An Introduction of Some Methods in Learning Gene Association Networks

Bo Li
Overview

- Motivation
- Naïve Method
- Graphical Lasso
Motivation

- Genes are never isolated
- Want to study them in a systematic way
- A good way to look at their interactions: build gene association networks
Data

- Microarrays
  - Traditional technology
  - Data are noisy

- RNA-Seq
  - Appears recently
  - Data are more accurate

*Pictures are found from internet, unauthorized*
Data

- Data Matrix $X \in \mathbb{R}^{p \times n}$
  - $p$ : number of genes
  - $n$ : number of samples
  - $p >> n$
Goal

- Build a undirected graph
  - Describing relationship between genes

- As simple as possible
Naïve Method

- For each pair of genes
  - Calculate Pearson Correlation Coefficient
    \[ \rho(X, Y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \]
  - Calculate P value
    - Permutation Tests
    - t–distribution
- Control False Discovery Rate (FDR)
Example Networks

Based on 431 mitosis cell cycle related microarray probe sets

Normal Leukocyte

Acute myeloid leukemia
If B, C depend on A, but are conditionally independent given A
  ◦ Say gene A regulates gene B and C
  ◦ Naïve method tends to connect B and C, too!

We want

Naïve method builds
Assume each sample is sampled from a multivariate Gaussian distribution $N(\mu, \Sigma)$.

$$N(x|\mu, \Sigma) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left\{-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}$$

Let $\Theta = \Sigma^{-1}$. This is precision matrix

- If $\theta_{ij}$ is 0, variable i & j are conditionally independent given all other variables
- Otherwise, connect i and j in the graph
Graphical Lasso
[Friedman et al. 2008]


- Assume $\mu$ is fixed, say $\mu = 0$
- Let $S = \frac{1}{n} XX^T$
- Want to maximize penalized log-likelihood

$$\log \text{det} \theta - \text{tr}(S\theta) - \rho \|\theta\|_1$$
Graphical Lasso

Define $W = \Theta^{-1}$

The subgradient equation is

$$W - S - \rho \cdot \Gamma = 0$$

$$\Gamma_{ij} = \begin{cases} 
\text{sgn}(\theta_{ij}), & \theta_{ij} \neq 0 \\
\gamma_{ij} \in [-1,1], & \theta_{ij} = 0 
\end{cases}$$

Partition $W$ as follows:

$$\begin{pmatrix} W_{11} & w_{12} \\
 w_{12}^T & w_{22} \end{pmatrix} \begin{pmatrix} \Theta_{11} & \theta_{12} \\
 \theta_{12}^T & \theta_{22} \end{pmatrix} = \begin{pmatrix} I & 0 \\
 0 & 1 \end{pmatrix}$$
Graphical Lasso

- Assume $W_{11}$ is known, we can find $w_{12}$ by solving

$$
\min_\beta \left\{ \frac{1}{2} \left\| W_{11}^{1/2} \beta - W_{11}^{-1/2} s_{12} \right\|^2 + \rho \| \beta \|_1 \right\} \quad (*)
$$

- $w_{12}$ is given by

$$
w_{12} = W_{11} \hat{\beta}, \quad \hat{\beta} \text{ solves } (*)
$$

- $(*)$ is a lasso problem!
Graphical Lasso

- To show it, the upper right block of penalized log-likelihood’s subgradient equation is

\[ w_{12} - s_{12} - \rho \cdot \gamma_{12} = 0 \]

- The subgradient equation of (*) is

\[ W_{11} \beta - s_{12} + \rho \cdot \nu = 0 \]

\[ \nu_i = \begin{cases} \text{sgn}(\beta_i), & \beta_i \neq 0 \\ \nu_i \in [-1,1], & \beta_i = 0 \end{cases} \]
Graphical Lasso

Algorithm
- Start with $W = S + \rho I$
- For $j = 1, 2, \ldots, p$
  - Permute $j$th diagonal item to the lower right corner of $W$
  - Solve the lasso problem (*) and get $w_{12} = W_{11} \hat{\beta}$
  - Recover the corresponding row and column with $w_{12}$
- Continue until convergence
Some Related References

Q&A

Thanks!