Mathematical Theories of Interaction with Oracles

Abstract:
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, Computational and Statistical Learning Theory, and Algorithmic Economics.

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. 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 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 two positive examples from a polynomial-sized sample satisfy at least one term in common in the target DNF. We begin with showing learning general DNF formulas under arbitrary distributions from this type of 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 (ask how many terms in common the two examples satisfy) 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.

We study the problem of active learning in a stream-based setting, allowing the distribution of the examples to change over time. We prove upper bounds on the number of prediction mistakes and the number of label requests for established disagreement-based active learning algorithms, both in the realizable case and under Tsybakov noise. We further prove minimax lower bounds for this problem. We also explore a transfer learning setting, in which a finite sequence of target concepts are sampled independently with an unknown distribution from a known family. We study the total number of labeled examples required to learn all targets to an arbitrary specified expected accuracy, focusing on the asymptotics in the number of tasks and the desired accuracy. We study in detail the benefits of transfer for self-verifying active learning; in this setting, we find that the number of labeled examples required for learning with transfer is often significantly smaller than that required for learning each target independently.

I discuss connections and applications of Machine Learning Theory to problems in Algorithmic Economics. We consider the problem of online pricing and allocating multiple goods of economics of scale or decreasing marginal costs to the seller. In particular, we analyze the case where buyers have unit-demand and arrive one at a time with valuations on items, sampled iid from some unknown underlying distribution over valuations. Our strategy operates by using initial buyers to learn enough about the distribution to determine how best to allocate to the future buyers. We show, for instance, if the goal of the seller is to give each buyer a toy of value 1, but in such a way that minimizes the total cost to the seller and if buyers have binary valuations, we can efficiently perform this allocation task with cost at most a constant factor greater than that of the optimal allocation of items in hindsight, so long as the marginal costs do not decrease too rapidly. We also study factor models for correlated auctions. The auctioneer wants to maximize his revenue. He does not know the exact valuations of the buyers. What he does know is a distribution from which the vector of valuations is drawn. For a special type of correlated distributions where the correlation between buyer valuations is driven by common factors, we construct an auction that is dominant strategy truthful, ex-post individually rational, and asymptotically optimal.

I discuss connections and applications of Machine Learning Theory to problems in Algorithmic Economics. We consider the problem of online pricing and allocating multiple goods of economics of scale or decreasing marginal costs to the seller. In particular, we analyze the case where buyers have unit-demand and arrive one at a time with valuations on items, sampled iid from some unknown underlying distribution over valuations. Our strategy operates by using initial buyers to learn enough about the distribution to determine how best to allocate to the future buyers. We show, for instance, if the goal of the seller is to give each buyer a toy of value 1, but in such a way that minimizes the total cost to the seller and if buyers have binary valuations, we can efficiently perform this allocation task with cost at most a constant factor greater than that of the optimal allocation of items in hindsight, so long as the marginal costs do not decrease too rapidly. We also study factor models for correlated auctions. The auctioneer wants to maximize his revenue. He does not know the exact valuations of the buyers. What he does know is a distribution from which the vector of valuations is drawn. For a special type of correlated distributions where the correlation between buyer valuations is driven by common factors, we construct an auction that is dominant strategy truthful, ex-post individually rational, and asymptotically optimal.