High dimensional Sparse Gaussian Graphical Mixture Model

A. Lotsi, E. Wit

University of Groningen
JBI of Mathematics and Computer Science

a.lotsi@rug.nl

February 1, 2013
Networks have taken the science by storm

System of interest
Graph representing the system

- Networks reconstruction → attractive paradigm of genomic science.
  Genomics system → complex.

- Graphical models (GM) are used for networks reconstruction
- Idea: Data is heterogeneous
  Can we find underline graphs → data?
1708 patients score 28 questions

Idea: depression is heterogeneous

Each type of depression is described as a sparse graph of connected answers to the questionnaire.
Graphical model (GM)

- Describe Conditional dependencies between variables
- Graph: $G = (V, E)$
- nodes $V$ are random variables
- links $E$ represent conditional independent relationships

**Graph (a):** X and Z are CI given Y

**Graph (b):** Z is ind. of each of X, Y, and W
Gaussian Graphical Model (GGM)

- r.v. come from **multivariate normal distribution** \( N(\mu, \Sigma) \)
- \( \Theta = \Sigma^{-1} \), \( \Theta \) contains all conditional (in) dependence information
- Pairwise Markov Property:
  \[ X_i \perp X_j | X_{V \setminus \{i,j\}} \iff \theta_{ij} = 0 \]

- Inv Cov. Mat. \( \Theta \) of any MVG is graph structured
Mixture of GGM
Mixture model

Graphical representation is

\[ Z_n : \text{latent variable} \]

\[
f_\Omega(y_i) = \sum_{k=1}^{K} \pi_k \phi_k(y_i; 0, \Theta_k^{-1})
\]

\[ \Omega = \{(\pi_k, \Theta_k)\} \]

Model-based likelihood

\[
l_y(\Omega) = \sum_{i=1}^{n} \log f_\Omega(y_i)
\]

**Goal:** Maximize log-likelihood with respect to \( \Omega \) using MLE approach.

few difficulties .....
Difficulties

**Lindsay[2005]**: Multimodal nature of likelihood. Set of modes not in one-to-one correspondence with distinct components.

A common approach to such data is mixture model because it provides a decomposition of sampled population into a set of homogeneous component in a way that is consistent with the multimodal density configuration

![The Topography of Mixtures](image)

**Solution**: Avoids to stay in local max. by starting with high penalty
**Difficulties**

small $n$ large $p$: $(n \prec p)$, $\Sigma$ is not invertible to yield estimate of $\Theta$

**Solution:** Penalized log-likelihood approach: [Friedman et al 08].

$$l_y^{\text{penalty}}(\Omega) = l_y(\Omega) - \lambda \sum_{k=1}^{K} ||\Theta_k||_1$$

Result of $l_1$ likelihood: Consistent network selection for certain rate of $\lambda$. Buhlmann van der Geer[2011]
Difficulties

trade off between small variance and high likelihood

Solution: Penalize diagonal
The E-step  Data augmentation with a latent $Z \Rightarrow (Y, Z)$

$$Q(\Omega|\Omega^{(t)}) = E_{Z_i} \left[ l_{YZ}(\Omega) - \frac{n\lambda}{2} \|\Theta\|_1 | Y; \Omega^{(t)} \right]$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{n} \left[ \log \phi_k(Y_i; 0, \Theta_k^{-1}) + \log \pi_k \right] \gamma_{ik}^{(t)} - \frac{n\lambda}{2} \|\Theta_k\|_1$$

where $\gamma_{ik}^{(t)} = P(Z_i = k | Y_i, \Omega^{(t)})$

- Calculate $\gamma_{ik}^{(t)}$ via Bayes’ theorem

$\Rightarrow$ naive Bayes classification
The M-step

\[ \pi_k^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} \gamma_i(t) \delta_{ik} \]

Maximization over \( \Theta \)

Formulation of mixture problem similar to GGM

\[ Q(\Theta_k^0) = \sum_{i=1}^{n} \gamma_i(t) \delta_{ik} 2 \log \Theta_{k0}^{ij} - 2 \text{tr} \left( \sum_{i=1}^{n} \gamma_i(t) \delta_{ik} Y_i Y_i' \right) \Theta_{k0}^{ij} + \lambda_{k0}^{ij} \Theta_{k0}^{ij} \]

\[ \hat{\Theta}_{k0} = \arg\max_{\Theta} \{ \log \Theta_{k0}^{ij} - \text{tr} (\tilde{S}_{k0} \Theta_{k0}^{ij}) - \lambda_{k0}^{ij} \Theta_{k0}^{ij} \} \]
Em algorithm

The M-step

- Maximization over $\pi_K$: $\pi_{k_0}^{(t+1)} = \gamma_{k_0}^{(t)}/n$
The M-step

- Maximization over $\pi_K$: $\pi_{k_0}^{(t+1)} = \gamma_{k_0}^{(t)} / n$
- Maximization over $\Theta_K$

Formulation of mixture problem similar to GGM
**The M-step**

- Maximization over $\pi_K$: $\pi_{k_0}^{(t+1)} = \gamma_{.k_0}^{(t)}/n$
- Maximization over $\Theta_K$

