THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

TRACY KE

UNIVERSITY OF CHICAGO

SPECTRAL APPROACH TO TOPIC LEARNING

MONDAY, October 9, 2017 at 4:30 PM

Eckhart 133, 5734 S. University Avenue

ABSTRACT

In the probabilistic topic models, the quantity of interest—a low-rank matrix consisting of topic vectors—is hidden in the text corpus matrix, masked by noise, and Singular Value Decomposition (SVD) is a potentially useful tool for learning such a low-rank matrix. However, the connection between this low-rank matrix and the singular vectors of the text corpus matrix are usually complicated and hard to spell out, so how to use SVD for learning topic models faces challenges. We overcome the challenge by revealing a surprising insight: there is a low-dimensional simplex structure which can be viewed as a bridge between the low-rank matrix of interest and the SVD of the text corpus matrix, and which allows us to conveniently reconstruct the former using the latter. Such an insight motivates a new SVD-based approach to learning topic models. For asymptotic analysis, we show that under a popular topic model (Hofmann, 1999), the convergence rate of the l1-error of our method matches that of the minimax lower bound, up to a multi-logarithmic term. In showing these results, we have derived new element-wise bounds on the singular vectors and several large deviation bounds for weakly dependent multinomial data. Our results on the convergence rate and asymptotical minimaxity are new. We have applied our method to two data sets, Associated Process (AP) and Statistics Literature Abstract (SLA), with encouraging results. In particular, there is a clear simplex structure associated with the SVD of the data matrices, which largely validates our discovery.

Joint work with Minzhe Wang

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.