A tutorial on spectral and predictive state learning*

Geoff Gordon

Abstract

A common problem in machine learning is to find a compact model that explains some high-dimensional, structured observations. For example, we might want to explain the sequence of sensor readings from a mobile robot, or the parse tree of some natural-language text, or other types of data such as images, video, biological sequences, or social interactions. An appealing way to design such a model is to suppose that a small set of underlying latent variables explains the dependences among our observations. This intuition leads to popular models such as latent Dirichlet allocation, Gaussian mixtures, latent tree graphical models, and hidden Markov models. Unfortunately, fitting these models is tricky: to find the right latent state representation and model parameters, we must solve difficult temporal and structural credit assignment problems, leading to a search space with a host of bad local optima. Often, researchers have attempted to fit these models via local exploration heuristics such as expectation-maximization (EM) or Gibbs sampling, with mixed success. In this talk, I'll describe a qualitatively different class of algorithms, called spectral, predictive state, or method-of-moments estimators. These methods are computationally efficient, statistically consistent, and have no local optima; in addition, they can be simple to implement, and have state-of-the-art practical performance for many interesting learning problems.

*Machine Learning External Seminar, Gatsby Unit, July 16, 2014.