Sparse Bayesian factor analysis

Ilsang Ohn

Department of statistics, SNU

March 3, 2017

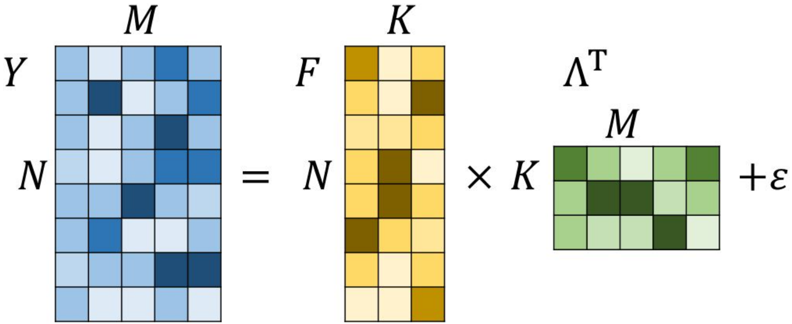
Factor analysis

- Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors, which have some practical meanings.
- Suppose there are N independent M-dimensional observations $\mathbf{y}_1, \ldots, \mathbf{y}_N$ with $\mathbf{y}_i = (y_{i1}, \ldots, y_{iM})^\top$. e.g.,
    - y_{ij} would be expression of gene j of sample i
    - y_{ij} would be user i's rating of item j
- $\mathbf{y}_i$ are said to have a factor structure if they are represented as

$$\mathbf{y}_i = \mathbf{\Lambda} \mathbf{f}_i + \varepsilon_i$$

  where $\mathbf{f}_i$ are K-dimensional unobserved random vectors of common factors with $K < M$, $\mathbf{\Lambda}$ is a $M \times K$ matrix of factor loadings and ε_i are noise vectors. Conventionally we assume $\mathbf{f}_i \sim \mathrm{N}_K(\mathbf{0}, \mathbf{I})$ and $\varepsilon_i \sim \mathrm{N}_M(\mathbf{0}, \mathbf{\Omega})$ with $\mathbf{\Omega}$ being diagonal.

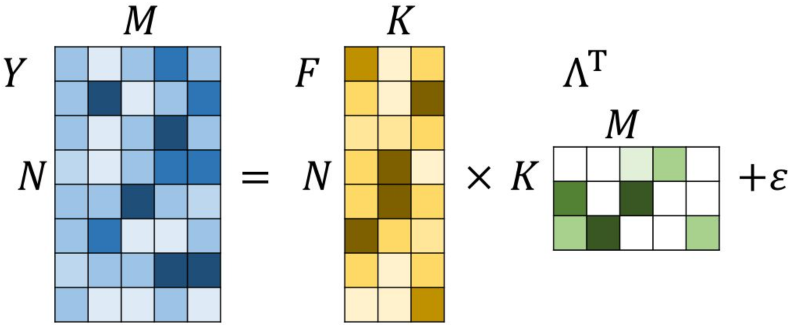
Factor analysis



Sparse factor analysis

- For several reasons, the loading matrix $\mathbf{\Lambda}$ is encouraged to be sparse, i.e., to have lots of zeros.
- Sparsity gives improved predictive performance, because factors irrelevant to a particular dimension are not included.
- Also, sparse models are more readily interpretable since a smaller number of factors are associated with observed dimensions.
    - genes link to some specific pathways
    - movies can be grouped into a small number of genres

Sparse factor analysis



Sparse Bayesian factor analysis

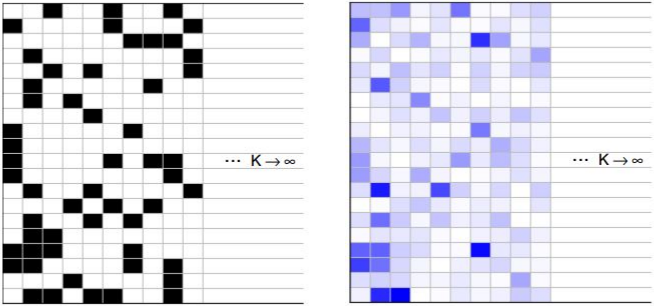
- In a Bayesian context, sparsity can be induced by using the spike and slab prior on the factor loadings. For each factor loading λ_{jk}, we use

