Principled Machine Learning for Practical Applications

Claire Monteleoni

Ph.D., MIT

Date: March 3rd, 2008
Time: 3:00 PM
Venue: 256 Coates Hall

Abstract

The rapidly increasing abundance of data generated by internet transactions, satellite measurements, and environmental sensors, among other sources, creates new and urgent challenges for machine learning. Such data can be vast in terms of the number of examples, e.g. image data recorded by satellites, or text data on the internet. It is often extremely high-dimensional, and it can be sparse, meaning that only a much smaller subset of the dimensions actually contains relevant information. It can be noisy, in that measurements may have been corrupted or lost, and thus some labels or features may be incorrect or missing. In some problems, such as forecasting the behavior of the internet, the data also changes over time. The goal of principled machine learning for practical applications is to provide provably cheap, efficient, and generic methods to detect patterns in real-world data. In this talk I will focus on two problems for which we provided principled machine learning algorithms for practical applications: (1) learning with online constraints, and (2) active learning.

(1) Learning with online constraints: Many practical problems such as forecasting, real-time decision making, streaming data applications, and resource-constrained learning, can be modeled as learning with online constraints. In this model, the data arrives in a stream and the learner must make immediate predictions, and is unable to store all the past data. In an online learning framework in which no statistical assumptions are made about the sequence of observations, algorithms can be evaluated based on their regret, i.e. their relative prediction loss with respect to the hindsight-optimal algorithm in a comparator class. I will present an algorithm that avoids the lower bound on regret that we demonstrated for previous algorithms in this framework, and yields better performance guarantees: regret is logarithmic, as opposed to linear, in the number of predictions. Our algorithm simultaneously learns the level of non-stationarity in the observation sequence, while performing the learning task. We applied our algorithm to real-time energy management in wireless networks, for mobile nodes using the 802.11 protocol, yielding encouraging resultsin simulation.

(2) Active learning: In many data-rich problems, unlabeled data is much easier to obtain than labeled data (e.g. images on the web, or unlabeled video signal). An active learner receives unlabeled data and can make intelligent choices as to which labels to obtain via

some mechanism (e.g. paying for annotation by a human). I will present an algorithm that, for uniformly distributed, linearly separable data, needs exponentially fewer labels than the analogous sample complexity of supervised learning. Our algorithm is extremely light-weight and easy to implement, and is in fact well-suited to active learning with online constraints. We demonstrate the empirical performance of online active learning algorithms on an optical character recognition task. Recently we worked on generalizing active learning to arbitrary distributions and hypothesis classes, and provided an algorithm for general active learning via reduction, which is as efficient (up to polylogarithmic factors) as the algorithm for the supervised version.

I will conclude with comments on my ongoing and future work on privacy-preserving machine learning.