New Directions in Privacy-Preserving Data Analysis

Kamalika Chaudhuri (UCSD)

Machine learning algorithms increasingly work with sensitive information on individuals, and hence the problem of privacy-preserving data analysis -- how to design data analysis algorithms that operate on the sensitive data of individuals while still guaranteeing the privacy of individuals in the data-- has achieved great practical importance. In this talk, we address two new problems we have been looking at in privacy-preserving data analysis.
First, we address the question of learning from sensitive correlated data, such as private information on users connected together in a social network, and measurements of physical activity of a single user across time. Unfortunately differential privacy cannot adequately address privacy challenges in this kind of data, and as such, these challenges have been largely ignored by existing literature. We consider a recent generalization of differential privacy, called Pufferfish, that can be used to address privacy in correlated data, and present new privacy mechanisms in this framework.
Second, while currently there exist a number of differentially private algorithms for frequentist machine learning, privacy-preserving Bayesian data analysis is not as well-understood. If time permits, I will talk about privacy-preserving Bayesian posterior sampling from exponential families, and show some new experiments on Wikileaks war logs that illustrate how they can be used to learn parameters of relatively complex graphical models.

Based on joint work with Yizhen Wang (UCSD), Shuang Song (UCSD), James Foulds (UCSD), Joseph Geumlek (UCSD), and Max Welling (Univ. of Amsterdam).