1 Graph Estimation

1.1 Regression Approach

Our Sparse Additive Convex/Concave Model (SCCAM) is expressed as

$$\min_{h,g} \frac{1}{2n}\sum_{i=1}^{n}(y_i-(h(\boldsymbol{x}_i)-g(\boldsymbol{x}_i)))^2+\lambda(\|\partial \boldsymbol{h}\|+\|\partial \boldsymbol{g}\|) \text{ s.t. } h(\boldsymbol{x}), g(\boldsymbol{x}) \text{ are additive concave, } \sum_{i=1}^{n} h(\boldsymbol{x}_i)=0, \sum_{i=1}^{n} g(\boldsymbol{x}_i)=0.$$

The above model can be applied to neighborhood selection for graph estimator, although we do not have a good theoretical justification for the joint distribution on the graph.

Here is some thought to justify the above program. Suppose the joint distribution $p(\boldsymbol{x}, y) = e^{f(\boldsymbol{x},y)}$ is log-concave, i.e., $f(\boldsymbol{x}, y)$ is concave. Then

$$E(e^y|\boldsymbol{x}) = \frac{\int e^y e^{f(\boldsymbol{x},y)}dy}{\int e^{f(\boldsymbol{x},y)}dy} \triangleq \frac{e^{h(\boldsymbol{x})}}{e^{g(\boldsymbol{x})}} = e^{h(\boldsymbol{x})-g(\boldsymbol{x})}.$$

According to the properties of log-concavity, $h(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are concave. The above equation is equivalent to $E(e^{y-(h(\boldsymbol{x})-g(\boldsymbol{x}))}|\boldsymbol{x}) = 1$ or expressed in the sample version $\frac{1}{n}\sum_{i=1}^{n} e^{y_i-(h(\boldsymbol{x}_i)-g(\boldsymbol{x}_i))} \to 1$. Hence we can formulate the following program

$$\min_{h,g} \frac{1}{n}\sum_{i=1}^{n} e^{y_i-(h(\boldsymbol{x}_i)-g(\boldsymbol{x}_i))} \text{ s.t. } h(\boldsymbol{x}), g(\boldsymbol{x}) \text{ are (additive) concave, } \sum_{i=1}^{n}(y_i-(h(\boldsymbol{x}_i)-g(\boldsymbol{x}_i)))=0.$$

The last equality constraint is for normalization purpose. The objective function will approach 1, since according to Jensen's inequality

$$\log\left(\frac{1}{n}\sum_{i=1}^{n} e^{y_i-(h(\boldsymbol{x}_i)-g(\boldsymbol{x}_i))}\right) \geq \frac{1}{n}\sum_{i=1}^{n}(y_i-(h(\boldsymbol{x}_i)-g(\boldsymbol{x}_i))) = 0,$$

which means the minimum value can be obtained by the objective function is 1. In addition, the current objective function is convex and reduces to the traditional ℓ_2 loss asymptotically when $y_i-(h(\boldsymbol{x}_i)-g(\boldsymbol{x}_i))$ is small. i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} e^{y_i-(h(\boldsymbol{x}_i)-g(\boldsymbol{x}_i))} = \frac{1}{n}\sum_{i=1}^{n}(1+(y_i-(h(\boldsymbol{x}_i)-g(\boldsymbol{x}_i)))+\frac{1}{2}(y_i-(h(\boldsymbol{x}_i)-g(\boldsymbol{x}_i)))^2+\cdots) = 1+\frac{1}{2n}\sum_{i=1}^{n}(y_i-(h(\boldsymbol{x}_i)-g(\boldsymbol{x}_i)))^2+\cdots.$$

We thought of using $\frac{1}{n}\sum_{i=1}^{n}(e^{y_i-(h(\boldsymbol{x}_i)-g(\boldsymbol{x}_i))}-1)^2$ as the objective function before, but it is not convex.

However, there might still be some issues in the above derivation. In particular, it is not clear whether we can assume $\frac{1}{n}\sum_{i=1}^{n}(y_i-(h(\boldsymbol{x}_i)-g(\boldsymbol{x}_i)))=0$, which implicitly assumes $E(y-(h(\boldsymbol{x})-g(\boldsymbol{x})))=0$. Other directions to find a justifiable graph estimation method include maximum pseudo-likelihood (see notes by John).

