

## SEMINAR

### *Theory for New Machine Learning Problems and Applications*

Yingyu Liang  
Princeton University

**Abstract:** Machine learning has recently achieved great empirical success. This comes along with new challenges, such as sophisticated models that lack rigorous analysis, simple algorithms with practical success on hard optimization problems, and handling large scale datasets under resource constraints. In this talk, I will present some of my work in addressing such challenges.

This first part of the talk focuses on learning semantic representations for text data. Recent advances in natural language processing build upon the approach of embedding words as low dimensional vectors. The fundamental observation that empirically justifies this approach is that these vectors can capture semantic relations. A probabilistic model for generating text is proposed to mathematically explain this observation and existing popular embedding algorithms. It also reveals surprising connections to classical notions such as Pointwise Mutual Information, and allows to design novel, simple, and practical algorithms for applications such as sentence embedding.

In the second part, I will describe my work on distributed unsupervised learning over large-scale data distributed over different locations. For the prototypical tasks clustering, Principal Component Analysis (PCA), and kernel PCA, I will present algorithms that have provable guarantees on solution quality, communication cost nearly optimal in key parameters, and strong empirical performance.

Bio: Yingyu Liang is an associate research scholar in the Computer Science Department at Princeton University. His research interests are providing rigorous analysis for machine learning models and designing efficient algorithms for applications. He received a B.S. in 2008 and an M.S. in 2010 in Computer Science from Tsinghua University, and a Ph.D. degree in Computer Science from Georgia Institute of Technology in 2014. He was a postdoctoral researcher in 2014-2016 in the Computer Science Department at Princeton University.

Thursday, March 30, 2017, 4:00 pm  
Mathematics and Science Center: W201

MATHEMATICS AND COMPUTER SCIENCE  
EMORY UNIVERSITY