OUTLINE OF PAPER

ABSTRACT Other people use discretization or linear assumptions, we do nonparametric continuous BN. It works well.

INTRO

* Bayesian network definition.
* Bayes nets are cool because \_\_\_\_\_\_.
* Sometimes people construct Bayes nets by hand, but in other cases we don’t know the relationships, hence instead we try to learn them from data. Formally, the problem is \_\_\_\_\_.
* BN structure learning is hard, but people do it anyways.
* One important field which motivates this work is computational biology. Here often the relationships between variables is not known, or at least controversial, and hence structure learning can prove to be very useful in answering these questions, that is, if one can build a reliable model. However, often the datasets generated are quite small for structure learning standards, adding another layer of difficulty to the problem.
* Often in this field, the measured variables are continuous, e.g. gene expression or proteomics data. However, when learning a Bayes net to model such data, researchers often discretize very coarsely, e.g. three bins denoting high, low, and mid-range values, or use a linear Gaussian model, e.g. the mean of each variable is modeled as a linear combination of its parents plus Gaussian noise. The discrete model of course throws out a lot of valuable information. The continuous, linear Gaussian model makes a strong assumption about the data, and also tends to be a difficult model for which to learn the structure, i.e. v-structures can be difficult to identify.
* BN Structure learning approached can be roughly divided into two categories- constraint-based, and score-and-search based. Our approach can be considered a hybrid of the two, incorporating the results of conditional independence tests in order to construct a more powerful scoring function.
* Recently, an ILP approach was presented, which means that now we can hope to learn structures which optimize a score for networks with some maximum in-degree.

RELATED WORK

Modeling Continuous Bayes Nets

{already done, should edit}

Scoring Functions

Talk about standard- BIC, BDe. Mention MIT score. Mention thesis incorporating constraints into scoring function.

BACKGROUND

Kernel ridge regression

Partial correlation

Kernelized Conditional Independence

KERNEL SPARSITY BOOST (KSB): A NEW SCORING FUNCTION

Overview:

Here is our score. The first term is the likelihood conditioned on w. The second term follows the SparsityBoost framework (explain that).

Likelihood term

Equivalent to ML under the assumption linear-in-RKHS, plus Gaussian noise.

SparsityBoost term

Explain that we want confidence about independence, have to invert hypothesis test. Describe Efron’s approach. Discuss how we fit distributions. **Present theorem!** Show a plot.

NUMERICAL EXPERIMENTS

1. Synthetic data:
   1. Describe toy networks
   2. Describe data model
   3. ROC curves: KCI compared with other standard BN learning tests (esp G^2 test). Add KCI-Bayes??
   4. Structure learning results.
2. Real data:
   1. Describe three datasets
   2. Show structure learning results (LL-test)
   3. Show network compared with others’ results.

CONCLUSION

KSB rocks. Low statistical complexity. Slow but could be parallelized. Test inversion could be applied in many ways. Other ways to construct sparsity boost term. Can add this to other scores, e.g. for learning Copula BNs. Can compute a BDE-like score, i.e. marginalize over prior over w.