



THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

JOINT SEMINAR WITH THE DEPARTMENT OF COMPUTER SCIENCE

SAMORY KPOTUFE

Department of Statistics
Columbia University

From Theory to Clustering

WEDNESDAY, February 17, 2021 at 4:00 PM
Via Zoom (session information will be e-mailed to subscribers)

ABSTRACT

Clustering is a basic problem in data analysis, consisting of partitioning data into meaningful groups called clusters. Practical clustering procedures tend to meet two criteria: flexibility in the shapes and number of clusters estimated, and efficient processing. While many practical procedures might meet either of these criteria in different applications, general guarantees often only hold for theoretical procedures that are hard if not impossible to implement. A main aim is to address this gap.

We will discuss two recent approaches that compete with state-of-the-art procedures, while at the same time relying on rigorous analysis of clustering. The first approach fits within the framework of density-based clustering, a family of flexible clustering approaches. It builds primarily on theoretical insights on nearest-neighbor graphs, a geometric data structure shown to encode local information on the data density. The second approach speeds up kernel k-means, a popular Hilbert space embedding and clustering method. This more efficient approach relies on a new interpretation - and alternative use - of kernel-sketching as a geometry-preserving random projection in Hilbert space.

Finally, we will present recent experimental results combining the benefits of both approaches in the IoT application domain.

The talk is based on various works with collaborators Sanjoy Dasgupta, Kamalika Chaudhuri, Ulrike von Luxburg, Heinrich Jiang, Bharath Sriperumbudur, Kun Yang, and Nick Feamster.

For further information and inquiries about building access for persons with disabilities, please contact Jonathan Rodriguez at 773.702.8333 or send him an email at jgrodriquez@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following website:
<https://lists.uchicago.edu/web/subscribe/statseminars>.