Consistent Vector-valued Regression on Probability Measures*

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

Abstract

I will focus on the distribution regression problem (DRP): our goal is to regress from probability measures to vector-valued outputs, in the two-stage sampled setup when only samples from the distributions are available. The studied DRP framework incorporates several important machine learning and statistical tasks, including multi-instance learning or point estimation problems without analytical solution (such as hyperparameter estimation). Obtaining theoretical guarantees, bounds on the generalization error of the estimated predictor is pretty challenging due to the two-stage sampled characteristic of the task. To the best of our knowledge, among the vast number of heuristic approaches in the literature, the only theoretically justified technique tackling the DRP problem requires that the domain of the distributions be compact Euclidean, and uses density estimation (which often performs poorly in practice). We present a simple, analytically tractable alternative: we embed the probability measures to a reproducing kernel Hilbert space, and perform ridge regression from the embedded distributions to the outputs. We prove that this method is consistent under mild conditions, on separable topological domains endowed with kernels. Specifically, we establish the consistency of the traditional set kernel in regression, which was a 15-year-old open question. We 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/`

*Tübingen, Germany; January 15, 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).