A similar approach for graph estimation was proposed in [Voorman, Shojaie & Witten, 2013], which also assumes that the regression function is additive. In addition, it assumes that each additive function can be represented by known functional bases, and group lasso type penalty is employed to select features. In the simulation below, we use linear, quadratic and cubic bases.

1.2 Simulation

We construct a simulation graph with log-concave density as follows. Consider the following joint distribution

$$p(\boldsymbol{x}, s) = p(\boldsymbol{x}|s) \times p(s) = \mathcal{N}(\boldsymbol{x}; \mathbf{0}, s\boldsymbol{I}_m) \times p(s) \quad (s > 0)$$

where $\boldsymbol{x} \in \mathbb{R}^m$ is drawn from a multivariate Gaussian with variance $s \in \mathbb{R}_+$ drawn from $p(s)$. Since

$$\log p(\boldsymbol{x}, s) = \log \mathcal{N}(\boldsymbol{x}; \mathbf{0}, s\boldsymbol{I}_m) + \log p(s) = -\frac{m}{2}\log(2\pi) - \frac{m}{2}\log s - \frac{\boldsymbol{x}^\top \boldsymbol{x}}{2s} + \log p(s)$$

and $\frac{\boldsymbol{x}^\top \boldsymbol{x}}{2s}$ is convex jointly for $(\boldsymbol{x}, s)$, it can be concluded that $p(\boldsymbol{x}, s)$ is log-concave if $\log p(s) - \frac{m}{2}\log s$ is concave [1]. Consider a Gamma prior on s:

$$p(s) = \mathrm{Ga}(s; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} s^{\alpha-1} \exp(-\beta s) \quad (s > 0).$$

Then according to the above analysis, $p(\boldsymbol{x}, s) = \mathcal{N}(\boldsymbol{x}; \mathbf{0}, s\boldsymbol{I}_m) \times \mathrm{Ga}(s; \alpha, \beta)$ is log-concave if $\alpha \geq 1 + \frac{m}{2}$. The conditional distribution $p(s|\boldsymbol{x})$ can be derived as

$$p(s|\boldsymbol{x}) \propto s^{-\frac{m}{2}} \exp(-\frac{\boldsymbol{x}^\top \boldsymbol{x}}{2s}) s^{\alpha-1} \exp(-\beta s) = \mathrm{GIG}(s; 2\beta, \boldsymbol{x}^\top \boldsymbol{x}, \alpha - \frac{m}{2})$$

where $\mathrm{GIG}(s; a, b, c) = \frac{(a/b)^{c/2}}{2K_c(\sqrt{ab})} s^{c-1} \exp(-\frac{1}{2}(as + \frac{b}{s}))$ $(s > 0)$ is the Generalized Inverse Gaussian (GIG) distribution with mean $E(s) = \frac{\sqrt{b}}{\sqrt{a}} \frac{K_{c+1}(\sqrt{ab})}{K_c(\sqrt{ab})}$, and $K_c(\cdot)$ is the modified Bessel function of the second kind. Hence the conditional mean can be expressed as

$$E(s|\boldsymbol{x}) = \frac{\|\boldsymbol{x}\|_2}{\sqrt{2\beta}} \frac{K_{\alpha - \frac{m}{2} + 1}(\sqrt{2\beta}\|\boldsymbol{x}\|_2)}{K_{\alpha - \frac{m}{2}}(\sqrt{2\beta}\|\boldsymbol{x}\|_2)}$$

