
Learning on Distributions*

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

Abstract

Problems formulated in terms of distributions have recently gained widespread attention. An important task that belongs to this family is distribution regression: regressing to a real-valued response from a probability distribution. One particularly challenging difficulty of the task is its two-stage sampled nature: in practise we only have samples from sampled distributions. In my presentation I am going to talk about two (intimately related) directions to tackle this difficulty. Firstly, I am going to present a recently released information theoretical estimators open source toolkit capable of estimating numerous dependency, similarity measures on distributions in a nonparametric way. Next, I will propose an algorithmically very simple approach to tackle the distribution regression: embed the distributions to a reproducing kernel Hilbert space, and learn a ridge regressor from the embeddings to the outputs. I will show that (i) this technique is consistent in the two-stage sampled setting under fairly mild conditions, and (ii) it gives state-of-the-art results on supervised entropy learning and the prediction problem of aerosol optical depth based on satellite images.

Preprint: <http://arxiv.org/pdf/1402.1754>

ITE toolbox: <https://bitbucket.org/szzoli/ite/>

*Kernel methods for big data workshop, Lille, France; Apr. 2, 2014; abstract.

[†]Joint work with Arthur Gretton (Gatsby Computational Neuroscience Unit, University College London), Barnabás Póczos (Machine Learning Department, Carnegie Mellon University), Bharath K. Sriperumbudur (Department of Pure Mathematics and Mathematical Statistics, University of Cambridge); the ordering of the second through fourth authors is alphabetical.