abstract

COMPUTER SCIENCE/DISCRETE MATH I
Topic:

Speaker:

Affiliation:

Date:

Time/Room:

Learning theory, and in particular PAC learning, was introduced by Valiant in 1984 and has since become a major area of research in theoretical and applied computer science. One natural question that was posed at the very inception of the field is whether there are classes of functions that are hard to learn. Here it is important to make a distinction between proper and improper learning: in proper learning one is required to output a hypothesis that is comparable to the function being learned (e.g. if we are trying to learn k-DNF then the hypothesis should also be a k-DNF), while in improper learning the hypothesis can be more complex (e.g. if we are trying to learn k-DNF then the hypothesis could be a circuit of size n^c).

Computational limitations to proper and improper learning have been extensively studied, starting with the seminal works of Pitt-Valiant (JACM '88) and Kearns-Valiant (JACM '94). However, while the hardness of proper learning is typically based on worst case assumptions on the power of NP (e.g., SAT \notin BPP), all known limitations on improper learning are based on cryptographic assumptions (e.g., the existence of one-way functions). It is natural to ask whether this gap is inherent, specifically: is it possible to base hardness of improper learning on worst case assumptions such as SAT \notin BPP?

We show that, unless the Polynomial Hierarchy collapses, such a statement cannot be proven via a large class of reductions including Karp reductions, truth-table reductions, and a restricted form of non-adaptive Turing reductions. Also, a proof that uses a Turing reduction of constant levels of adaptivity would imply an important consequence in cryptography as it yields a transformation from any average-case hard problem in NP to a (standard) one-way function.

abstract

These results are obtained by showing that lower bounds for improper learning are intimately related to the complexity of zero-knowledge arguments and to the existence of weak cryptographic primitives. In particular, we prove that if a language L reduces to the task of improper learning circuits, then, depending on the type of the reduction in use, either (1) L has a statistical zero-knowledge argument system, or (2) the worst-case hardness of L implies the existence of a weak variant of one-way functions (as defined by Ostrovsky-Wigderson). Interestingly, we observe that the converse implication also holds. Namely, if (1) or (2) hold then the intractability of L implies that improper learning is hard. Our results hold even in the stronger model of agnostic learning.

This is joint work with Boaz Barak and David Xiao.