SEP Instructions

Nevena Lazic
nevena.lazic@gmail.com

Structural Expectation Propagation (SEP) [Lazic et al., 2013] is a method for learning the structure of discrete Bayesian networks containing latent variables. This software is an implementation of SEP under the Infer.NET framework [Minka et al., 2010]. It supports learning bipartite networks in which latent variables are parents of observed variables and each observed variable has up to two latent parents, as in Figure 1.

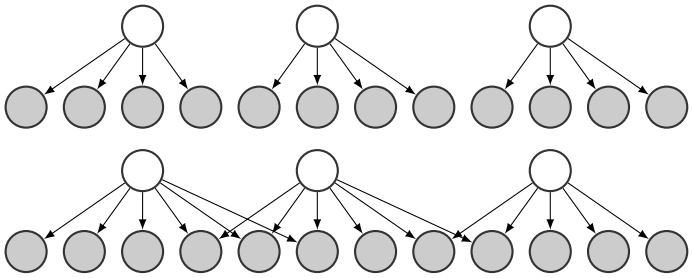

Figure 1: Top: a Bayesian network in which each observed variable (shaded) has a single latent parent (white). Bottom: a Bayesian network in which each observed variable has up to two latent parents.

Commands

`SampleData nPa nX dX nU dU N fname`

`SampleData` randomly generates a bipartite network containing `nX` observed variables of cardinality `dX` and `nU` latent variables of cardinality `dU`. Each observed variable has maximum degree `nPa`. Network parameters are set according to [Chickering and Meek, 2002].
`fnameG.txt` lists the latent parents of each observed variable, one variable per line. `fnameX.txt` contains `N` data points sampled from the network, one per line. `SampleData` also writes the data and network constraints in the format required by an implementation of the Structural EM algorithm [Friedman, 1998].

`ModelEvidence datafile modelfile nU dU`

`ModelEvidence` calculates EP model evidence for observations in `datafile`, given a bipartite network with `nU` latent variables of cardinality `dU` and structure given in `modelfile`. `datafile`should contain comma-delimited variables for each observation, and one observation per line. `modelfile` should list comma-delimited indices of the latent parents of each observed variable, one variable per line.

```
SEP1 datafile outname nU dU
SEP2 datafile outname nU dU
SEP2batch datafile outname nU dU K I
```

`SEP*` commands learn the structure of discrete bipartite networks containing `nU` latent variables of cardinality `dU`, based on data in `datafile.` Each observed variable has one latent parent (`SEP1`) or up to two latent parents (`SEP2`). `datafile` should contain comma-delimited variables and one observation per line. For large datasets, `SEP2batch` splits the data into `K` batches and cycles over them `I` times, producing output at each iteration.
`outname_qG.txt` contains the posterior distribution over network edges.
`outname_qGmap.txt` contains the MAP network.
`outname_qUk.txt` contains posteriors of latent variables U_k^n, one per line.
`outname_qT.txt` contains the posteriors (pseudocounts) of Dirichlet parameters `Ti_j` for the conditional probability of an observed variable X_i given that its parent takes on the value j.

References

[Chickering and Meek, 2002] Chickering, D. and Meek, C. (2002). Finding optimal bayesian networks. In *Proc. 18th Conference on Uncertainty in Artificial Intelligence (UAI)*.

[Friedman, 1998] Friedman, N. (1998). The bayesian structurall em algorithm. In *Proc. 14th Conference on Uncertainty in Artificial Intelligence (UAI)*.

[Lazic et al., 2013] Lazic, N., Bishop, C. M., and Winn, J. (2013). Structural expectation propagation: Bayesian structure learning for networks with latent variables. In *Proc. 16th Conference on Artificial Intelligence and Statistics (AISTATS)*.

[Minka et al., 2010] Minka, T., Winn, J., Guiver, J., and Knowles, D. (2010). Infer.NET 2.4. Microsoft Research Cambridge. http://research.microsoft.com/infernet.