

abstract

COMPUTER SCIENCE/DISCRETE MATH SEMINAR, I
Topic:

Speaker:

Affiliation:

Date:

Time/Room:

The problem of average-case learning for decision trees is as follows: a decision tree T (over Boolean features x_1, \dots, x_n) is chosen at random from some natural distribution over decision trees. The learner is given access to uniform random examples from $\{0,1\}^n$ each of which is labeled according to T . The job of the learning algorithm is to efficiently construct a hypothesis function f which closely agrees with the unknown randomly chosen tree T .

Decision trees are an expressive representation class for Boolean functions which have proven hard to learn in worst-case models of learning from random examples. An even more challenging problem is that of learning an unknown DNF (Disjunctive Normal Form formula, i.e. an AND of ORs of Boolean variables) or CNF in the worst case.

In this work we give the first efficient algorithms for the average-case versions of these learning problems. Our algorithm for decision trees runs in polynomial time for random trees of logarithmic depth. Our algorithm for DNF and CNF formulas runs in polynomial time for random formulas with a subquadratic number of AND gates.

Joint work with Jeff Jackson.