**Formulation of mixture problem similar to GGM**

$$Q(\Theta_{k_0}) = - \sum_{i=1}^{n} \frac{\gamma_{ik_0}^{(t)}}{2} \log |\Theta_{k_0}| - \frac{1}{2} tr \left( \frac{\sum_{i=1}^{n} \gamma_{ik_0}^{(t)} (Y_i Y'_i)}{\gamma_{ik_0}^{(t)}} \Theta_{k_0} \right) - \frac{n \lambda}{2} \| \Theta_{k_0} \|$$
The **M-step**

- Maximization over $\pi_K$: $\pi_{k_0}^{(t+1)} = \gamma_{.k_0}^{(t)}/n$
- Maximization over $\Theta_K$

**Formulation of mixture problem similar to GGM**

$$Q(\Theta_{k_0}) = - \sum_{i=1}^{n} \frac{\gamma_{ik_0}^{(t)}}{2} \log |\Theta_{k_0}| - \frac{1}{2} tr \left( \sum_{i=1}^{n} \frac{\gamma_{ik_0}^{(t)} (Y_i Y_i')}{\gamma_{ik_0}^{(t)}} \Theta_{k_0} \right) - \frac{n \lambda}{2} \|\Theta_{k_0}\|_1$$

$$= \frac{\gamma_{.k_0}^{(t)}}{2} \left[ \log |\Theta_{k_0}| - tr (\hat{S}_{k_0} \Theta_{k_0}) - \hat{\lambda}_{k_0} \|\Theta_{k_0}\|_1 \right]$$
**The M-step**

- Maximization over $\pi_K$: $\pi_{k_0}^{(t+1)} = \gamma_{k_0}^{(t)}/n$
- Maximization over $\Theta_K$

**Formulation of mixture problem similar to GGM**

\[
Q(\Theta_{k_0}) = -\sum_{i=1}^{n} \frac{\gamma_{ik_0}^{(t)}}{2} \log |\Theta_{k_0}| - \frac{1}{2} \text{tr} \left( \sum_{i=1}^{n} \frac{\gamma_{ik_0}^{(t)} (Y_i Y_i')}{\gamma_{ik_0}^{(t)}} \Theta_{k_0} \right) - \frac{n \lambda}{2} \|\Theta_{k_0}\|_1
\]

\[
= \frac{\gamma_{k_0}^{(t)}}{2} \left[ \log |\Theta_{k_0}| - \text{tr} (\tilde{S}_{k_0} \Theta_{k_0}) - \tilde{\lambda}_{k_0} \|\Theta_{k_0}\|_1 \right]
\]

$\hat{\Theta}_{k_0} = \arg\max_{\Theta} \left\{ \log |\Theta_{k_0}| - \text{tr}(\tilde{S}_{k_0} \Theta_{k_0}) - \tilde{\lambda}_{k_0} \|\Theta_{k_0}\|_1 \right\}$
Model selection

- Ruan [2011], Pan, Xiaotong Shen [2007] use BIC.
- Akshay Krishnamurthy. [2011] combined AIC/ CV
- We propose:
  - Initial search: $0 < \lambda < \max(S_{ij})_{i \neq j}$
  - Use **EBIC** of Chen et al. [2008]

  $$EBIC_\gamma(K, \lambda) = -2l_Y(\hat{\Theta}(K, \lambda)) + P(K, \lambda)\log(n) + 4P(K, \lambda)\gamma\log(p)$$

  $$(\hat{K}, \hat{\lambda}) = \arg\min_{K, \lambda}(EBIC_\gamma(K, \lambda))$$

- Takes into consideration $p$. 

Simulation study

Simulate 2 gaussian distributions from true precision matrices with \( \pi = 0.5 \)

\[
\begin{array}{cccc}
100 & 0.17 & 0.82 & 2 \\
200 & 0.08 & 0.91 & 2 \\
500 & 0.5 & 0.49 & 2 \\
1000 & 0.5 & 0.49 & 2 \\
\end{array}
\]

Table: Estimates of mixing coefficients \( \hat{\pi}_K \), the mixture component \( \hat{K} \)

\[
\begin{array}{ccc}
 & 1 & 2 \\
0 & 410 & 110 \\
1 & 110 & 310 \\
\end{array}
\]

Table: Misclassification rate at \( n = 1000 \)
• Recovered 2 main components
• Underline networks

Cluster 1

Cluster 2
- Q2 influences negatively Q10 in cluster 1

- Positive influence between Q1 and Q4 in Cluster 1

- Negative interactions between Q4 and Q22 in cluster 1 BUT positive interaction in cluster 2
conclusions

- Developed GGMM algorithm with $l_1$ penalty.
- Simultaneously estimate hyper-parameters through EBIC.
- Simulation shows that algorithm is suitable for HDD from MM.
- Depression data reveals 2 components with underlying networks.
- Determining optimum $K$ is still a major contemporary issue.
- glasso.mix package will be available.
The End