Adaptive Nonparametric Classification

Alexandre B. Tsybakov
Université Paris VI

Abstract

We consider the classification in 2 classes based on the i.i.d. data $(X_1, Y_1), \ldots, (X_n, Y_n)$, with $Y_i \in \{0, 1\}$ and the predictors $X_i \in \mathbf{R}^d$. This problem can be viewed as a set estimation problem with a particular risk function: the Bayes regret. Devroye, Györfi and Lugosi (1996) proved that the Bayes regret of any classification rule cannot decrease faster than $1/\sqrt{n}$ in a minimax sense, as soon as there is no restriction on the class of joint distributions of (X, Y). The literature suggests a large variety of classification rules that converge with the rate $1/\sqrt{n}$ (up to a log-factor) or slower. These results take into account only the complexity of the underlying "candidate" boundaries between classes and neglect the structure of the margin (i.e. the behaviour of the regression function $\eta(x) = P(Y = 1|X = x)$ near the boundary curve $\{x : \eta(x) = 1/2\}$). The first result of this talk shows that the structure of the margin plays a crucial role in the convergence of the Bayes regret. In particular, fast rates, up to $1/n$, can be attained for "good" margins by using a simple empirical risk minimization rule or other familiar methods. The second result suggests a classifier which is adaptive both to the complexity of the boundary and to the margin. It is shown that this classifier:

- shares the properties of usual classifiers, i.e. attains the optimal rates up to $1/\sqrt{n}$ when there is no restriction on the joint distribution of (X, Y),
- attains the fast optimal rates up to $1/n$ (to within a log-factor) when the margins are " good", whatever is the complexity of the boundary in a given range.

These results explain the role of the margin in nonparametric classification and propose the first theoretically justified procedure that allows to adapt to the margin. Constructively, the suggested method is explicit but computationally difficult. It is based on multiple pre-testing schemes.