Latent Variable Selection with Convex Optimization

Tina Behrouzi, Yuan Liu

ABSTRACT

todo
include motivation, challenge, contributions

In this project, we aim to explore how to choose good topology for graphical models using convex optimization techniques. Performance of graphical models used in methods such as Bayesian inference is dependent largely on the structure of the model and parameters such as latent variables, thus it is important to pick a good model topology. However, typical realization of graphical models uses some fixed prior or greedy search which are suboptimal and/or non-convergent in general. In several literature (cite sources), a different approach uses sparsity of graphical models and regularization terms in order to generate graph decomposition using convex invariants of graphs. Such general approach of sparsity and dimensionality reduction for selection of graphical model topology is our topic of interest.

PRELIMINARY SURVEY

Graph modeling is used in communication models, stock market, bio information, etc. Finding the structure of latent variables subject to observed ones has been the interest of many studies. The challenge faced when solving this optimization problem on recent data, which usually has high dimensionality and size. Previously, the Expectation-Maximization (EM) algorithm was widely used to learn latent variables of tree-like graphs (cite). However, it has a very slow convergence and poor local optima.

In (cite), it is considered that latent and observed variables are jointly Gaussian, and the estimation of the covariance matrix between nodes is considered as the representation of the model. The assumption of the sparsity of the matrices has an important role in this method. L1 and nuclear norm regularizers are considered to find the best covariance matrix representation of the latent variable. Some papers have used the Alternating Direction Method of Multipliers (ADMM) to alter the method to the more class scaleable problem (cite).

OTHER MAIN SECTIONS (TBD)

todo..

Derivations:

Likelihood, assuming normal distribution:

$L(\theta; x) = p_\theta(X = x|\theta)$

$$L(\theta; x) = \frac{1}{(2\pi)^{\frac{D}{2}} (det\Sigma)^{\frac{1}{2}}} exp(-\frac{1}{2}(x - \mu)^T\Sigma^{-1}(x - \mu))$$

$\theta = (\mu, \Sigma)$

Log likelihood:

$$l(\theta; x) = logL(\theta; x)$$
$$= -log((\frac{1}{2\pi})^{\frac{D}{2}}) - log(det\Sigma)^{\frac{1}{2}}$$
$$- \frac{1}{2}(x - \mu)^T\Sigma^{-1}(x - \mu)$$

Optimize over θ:

$$\max_\theta \ logL(\theta; x)$$
$$\max_\theta \ -\frac{1}{2}log(det\Sigma) - \frac{1}{2}(x - \mu)^T\Sigma^{-1}(x - \mu)$$

Optimal solution:

$$\underset{\theta}{argmax} \ -log(det\Sigma) - (x - \mu)^T\Sigma^{-1}(x - \mu)$$

$let \ z = x - \mu$

$let \ \Sigma_s = zz^T$

$let \ K = \Sigma^{-1}$

$(x - \mu)^T\Sigma^{-1}(x - \mu) = tr(z^T\Sigma z) = tr(zz^T\Sigma^{-1}) = tr(\Sigma_s K)$

$$\underset{\theta}{argmax} \ log(detK) - tr(\Sigma_s K)$$

For K positive definite, the above problem can be cast as a convex optimzation.

$let \ K = S - L$

$let \ S = X_O \ given \ X_H$

$let \ L \ be \ marginalization \ over \ X_H$

$$\max_{S,L} \ l(\theta = S - L; \Sigma_s) = \max_{S,L} logdet(S - L) - tr(\Sigma_s(S - L))$$

$s.t. \ S - L \succ 0, L \succeq 0$

Adding a regularization term (from main paper):

$$\max_{S,L} l(\theta = S - L; \Sigma_s) + \lambda(\gamma\|S\|_1 + tr(L))$$
$$\max_{S,L} logdet(S - L) - tr(\Sigma_s(S - L))$$
$$+ \lambda(\gamma\|S\|_1 + tr(L))$$
$$s.t. \ S - L \succ 0, L \succeq 0$$

Once the structure is receovered, a model can be trained using one of existing techniques such as loopy belief propagation, mean field, evidence lower bound, etc.

Citation test only [1].

CONCLUSION

todo

REFERENCES

[1] V Chandrasekaran, P.A. Parrilo, and A.S. Willsky. "Latent Variable Graphical Model Selection via Convex Optimization". In: *The Annals of Statistics, Vol. 40, No. 4, 1935-1967* (2012).