On Sparse Nonparametric Conditional Covariance Selection

Mladen Kolar, Ankur Parikh, Eric Xing

{mladenk,apparikh,epxing}@cs.cmu.edu

Machine Learning Department
Carnegie Mellon University
Estimating Associations among Variables

Important to many researchers in......
Biology: Gene Regulatory Networks
Social Networks: Twitter

- ivanmurat
  - lezanje u stanu, gledanje seriija i slusanje maidena.
- mbudisic
  - u mom kvartu status quo, planovi u vakumu. #fb
- cmucycling
  - Congratulations to Travis Wolfe elected new President of the Cycling Club!
- mknar
  - needs to organize his desk. maybe when there is no space left to put a coffee mug...
- webUGhr
  - RT @hhvrjoje: For all asp.net mvc folks, check out videos from C4MVC.
    http://bit.ly/P3ibJ #aspnetmvc
- dev102
  - Check out our latest article, "DebuggerDisplay and DebuggerBrowsable &nadash; True Debugging? All you should know.
- elijahmanor
  - RT @NETTuTs: This will blow your mind.
- igorcandadi
  - Mathematika 2K10 - online natjecanje za sve koji vole spoj matematike i programiranja.
Finance: Stock Associations

Return on Investment

Jan 1, 2000 – Oct 1, 2009

Source: MSN Money.com, Case Shiller


http://minnesota.publicradio.org/display/web/2008/04/15/nwa2/
Covariance Selection

- Assume data is iid and system is isolated

- Estimate non-zero elements of inverse covariance matrix
  [Meinshausen, Buhlmann 06, Ravikumar et al. 08]

- Can be visualized as network where edges are non-zero precision matrix elements
Environmental Variables

- More realistic to model the associations as **functions of the environment**
  - **Genes** in a regulatory network **conditioned on blood pressure**
  - **Stocks** in the NYSE **conditioned on economic indicator**
  - **People** in a social network **conditioned on time**
- Environmental variable can be **continuous and random**
Conditional Covariance Estimation

- Assume that the set of edges remains fixed, but the values can change.

Low oil price

![Graph with GOOG, INTC, AAPL, LUV, XOM nodes in a network]  

high oil price

![Graph with GOOG, INTC, AAPL, LUV, XOM nodes in a network]
Conditional Covariance Estimation

\[ \mathbf{X} \in \mathbb{R}^p \]  Set of variables (stocks)

\[ \mathbf{Z} \in \mathbb{R} \]  Environmental variable (oil price)

\[ \Sigma(\mathbf{z}) := \text{Cov}(\mathbf{X}|\mathbf{Z} = \mathbf{z}) \]  Conditional covariance matrix

\[ \Omega(\mathbf{z}) := \Sigma(\mathbf{z})^{-1} \]  Conditional precision matrix

Goal is to select non-zero elements of conditional precision matrix (network edges)
Partial Correlation

- Partial Correlation - correlation between two variables conditioned on the rest

- Related to precision matrix elements

\[ \rho_{uv}(z) = -\frac{\omega_{uv}(z)}{\sqrt{\omega_{uu}(z)\omega_{vv}(z)}} \]
Neighborhood selection

\[ X_u = \sum_{v \neq u} X_v b_{uv}(z) + \epsilon_u(z), \quad u \in [p] \]

\[ \rho_{uv}(z) = \text{sign}(b_{uv}(z)) \sqrt{b_{uv}(z)b_{vu}(z)} \]
Neighborhood selection

\[ X_u = \sum_{v \neq u} X_v b_{uv}(z) + \epsilon_u(z), \quad u \in [p] \]

\[ \rho_{uv}(z) = \text{sign}(b_{uv}(z)) \sqrt{b_{uv}(z)b_{vu}(z)} \]
Neighborhood selection

\[ X_u = \sum_{v \neq u} X_v b_{uv}(z) + \epsilon_u(z), \quad u \in [p] \]

\[ \rho_{uv}(z) = \text{sign}(b_{uv}(z)) \sqrt{b_{uv}(z)b_{vu}(z)} \]
Loss Function

\[ \mathcal{L}_u(B_u; \mathcal{D}^n) := \sum_{z \in \{z^j\}_{j \in [n]}} \left( \sum_{i \in [n]} (x^i_u - \sum_{v \neq u} x^i_v b_{uv}(z))^2 \right) + 2\lambda \sum_{v \neq u} \sqrt{\sum_{z \in \{z^j\}_{j \in [n]}} b_{uv}(z)^2} \]

