L2-Boosting in High- Dimensions: Rate of Convergence

Martin Spindler
Munich Center for the Economics of Aging

January 27, 2016, 16:00 Uhr s.t.
Department of Statistics, Ludwigstr. 33, Room 144, Munich

Boosting is one of the most significant developments in machine learning. This paper studies the rate of convergence of L_2Boosting, which is tailored for regression, in a high-dimensional setting. Moreover, we introduce so-called "post-Boosting". This is a post-selection estimator which applies ordinary least squares to the variables selected in the first stage by L_2Boosting. Another variant is orthogonal boosting where after each step an orthogonal projected is conducted. We show that both post-L_2Boosting and the orthogonal boosting achieve the same rate of convergence as Lasso in a sparse, high-dimensional setting. The "classical" L_2Boosting achieves a slower convergence rate for prediction, but no assumptions on the design matrix are imposed for this result in contrast to rates e.g. established with LASSO. We also introduce rules for early stopping which can be easily implemented and can be used in applied work. Moreover, our results also allow a direct comparison between LASSO and boosting which has been missing in the literature. Finally, simulation studies are presented to illustrate the relevance of our theoretical results and for providing insights to practical aspects of boosting. In the simulation studies post-L_2Boosting clearly outperforms LASSO.