Dissertation Defense

Differentially private data release and learning mechanisms

Haoran Li
Emory University

Abstract: Nowadays data sharing is important for application domain, such as scientific discoveries, business strategies, commercial interests, and social goods, especially when there are not enough local samples to test a hypothesis. However, data in its raw format are sensitive as they essentially contains individual specific information, and publishing such data without proper protection may disclose personal privacy. Netflix canceled their recommendation system contest because the released customers data can identify special individuals with high probability. In order to promote data sharing, it is important to develop privacy-preserving algorithms that respect data confidentiality while present data utility. In this dissertation, we address the privacy concerns in publishing high-dimensional data and dynamic datasets, releasing support vector machine classification model, and optimizing utility on data with records of various privacy preferences. It can be shown that all of our privacy preserving algorithms satisfy a rigorous privacy guarantee known as differential privacy, which has been the de facto standard for privacy protection. Extensive empirical studies confirm that they will enable privacy-preserving data release and analytical tasks in a broad range of application domains.

Thursday, November 17, 2016, 1:45 pm
Mathematics and Science Center: E408

Advisor: Li Xiong

Mathematics and Computer Science
Emory University