

## **abstract**

Computer Science/Discrete Mathematics Seminar I  
Topic:

Speaker:

Affiliation:

Date:

Time/Room:

---

With the notion of interaction with oracles as a unifying theme of much of my dissertation work, I discuss novel models and results for property testing and computational learning, with the use of Fourier analytic and probabilistic methods.

One motivation for property testing is that testing can provide a fast preprocessing step before learning. However, algorithms based on membership queries (i.e., the ability to query functions on arbitrary points) tend to query highly ambiguous or unnatural points that can be impossible for a human oracle to label. In this work, we analyze a more realistic model where the algorithm may query for labels only on points in a given (polynomial-sized) unlabeled sample drawn from some underlying distribution. Further, we develop a general notion of the  $\{\text{testing dimension}\}$  of a given property with respect to a given distribution and use that to establish a number of lower bounds for the class of dictator functions, linear separators, etc.

We also study new models for PAC learning the class of DNF formulas. We consider a type of natural Boolean pairwise query that asks whether a pair of positive examples from a polynomial-sized sample satisfy at least one term in common in the target DNF, as well as numerical queries that ask how many terms in common the two examples satisfy. We begin with a somewhat surprising negative result: that learning general DNF formulas under arbitrary distributions from the above Boolean queries is as hard as PAC-learning DNF formulas without them. However, on the positive side we show general DNF can be learned from numerical queries over the uniform distribution, and give positive results for a number of other natural DNF classes. In the process, we give an algorithm for learning a sum of monotone terms from labeled data only.

