



Estimating modes with k -NN graphs

Samory Kpotufe

Princeton University

samory@princeton.edu

Estimating the mode(s) of an unknown density from its sample is a basic problem in data analysis. Mode estimation is relevant to other problems such as clustering and outlier detection. It is also important in its own right as a measure of central tendency. Theoretical work on mode-estimation has concentrated largely on understanding its statistical difficulty. Less attention has been given to implementable procedures.

I will present two related contributions of independent interest:

- (1) Implementable mode estimators – based on subgraphs of a k -Nearest Neighbors (k -NN) graph – which attain minimax-optimal rates under surprisingly general distributional conditions.
- (2) High-probability finite sample rates for k -NN density estimation.

Finally, I will discuss ongoing work towards the deployment of these mode estimators for real-world clustering applications.

Much of the talk is based on a series of works with collaborators S. Dasgupta, K. Chaudhuri, and U. von Luxburg.

Biography

Samory Kpotufe is an Assistant Professor at *ORFE, Princeton University*. In the recent past he was an Assistant Research Professor at the *Toyota Technological Institute Chicago*. Prior to this, he was a researcher at the *Max Planck Institute for Intelligent Systems (MPI)*. At the MPI he worked in the department of Bernhard Schoelkopf and in the learning theory group of Ulrike von Luxburg. He obtained his PhD (Sept 2010) in Computer Science at the University of California, San Diego, advised by Sanjoy Dasgupta.

His research interests lie at the intersection of Machine Learning and more traditional areas of Statistics such as Nonparametric and High-Dimensional Statistics.