Center for Responsible Machine Learning

UC SANTA BARBARA

CRML Distinguished Lecture Series:

Adam Smith

Professor of Computer Science, Boston University

When is Memorization Necessary for Machine Learning?

Friday, April 15, 2022 | 1:00 PM | Webinar / Free
Registration: https://ucsb.zoom.us/webinar/register/WN_KlJ2rxKYRly7Xa8jJnpKjg

Abstract: Modern machine learning models are complex, and frequently encode surprising amounts of information about individual inputs. In extreme cases, complex models appear to memorize entire input examples, including seemingly irrelevant information (exact addresses from text, for example). In this talk, we aim to understand whether this sort of memorization is necessary for accurate learning, and what the implications are for privacy.

We describe two results that explore different aspects of this phenomenon. In the first, from STOC 2021, we give natural prediction problems in which every sufficiently accurate training algorithm must encode, in the prediction model, essentially all the information about a large subset of its training examples. This remains true even when the examples are high-dimensional and have entropy much larger than the sample size, and even when most of that information is ultimately irrelevant to the task at hand. Further, our results do not depend on the training algorithm or the class of models used for learning.

Our second, unpublished result shows how memorization must occur during the training process, even when the final model is succinct and depends only on the underlying distribution. This leads to new lower bounds on the memory size of one-pass streaming algorithms for fitting natural models.

Joint work with Gavin Brown, Mark Bun, Vitaly Feldman, and Kunal Talwar.

Bio: Adam Smith is a professor of computer science at Boston University. From 2007 to 2017, he served on the faculty of the Computer Science and Engineering Department at Penn State. His research interests lie in data privacy and cryptography, and their connections to machine learning, statistics, information theory, and quantum computing. He obtained his Ph.D. from MIT in 2004 and has held postdoc and visiting positions at the Weizmann Institute of Science, UCLA, Boston University and Harvard. He received a Presidential Early Career Award for Scientists and Engineers (PECASE) in 2009; a Theory of Cryptography Test of Time award in 2016; the Eurocrypt 2019 Test of Time award; and the 2017 Gödel Prize. He is a Fellow of the ACM.