A Simple and Consistent Technique for Vector-valued Distribution Regression*

Zoltán Szabó (Gatsby Computational Neuroscience Unit, University College London)†

Abstract

I am going to tackle the regression task of vector-valued outputs from probability distributions in the two-stage sampled setting, when only sets of samples from the distributions are observable. The studied distribution regression problem (DRP) covers several important and challenging tasks in machine learning and statistics, including multi-instance regression or point estimation problems (such as hyperparameter identification). The inherent two-stage sampled nature of the setup makes the derivation of theoretical performance guarantees rather difficult: to the best of our knowledge the only available method from the large number of existing techniques performs density estimation (which typically performs poorly in practise), and restricts the problem to distributions with compact Euclidean support. In my talk, I will present a simple, ridge regression-based alternative to solving the DRP problem: we embed the distribution to a reproducing kernel Hilbert space, and learn the regressor from the embedded distribution to the outputs. We prove that under mild assumptions (on separable topological domains enriched with kernels), this scheme is consistent; moreover, we derive explicit rates of convergence in terms of the problem difficulty. Specifically, we prove that the set kernel is consistent in regression, which was a 15-year-old open, and demonstrate the efficiency of our method in supervised entropy learning and aerosol prediction based on multispectral satellite images.

Preprint: `http://arxiv.org/abs/1411.2066`

Code: `https://bitbucket.org/szzoli/ite/`

*Artificial Intelligence and Natural Computation seminars, University of Birmingham, UK; January 26, 2015; abstract.

†Joint work with Bharath K. Sriperumbudur (Department of Statistics, Pennsylvania State University), Barnabás Póczos (Machine Learning Department, Carnegie Mellon University), Arthur Gretton (Gatsby Computational Neuroscience Unit, University College London).