(f) = \mathbb{E}Q(Z, f)$$

assuming that the expectation $\mathbb{E}Q(Z, f)$ is finite for all $f \in \mathcal{F}_0$ where Z has the same distribution as Z_i. Introduce two oracle risks: $R_{\mathrm{MS}} = \min_{1 \le j \le M} R(f_j)$ corresponding to model selection-type aggregation (MS-aggregation), and $R_{\mathrm{C}} = \inf_{f \in \mathrm{conv}\{f_1, \dots, f_M\}} R(f)$ corresponding to convex aggregation (C-aggregation). The excess risk of a statistic $\tilde{f}_n(Z_1, \dots, Z_n)$ is defined by

$$\mathcal{E}(\tilde{f}_n) = \mathbb{E}\{R(\tilde{f}_n)\} - R_{\mathrm{OR}},$$

where the oracle risk R_{OR} equals either R_{MS} or R_{C}. A natural question about optimality is how to find an estimator $\tilde{f}_n$ for which the excess risk is as small

Received March 2006.

[1]Supported by Miller Institute for Basic Research in Science, University of California, Berkeley.