THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

XIAOJIN (JERRY) ZHU

Department of Computer Science

University of Wisconsin-Madison

Machine Teaching: Frenemy of Machine Learning

MONDAY, April 28, 2014 at 4:00 PM

133 Eckhart Hall, 5734 S. University Avenue

Refreshments following the seminar in Eckhart 110

ABSTRACT

Consider the inverse problem of machine learning: a teacher knows a learner's learning algorithm and wants to construct the smallest (non-iid) training set to guide the learner to a specific target model. This problem, which we call machine teaching, is about designing the optimal "lesson" to maximally influence the learner. One application of machine teaching is in the security of machine learning systems that accept online training data. Here, the teacher is an attacker who can manipulate the training data. The computational problem is for the attacker to identify the minimum-cost manipulation so that the machine learner will be misled to a model that is beneficial to the attacker. A more friendly application of machine teaching is in education, where the teacher wishes to design the best lesson for a human student. Note that in both applications the teacher may only interact with the learner via the training data. I will introduce an optimization-based framework for machine teaching, balancing the goals of "teaching well" and "minimizing teaching effort." For certain learners, machine teaching has a closed-form solution. But in general the optimization problem is combinatorial. I will discuss two approximate solution techniques based on conjugate duality and submodularity. I will also discuss the relation between machine teaching, active learning, and teaching dimensions. Finally, I demonstrate the application of machine teaching with attacks on several popular machine learning models.

Bio: Xiaojin Zhu is an Associate Professor in the Department of Computer Sciences at the University of Wisconsin-Madison. Dr. Zhu received his B.S. and M.S. degrees in Computer Science from Shanghai Jiao Tong University in 1993 and 1996, respectively, and a Ph.D. degree in Language Technologies from Carnegie Mellon University in 2005. He was a research staff member at IBM China Research Laboratory from 1996 to 1998. Dr. Zhu received the National Science Foundation CAREER Award in 2010, and best paper awards at ICML, ECML/PKDD, and SIGSOFT. His research interest is in machine learning, with applications in natural language processing, cognitive science, and social media analysis.
