Learning optimal independence tests for Bayesian Networks

Anonymous Author(s)
Affiliation
Address
email

Abstract

1 Introduction

Learning Bayesian networks for general distributions is intractable task Chickering [1996]. However, in practise for distributions appearing in real data, still we are able to recover Bayesian network structure. This implies that real data distribution has some special properties, which simplify process of structure recovery. This process is almost always based on computation local statistics, and then reasoning about global structure Jaakkola et al. [2010]; Tsamardinos et al. [2006]. Local statistics describe complexity (e.g. number of parameters in case of BIC), and level of independence between nodes (e.g. mutual information tests, conditional independence tests). We focus in this work on improvement of independence tests by learning it for CPDs present in data.

We consider parameterized family of independence tests. Before inferring about structure of Bayesian network, we tune classifier predict independence or dependence on CPDs present in data. This way, we obtain highly sensitive independence test, which fires on dependence / independence phenomenas present in our data.

2 Related work

Independence tests used in Bayesian networks

- Schäfer and Strimmer [2005] Here they learn a Gaussian Graphical Model (GGM) using estimates of the partial correlation matrix.
- Opgen-Rhein and Strimmer [2007] Here they learn an approximate causal structure on gene expression based on full-order partial correlation (as an approximation to lower-order partial correlation that is called for theoretically for a Bayesian network).
- Tsamardinos et al. [2006] MMHC algorithm. They use a test based on what is called the G^2 statistic (asymptotically distributed as chi^2) and they also talk about a couple other independence tests that we may want to look into.

Existing independence tests:

- Pearson's χ-squared. The problem is the null hypothesis is independence, but independence is what we're trying to show.
-

LL (Log-likelihood) (1912-22) MDL/BIC (Minimum description length/Bayesian Information Criterion) (1978) AIC (Akaike Information Criterion) (1974) NML (Normalized Minimum Likelihood) (2008) MIT (Mutual Information Tests) (2006)

Margaritis [2003]

3 Independence testing

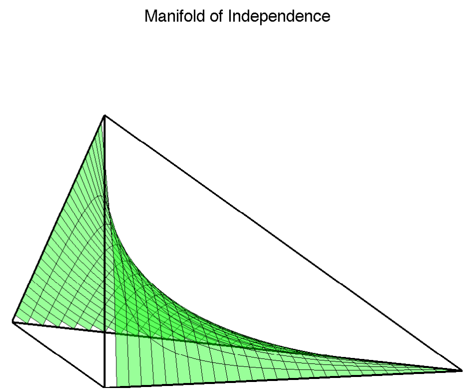

Figure 1: A

3.1 Discrete variables formulation

3.2 Continuous variables formulation

4 Experiments

4.1 Classification of Synthetic CPDs

4.2 Classification of CPDs from Gene Expression

4.3 Synthetic Bayesian networks

4.4 Gene expression data

5 Discussion

References

D. M. Chickering. Learning bayesian networks is np-complete. In *Learning from data*, pages 121–130. Springer, 1996.

T. Jaakkola, D. Sontag, A. Globerson, and M. Meila. Learning bayesian network structure using lp relaxations. In *International Conference on Artificial Intelligence and Statistics*, pages 358–365, 2010.

D. Margaritis. *Learning Bayesian network model structure from data*. PhD thesis, University of Pittsburgh, 2003.

R. Opgen-Rhein and K. Strimmer. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC systems biology*, 1(1): 37, 2007.

J. Schäfer and K. Strimmer. An empirical bayes approach to inferring large-scale gene association networks. *Bioinformatics*, 21(6):754–764, 2005.

I. Tsamardinos, L. E. Brown, and C. F. Aliferis. The max-min hill-climbing bayesian network structure learning algorithm. *Machine learning*, 65(1):31–78, 2006.