Topology Learning of High Dimensional Probability Density Functions

Yuan Yao
Department of Mathematics
Stanford University
Email: yuany@math.stanford.edu

Abstract

We present some technique on learning from random examples the topology of level sets of a probability distribution or density function, possibly in high dimensional spaces. Given a set of random samples of high dimension, due to the curse of dimensionality it is generally a difficult task to estimate the underlying probability distribution or density, however it may be relatively easier to estimate the topology of density level sets, such as the number of local maxima, the way of connectivity, etc. With this motivation, we utilize the Equi-Energy sampler to sample the density level sets uniformly and develop new tools based on computational topology to analyze density level sets, which leads to a cellular decomposition of the pre-image of probability density functions, in a spirit of Morse theory for manifolds. The new tools are certain generalizations of cluster trees in statistics to higher dimensional complexes.

This work is in collaborationwith Gunnar Carlsson (Math, Stanford), Wing Wong (Stat, Stanford), and Qing Zhou (Stat, UCLA).