Fall 2017 Mathematical Sciences Colloquium

"Three Applications of Randomized Numerical Linear Algebra in Data Analytics "

This talk presents recent results on the advantages and limitations of randomized numerical linear algebra in three data analysis applications. First we characterize the statistical and optimization tradeoffs in using randomized sketching to approximately solve the matrix ridge regression (MRR) problem; we show that in the distributed setting, model averaging combined with sketching obtains near-optimal solutions to the MRR problem while mitigating the statistical risk incurred by sketching. Second, we introduce a novel class of randomized low-rank approximations that can be used to obtain approximate non-linear k-means solutions with (1+epsilon) approximation ratio. We argue that k-means clustering with these randomized approximations is a more theoretically sound---and in practice more effective---approach to non-linear clustering than randomized spectral clustering. Finally, we establish a rate of convergence for randomized Gauss-Siedel that captures the fact that randomized partitioning can outperform a fixed partition scheme when the submatrices are well-conditioned. We provide experimental evidence that randomized Gauss-Siedel outperforms (even accelerated) fixed-partition Gauss-Siedel on large-scale machine learning tasks.

Speaker: Alex Gittens

(Rensselaer Polytechnic Institute)

Monday, November 13, 2017

Time: 4:00 – 5:00 PM

Location: AE214

4:00-5:00pm

Host: Rongjie Lai