Prof. Homrighausen gave a lecture on nonlinear embeddings, focusing on kernel PCA and Laplacian eigenmaps. These are manifold learning algorithms that respect the underlying geometry better than classical PCA, for example. The lecture notes are available on the class website and are quite straightforward. Thus the scribe will not add to them but rather turn to one of the student presentations given in class.

0   Topics in Graphical Models

Ryan Haunfelder gave a presentation to supplement our lecture materials on graphical models. He stepped back to square one to explain the ideas behind discrete graphical models. He then reviewed Gaussian graphical models which we discussed in the previous class. In a Gaussian graphical model our goal was to estimate the precision matrix to infer conditional independence between variables. There are two approaches: 1) global, i.e. estimating the precision matrix via maximum likelihood and 2) local, i.e. using regression techniques to take advantage of the fact that the conditional expectation of each variable can be written in terms of the other variables. However, when our data is not Gaussian we can't use the global approach, and there is no guarantee that the conditional means of the variables are linear. Voorman et al. (2014) suggested using jointly additive models (JAMs) for conditional means of variables, thus resulting in Sparse Conditional Estimation using JAMs, also known as SpaCE JAM. Ryan applied this new method to a psychology dataset and demonstrated the graph visualization software Gephi (check it out at: http://gephi.github.io/).

1   SpaCE JAM

Since I once went on a date with one of the authors of the SpaCE JAM paper (I'll leave it to you to guess which one), I decided to read the paper and summarize the highlights here.

- Estimating conditional independence graphs in the high-dimensional setting is a hot topic, but most prior work assumes either that variables are multivariate Gaussian or conditional means of the variables are linear (these two assumptions are essentially equivalent).

- However, this doesn't work for many real datasets. Voorman et al. (2014) give the example of a cell signalling dataset where there is a clear lack of multivariate normality even when transformed to be marginally normal. In addition, there is strong evidence that the conditional mean of protein P38 given PKC and PJNK is nonlinear.

- This leads the authors to propose a semi-parametric method which allows conditional means to take on an arbitrary additive form. They emphasize that "modelling the conditional distribution rather than the joint distribution is a more direct approach to graph estimation." To this end, they consider pseudolikelihoods of the form

$$\log\{p_{PL}(x;\theta)\} = \sum_{j=1}^{d} \log p_j(x_j | \{x_k : (j,k) \in E\};\theta)$$

where d is the dimension of the random variable X and we are estimating the graph $\Gamma = (V,E)$.

- Along the way, the authors briefly summarize/ criticize previous methods for estimating graphs with Gaussian data (neighborhood selection, sparse partial correlation estimation, and graphical lasso) as well as with non-Gaussian data (normal copula or nonparanormal model, forest density estimation, and graphical random forests).

- Voorman et al. (2014) model the conditional means with generalized additive models, such that

$$X_j | \{X_k : k \neq j\} = \sum_{k \neq j} f_{jk}(X_k) + \epsilon_j$$

  where $f_{jk}(\cdot) \in \mathcal{F}$ for some function space $\mathcal{F}$. They do not assume that the ϵ_j are independent of the $f_{jk}(\cdot)$, just that the conditional independence structure can be recovered from the $f_{jk}(\cdot)$.

- Since they believe that the conditional independence graph is sparse (this is also implied in the graphical lasso and other methods), the authors link together d sparse additive models using a penalty that groups the parameters corresponding to a single edge in the graph. This results in the optimization problem:

$$\min_{f_{jk} \in \mathcal{F}, 1 \leq j, k \leq d} \left[\frac{1}{2n} \sum_{j=1}^{d} ||x_j - \sum_{k \neq j} f_{jk}(x_k)||_2^2 + \lambda \sum_{k > j} \left\{ ||f_{jk}(x_k)||_2^2 + ||f_{kj}(x_j)||_2^2 \right\}^{1/2} \right]$$

  The functions under consideration are of the form $f_{jk}(x_k) = \Psi_{jk}\beta_{jk}$ where Ψ_{jk} is a $n \times r$ matrix whose columns are basis functions and β_{jk} is a vector of associated coefficients.

- The authors devise an estimation algorithm, which is implemented in the R package 'spacejam.' For tuning they take an approach motivated by BIC. They also propose an approximation to their method to speed up computation in high dimensions (see the paper for more details).

- In a simulation study, they show that SpaCE JAM performs better than existing methods when there are nonlinear relationships among the features and has comparable performance to methods that assume multivariate normality when the conditional means are linear.

And there you have it!

Figure 1: (Source: http://www.businessinsider.com/space-jam-sequel-2014-2)

References

[1] Arend Voorman, Ali Shojaie, and Daniela Witten. Graph estimation with joint additive models. *Biometrika*, 101(1):85–101, 2014.