**RSS:** residual sum of squares

**Kernel:** smooth across \( z \)

**Group penalty:** fixed set of non-zero elements
Determining Structure

Structure is fixed for all $z$, so

$$\hat{S} := \{(u, v) : \max\{\|\hat{b}_{uv}(\cdot)\|_2, \|\hat{b}_{vu}(\cdot)\|_2\} > 0\}$$
Optimization

- Modified active shooting algorithm /coordinate descent [Friedman et al. 2010]

- Optimize one group while holding others fixed.

\[
\mathcal{L}_u^v(\{b_{uv}(z^j)\}_{j \in [n]}; D^n) := \\
\sum_{z \in \{z^j\}_{j \in [n]}} \sum_{i \in [n]} \left( r_{uv}^i(z) - x_v^i b_{uv}(z) \right)^2 K_h(z - z^i) \\
+ 2\lambda \|b_{uv}(\cdot)\|_2,
\]

\[
r_{uv}^i(z) = x_u^i - \sum_{v' \neq u, v} x_{v'}^i b_{uv'}(z)
\]
Optimization

- Fast way to check if group is zero

\[
\frac{1}{\lambda^2} \sum_{z \in \{z^j\}} \left( \sum_{i \in [n]} x^i_u r^i_u(z) K_h(z - z^i) \right)^2 \leq 1.
\]

- Otherwise optimize block (i.e. using gradient descent)
Consistent Structure Estimation

Let $h = O(n^{-1/5})$, $\lambda = O(n^{7/10} \sqrt{\log p})$ and $n^{-9/5} \lambda \to 0$.

If $\frac{n^{11/10}}{\sqrt{\log p}} \min_{u,v \in S} \|b_{uv}(\cdot)\|_2 \to \infty$, then $\mathbb{P}[\hat{S} = S] \to 1$. 
Selecting Regularization

\[ \text{BIC}_u(\lambda) = \log(\text{RSS}_u(\lambda)) + \frac{\hat{df}_u, \lambda (\log(nh) + 2 \log p)}{nh} \]

Extra penalty (Chen & Chen 08)

- Log residual sum of squares
- Number of non-zero elements in row
Selecting Regularization

\[ \hat{\lambda} = \arg\min_\lambda \sum_{u \in [p]} \text{BIC}_u(\lambda) \]

\[ \mathbb{P}[\hat{S}(\hat{\lambda}) = S] \to 1. \]
Simulations: A Toy Example

1) Constant
2) Constant
3) Piecewise Constant
4) Linear
5) Sinusoid

\[ \Omega(z) = \begin{pmatrix}
1 & 4 \\
1 & 3 \\
3 & 2 \\
4 & 5
\end{pmatrix} \]
Toy Example Results

- **Ideal**
  - Frequency distribution for Precision Matrix element.

- **MB (static)**
  - Frequency distribution for Precision Matrix element.

- **Kernel + L1 penalty**
  - Frequency distribution for Precision Matrix element.

- **Kernel + Group penalty**
  - Frequency distribution for Precision Matrix element.
Larger Simulations

- 8x8 grid
- All non-zero precision matrix elements are sinusoids.
Analyzing the S&P 500

- Examine associations among stocks
  - Help an economist studying the market
  - Assist an investor building a diverse portfolio


- Condition on oil price, an economic indicator
Network
Clusters

- Industrial/manufacturing
- Home/office
- Oil/drilling/energy
- Energy/utilities
- Retail stores
- Defense
- Communications
- Financial
- Real estate
- Hotels/resorts
- Technology/semiconductor
- Railroads
Hubs

Citigroup

JP Morgan Chase
**Edge Weights**

- Edge weights (proportional to partial correlations) reflect changes in associations.
Discussion

- Estimate associations among variables conditioned on the environment

- Applicable to biology, finance, social networks etc.

- Our method is simple, nonparametric, and works well in high dimensions
Acknowledgements

ONR

NSF

NIH

Alfred P. Sloan Fellowship to EPX
Thank you!