Minimax-optimal Distribution Regression*

Zoltán Szabó (CMAP, École Polytechnique)

Abstract

In my talk, I am going to focus on the distribution regression problem (DRP): we regress from probability measures to Hilbert-space valued outputs, where the input distributions are only available through samples (this is the 'two-stage sampled' setting). Several important statistical and machine learning tasks can be phrased within this framework including point estimation tasks without analytical solution (such as hyperparameter or entropy estimation) and multi-instance learning. However, due to the two-stage sampled nature of the problem, the theoretical analysis becomes quite challenging: to the best of our knowledge the only existing method with performance guarantees to solve the DRP task requires density estimation (which often performs poorly in practise) and the distributions to be defined on a compact Euclidean domain. We present a simple, analytically tractable alternative to solve the DRP task: we embed the distributions to a reproducing kernel Hilbert space and perform ridge regression from the embedded distributions to the outputs. Our main contribution is to prove that this scheme is consistent in the two-stage sampled setup under mild conditions: we present an exact computational-statistical efficiency tradeoff analysis showing that the studied estimator is able to match the one-stage sampled minimax-optimal rate. This result answers a 17-year-old open question, by establishing the consistency of the classical set kernel [Haussler, 1999; Gaertner et. al, 2002] in regression. We also cover consistency for more recent kernels on distributions, including those due to [Christmann and Steinwart, 2010]. The practical efficiency of the studied technique is demonstrated in supervised entropy learning and aerosol prediction using multispectral satellite images.

Paper: Zoltn Szab, Bharath K. Sriperumbudur, Barnabs Pczos, Arthur Gretton. Learning Theory for Distribution Regression. Journal of Machine Learning Research, 17(152):1-40, 2016. (`http://jmlr.org/papers/v17/14-510.html`)
Code: in ITE toolbox (`https://bitbucket.org/szzoli/ite/`).

*Probability and Statistics Seminar, Orsay, France, 16 March 2017; abstract.