Asymptotic Properties of High-Dimensional Random Forests

Jinchi Lv[1]; Chien-Ming Chi[1]; Yingying Fan[1]; Patrick Vossler[1]

[1] University of Southern California

Abstract:
As a flexible nonparametric learning tool, random forests has been widely applied to various real applications with appealing empirical performance, even in the presence of high-dimensional feature space. Unveiling the underlying mechanisms has led to some important recent theoretical results on the consistency of random forests algorithm and its variants. However, to our knowledge, all existing works concerning random forests consistency under the setting of high dimensionality were done for various modified random forests models where the splitting rules are independent of the response. In light of this, in this paper we derive the consistency rates of the original version of the random forests algorithm in a general high-dimensional nonparametric regression setting through a bias-variance decomposition analysis. Our new theoretical results show that random forests can indeed adapt to high dimensionality and provide new insights into the effect of column subsampling on the bias of random forests. In particular, for shallow random forests, our theory provides guidance and theoretical support on the tuning of column subsampling parameter. This is a joint work with Chien-Ming Chi, Yingying Fan and Patrick Vossler.

Keywords:
Random forests; Nonparametric learning; High dimensionality; Consistency; Rate of convergence; Sparsity