$$\begin{aligned} \lambda_{jh} | \gamma_{jh} &\sim (1-\gamma_{jh})\delta_0 + \gamma_{jh}\mathrm{N}(0, \tau^{-1}) \\ \gamma_{jh} &\sim \mathrm{Bernoulli}(\theta) \end{aligned}$$

  where δ_0 denotes a point mass at zero.
- The matrix $\boldsymbol{\Gamma} = \{\gamma_{jk}\}_{j,k=1}^{M,K} \in \mathbb{R}^{M \times K}$ includes binary allocation indicators that characterize which factors are associated with each variable.

Choice of a number of factors

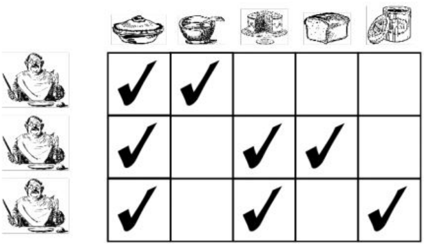
- Inference on the number of factors K in factor models is both conceptually and computationally challenging.
  - *traditional factor models*: fix K in advance and estimate $\mathbf{\Lambda} \in \mathbb{R}^{M \times K}$
- A possible solution is using the Indian Buffet Process (IBP) which defines a distribution over infinite binary matrices.
  - *infinitely many factor models*: give the IBP prior on $\mathbf{\Gamma} \in \mathbb{R}^{M \times \infty}$ and derive the posterior of $\mathbf{\Lambda} \in \mathbb{R}^{M \times \infty}$



(a) $\pi(\mathbf{\Gamma})$ (b) $\pi(\mathbf{\Lambda}|\mathbf{\Gamma})$

Indian buffet process

- The IBP is a distribution over infinite binary matrices.
- We can describe IBP in terms of the following restaurant analogy.
  - A customer enters a restaurant with an infinitely large buffet.
  - He helps himself to $\mathrm{Poisson}(\alpha)$ dishes.
  - The j-th customer enters the restaurant.
  - She helps herself to each dish with probability m_k/j, where m_k is the number of people whove tried dish k.
  - She then tries $\mathrm{Poisson}(\alpha/j)$ new dishes.



Sparse Bayesian factor models with the IBP prior

- Consider the infinite factor model:

$$\mathbf{y}_i = \boldsymbol{\Lambda} \mathbf{f}_i + \boldsymbol{\varepsilon}_i$$

where $\boldsymbol{\Lambda} \in \mathbb{R}^{p \times \infty}$.

- Use the IBP prior:

$$\begin{aligned} \lambda_{jh} | \gamma_{jh}, \tau_h &\sim (1 - \gamma_{jh})\delta_0 + \gamma_{jh} \mathrm{N}(0, \tau_h^{-1}) \\ \gamma_{jh} | \alpha &\sim \mathrm{IBP}(\alpha) \\ \alpha &\sim \mathrm{Gamma}(a_\alpha, b_\alpha) \\ \tau_h &\sim \mathrm{Gamma}(a_\tau, b_\tau) \end{aligned}$$

- The posterior distribution can be computed by using a straightforward Gibbs sampling algorithm.

Sparse Bayesian factor models with the IBP prior

The sparse Bayesian factor model with the IBP prior has many appealing properties, but it lacks

- theoretical support
- an efficient algorithm

Applications to marketing

- By the factor analysis, we can discover the descriptive features for predicting the item preferences. The loading vectors $\boldsymbol{\lambda}_j = (\boldsymbol{\lambda}_{j1}, \ldots, \boldsymbol{\lambda}_{jK})$ can represent item's feature (e.g., relative scores for movie j in K genres) and the factors $\mathbf{f}_i$ users' feature (e.g., user i's affinity for K genres).