A flexible approach to high-dimensional conditional density estimation

Rafael Izbicki*
Federal University of São Carlos, São Carlos, Brazil - rafaelizbicki@gmail.com

Ann B. Lee
Carnegie Mellon University, Pittsburgh, USA- annlee@cmu.edu

Abstract

There is a growing demand for nonparametric conditional density estimators (CDEs) in fields such as astronomy and economics. In astronomy, for example, one can dramatically improve estimates of the parameters that dictate the evolution of the Universe by working with full conditional densities instead of regression (i.e., conditional mean) estimates. More generally, standard regression falls short in any prediction problem where the distribution of the response is more complex with multi-modality, asymmetry or heteroscedastic noise. Nevertheless, much of the work on high-dimensional inference concerns regression and classification only. Here we propose a fully nonparametric approach to conditional density estimation that reformulates CDE as a non-parametric orthogonal series problem where the expansion coefficients are estimated by regression. By taking such an approach, one can efficiently estimate conditional densities in high dimensions by drawing upon the success in high-dimensional regression. Depending on the choice of regression procedure, our method can adapt to a variety of challenging high-dimensional settings with, for example, a large number of irrelevant components or nonlinear manifold structure in the data. We study the theoretical and empirical performance of our proposed method, and we compare our approach with traditional conditional density estimators on real-world as well as simulated data.

Keywords: Nonparametric Statistics; Conditional Density; High-Dimensional Data.