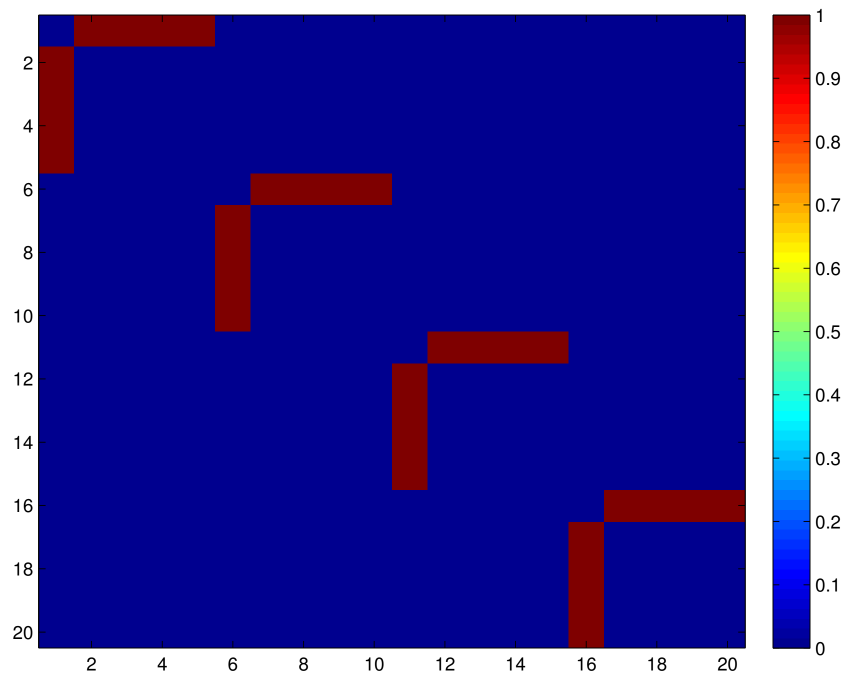
which is a nonlinear and non-additive function of $\boldsymbol{x}$.

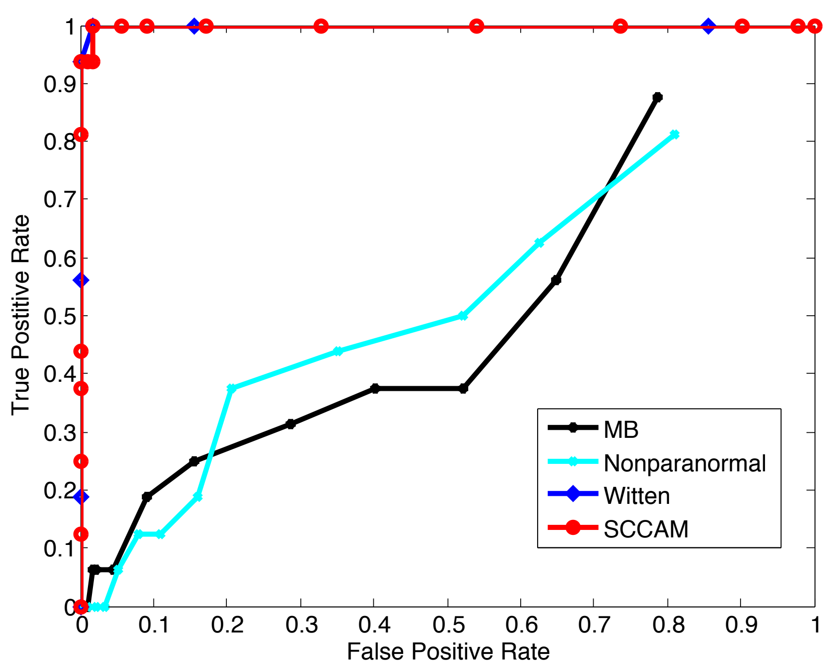
Motivated by the above analysis, we generate data as

$$s_{1i} \sim \mathrm{Ga}(\alpha, \beta), \boldsymbol{x}_{1i} \sim \mathcal{N}(\mathbf{0}, s_{1i}\boldsymbol{I}_m); \quad s_{2i} \sim \mathrm{Ga}(\alpha, \beta), \boldsymbol{x}_{2i} \sim \mathcal{N}(\mathbf{0}, s_{2i}\boldsymbol{I}_m); \quad \cdots\cdots \quad (i = 1, 2, \cdots, n)$$

Then the vector $(s_{1i}, \boldsymbol{x}_{1i}, s_{2i}, \boldsymbol{x}_{2i}, \cdots) \in \mathbb{R}^p$ is drawn from a log-concave graph with linkage structure depicted in Figure 1. In the experiment we set $p = 20, n = 500, m = 4, \alpha = 5, \beta = 0.5$. We then do neighborhood regression using SCCAM, Meinshausen-Bhlmann, and nonparanormal to identify the graph structure. We declare a edge between two nodes if an edge exists in either direction. As we observe in the ROC curve in Figure 1, both SCCAM and [Voorman, Shojaie & Witten, 2013] perform very well in this simulation example.

[1] Matthias Seeger, A Note on Log-Concavity, 2007.





True Postitive Rate
False Positive Rate
MB
Nonparanormal
Witten
SCCAM