High Dimensional Models

Time-Varying Graphical Lasso

Andrew Boomer & Jacob Pichelmann

Toulouse School of Economics
M2 EEE

January 29, 2021

Introduction to Graphs

- A graph is represented by a set of vertices $V = \{1, 2, \cdots, p\}$ and by a set of edges $E \subset VxV$
- A graph is undirected when there is no distinction between the edge (s, t) and $(t, s) \quad \forall\ s, t \in E$
- Graphs can also be thought of in terms of Markov chains, and some of the properties and representations of Markov Chains apply to graphs as well
- For example, the edge set of a graph can be represented as a Markov transition matrix. For an edge set $E \subset (a, b)x(a, b)$ the matrix would be:

$$\begin{bmatrix} W_{(a,a)} & W_{(a,b)} \\ W_{(b,a)} & W_{(b,b)} \end{bmatrix}$$

Extension to Graphical Models

- This graphical representation can be extended to a high dimensional set of random variables X_s.
- In this example, s corresponds to a single vertex with the whole vertex set V of the total graph. The connections between each vertex in the Markov transition matrix representation quanity the relationship between this random variables.

The two in Figure 1 are equivalent, where white spaces indicate no relationship, and grey spaces indicate a 1-to-1 relationship.

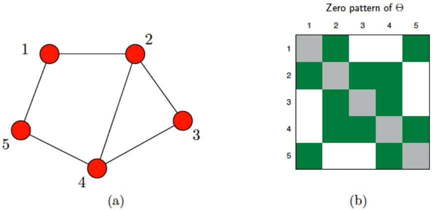

Figure: Graph and Sparsity Matrix

Gaussian Graphical Models

- For a guassian graphical model, we start from a gaussian distribution with dimension equal to the number of vertices $X \sim \mathcal{N}(\mu, \Sigma)$
- The sparsity matrix in the slide above, for a gaussian graphical model, is given by $\Theta = \Sigma^{-1}$, and is also known as the precision matrix. The goal of a gaussian graphical model is to solve for Θ.
- The gaussian distribution can be rewritten in terms of a maximum likelihood problem to solve for $\hat{\Theta}_{MLE}$, where S is the empirical covariance matrix, rearranged and simplified to:

$$\mathcal{L}(\Theta; X) = log\ det\ \Theta - trace(S\Theta)$$

- This MLE problem converges to the true precision matrix when $N \to \infty$. The MLE problem is, however, not even feasible if $p > N$, where p is the number of dimensions.

Graphical LASSO

- As with a regression LASSO, adding a regularization term solves the issue of a non-full rank matrix. Additionally, the LASSO term induces low-importance terms to zero.
- In the case of the graphical LASSO, weak edge connections in the precision matrix will go to zero, increasing the sparsity of the resulting solution.
- Whereas in the regression LASSO, large values of the β parameter were penalized, in the graphical LASSO, large values in the off-diagonal entries of the precision matrix will be penalized.
- Where λ is the LASSO penalty and $p_1(\Theta) = \sum_{s \neq t} |\theta_{st}|$, The maximization problem is:

$$\hat{\Theta} \in argmax\{log\ det\ \Theta - trace(S\Theta) - \lambda p_1(\Theta)\}$$

Time Varying Graphical LASSO

- In their paper [1], the authors discuss a method to estimate the change in a network's structure over time. This is accomplished with the addition of another parameter to penalize the change in network structure over time. The time-varying maximization problem is:

$$\hat{\Theta}_T \in argmax\{log\ det\ \Theta - trace(S\Theta) - \lambda p_1(\Theta) - \beta \sum_{t=2}^{T} \psi(\Theta_t - \Theta_{t-1}\}$$

- Here, ψ is a convex function which encourages similarity between Θ_t and Θ_{t-1}. Different choices of ψ can penalize different changes in network structure.

Time Penalty Functions

The authors lay out 5 different choices for the function ψ in [1].

1. **A few edges changing at a time:** $\psi(X) = \sum_{i,j} |X_{ij}|$
2. **Global Restructuring:** $\psi(X) = \sum_{j} ||[X]_j||_2$
3. **Smoothly Varying Over Time:** $\psi(X) = \sum_{i,j} X_{ij}^2$

David Hallac et al. "Network inference via the time-varying graphical lasso". In: (2017), pp. 205–213.