THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

BOAZ NADLER

Department of Computer Science and Applied Mathematics
Weizmann Institute of Science

Unsupervised Ensemble Learning

MONDAY, October 22, 2018 at 4:30 PM
Eckhart 133, 5734 S. University Avenue
Refreshments before the seminar at 4:00PM in Jones 111

ABSTRACT

In various applications, crowdsourcing in particular, one is given the advice or predictions of several classifiers of unknown reliability, over multiple questions or queries. This scenario is different from standard supervised learning where classifier accuracy can be assessed from available labeled training or validation data, and raises several questions: given only the predictions of several classifiers of unknown accuracies, over a large set of unlabeled test data, is it possible to a) reliably rank them, and b) construct a meta-classifier more accurate than any individual classifier in the ensemble? In this talk we'll show that under various independence assumptions between classifier errors, this high dimensional data hides simple low dimensional structures. Exploiting these, we will present simple spectral methods to address the above questions, and derive new unsupervised spectral meta-learners. We'll prove these methods are asymptotically consistent when the model assumptions hold, and present their empirical success on a variety of unsupervised learning problems.

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 jgrodriguez@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following website: https://lists.uchicago.edu/web/subscribe/statseminars.