Computer Science > Learning

Spectral Methods for Learning Multivariate Latent Tree Structure

Animashree Anandkumar, Kamalika Chaudhuri, Daniel Hsu, Sham M. Kakade, Le Song, Tong Zhang

(Submitted on 7 Jul 2011 (v1), last revised 8 Nov 2011 (this version, v2))

This work considers the problem of learning the structure of multivariate linear tree models, which include a variety of directed tree graphical models with continuous, discrete, and mixed latent variables such as linear-Gaussian models, hidden Markov models, Gaussian mixture models, and Markov evolutionary trees. The setting is one where we only have samples from certain observed variables in the tree, and our goal is to estimate the tree structure (i.e., the graph of how the underlying hidden variables are connected to each other and to the observed variables). We propose the Spectral Recursive Grouping algorithm, an efficient and simple bottom-up procedure for recovering the tree structure from independent samples of the observed variables. Our finite sample size bounds for exact recovery of the tree structure reveal certain natural dependencies on underlying statistical and structural properties of the underlying joint distribution. Furthermore, our sample complexity guarantees have no explicit dependence on the dimensionality of the observed variables, making the algorithm applicable to many high-dimensional settings. At the heart of our algorithm is a spectral quartet test for determining the relative topology of a quartet of variables from second-order statistics.

Subjects: **Learning (cs.LG)**; Machine Learning (stat.ML)

Cite as: **arXiv:1107.1283 [cs.LG]**

(or **arXiv:1107.1283v2 [cs.LG]** for this version)

Submission history

From: Daniel Hsu [view email]

[v1] Thu, 7 Jul 2011 02:33:31 GMT (367kb,D)

[v2] Tue, 8 Nov 2011 15:42:32 GMT (172kb,D)