Latent Variable Graphical Model Selection

Dongfang Wang

August 21, 2014

Outline

Outline

Gaussian Gaphical Model

What graphical model is concerned about?

Independentce,Correlation,and possible Causality between variables

Gaussian Graphical Model(Gaussian-Markov Random Field)

1. If all variables have a multivariate Gaussian distribution with
   - mean μ
   - covariance matrix Σ (positive definite)
   - concentration matrix $K = \Sigma^{-1}$

   the ijth component of Σ^{-1} is zero $\Rightarrow$ variables i and j are **conditionally independent**, given the other variables.
2. Undirected graph: no link means conditionally independent.
3. Parameter estimation and model selection:
   estimating parameters and identifying zeros in K

MLE for Gaussian

$$f(\mathbf{x};\mu,\Sigma) = \frac{1}{(2\pi)^{p/2}(det\Sigma)^{1/2}} exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

$$\log L(\mu,\Sigma;\mathbf{x}) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log det(\Sigma) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)$$

$$\max \quad \frac{n}{2}\log det(K) - \frac{1}{2}(x_i-\mu)^T C(x_i-\mu)$$

$$\mathbf{u}^* = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$S_n = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu^*)^T(x_i-\mu^*)$$

MLE

$$\max \quad \frac{n}{2}\log K - \frac{n}{2}Tr(KS_n)$$

Graphical Lasso

What the problem of MLE?

1. Large number of unknown parameters to be estimated
2. high-dimensional model
3. not a stable estimate of Σ^{-1}
4. not lead to "sparse" graph structure

Lasso for covariance selection

$$-\log det(K) + Tr(KS_n) + \lambda \sum_{i \neq j} |k_{ij}|$$

$$\begin{aligned} \min \quad & -\log det(K) + Tr(KS_n) \\ s.t. \quad & \sum_{i \neq j} |k_{ij}| \leq t \end{aligned}$$

Any problem?

Latent variables

"Given all other variables" ?

- not all other variables observed
- latent variables lead to failure of graphical lasso

Outline

Setup

All variables: $X \in R^{p+h}$ **(jointly Gaussian)**

- Observed variables: O, $|O| = p$, $X_O \in R^p$
- Latent Variables: H, $|H| = h$, $X_H \in R^h$

Covariance matrix and Concentration matrix:

$$\Sigma_{(O,H)} = \begin{pmatrix} \Sigma_O & \Sigma_{O,H} \\ \Sigma_{H,O} & \Sigma_H \end{pmatrix} \quad K_{(O,H)} = \begin{pmatrix} K_O & K_{O,H} \\ K_{H,O} & K_H \end{pmatrix}$$

Schur complement

$$\tilde{K_O} = (\Sigma_O)^{-1} = K_O - K_{O,H}(K_H)^{-1}K_{H,O}$$

Decomposition

Schur complement

$$\tilde{K_O} = (\Sigma_O)^{-1} = K_O - K_{O,H}(K_H)^{-1}K_{H,O}$$

- $S = K_O$: the concentration matrix of the *conditional statistics* of the observed variables given the latent variables.

  sparse
  (conditional statistics are given by a sparse graphical model)

- $L = K_{O,H}(K_H)^{-1}K_{H,O}$:a *summary* of the effect of marginalization over the latent variables X_H

  small rank
  (number of latent variables is small relative to the number of observed variables)

$\tilde{K_O}$: *not sparse* due to the additional low-rank term

Latent Components

- The sum of a sparse matrix and a low-rank matrix
- the *conditional* graphical model structure in the observed variables as well as the number of and effect due to the unobserved latent variables

Low rank : Invariant property

$infinitely$ many configurations of the latent variables:

$$B \in R^{h \times h} \ nonsigular$$
$$\hat{K_H} = BK_H B^T, \hat{K_{O,H}} = K_{O,H} B^T$$

X_O: marginal statistics remain the same upon marginalization over the latent variables

Low-rank summarizes the effect of marginalization over the latent variables

Algebraically correct

An estimate $(\hat{S},\hat{L})$ of a latent-variable Gaussian graphically model given by the concentration matrix $K_{(O,H)}$:

1. sign-pattern of $\hat{S}$ is the same as that of S:
correct structural estimation

$$sign(\hat{S}_{ij}) = sign(S_{i,j})$$

2. the rank of $\hat{L}$ is the same as that of L
the numberof latent variables(rank) is properly estimated

$$rank(\hat{L}) = rank(L)$$

3. a realizable graphical model

$$\hat{S} - \hat{L} > 0 \quad \hat{L} \geq 0$$

Problem Statement

Goal

Suppose we have n samples $\{X_O^i\}_{i=1}^n$ of the observed variables X_O.

Produce estimates $(\hat{S}_n, \hat{L}_n)$ with high probability are algebraically correct and have bounded estimation error (in some norm)

Approach

$$(\hat{S}_n, \hat{L}_n) = arg \min_{S,L} \; -l(S-L; \Sigma_O^n) + \lambda_n(\gamma||S||_1 + tr(L))$$
$$s.t. \; S-L > 0, L \geq 0$$

Regularized Maximum-likelihood Decompostion Framework

$$(\hat{S}_n, \hat{L}_n) = arg \min_{S,L} \ -l(S-L;\Sigma_O^n) + \lambda_n(\gamma||S||_1 + tr(L))$$
$$s.t. \ S-L>0, L \geq 0$$

log-likelihood

$$-\log det(K) + Tr(KS_n)$$

$K = S - L > 0$

Penalization

trace functional is the usual convex surrogate for the rank over the cone of positive semidefinite matrices

Adaptivity

Suppose there is no hidden variable

The $L+S$ model behaves as well or nearly as well as the graphical lasso?

$The\ analysis\ in\ the\ paper\ doesn't\ reach\ this\ conclusion$

From related problems,suppose we have incomplete and corrupted information about an $n_1 \times n_2$ low-rank matrix L^0:

$$M_{ij} = L_{ij}^0 + S_{ij}^0$$

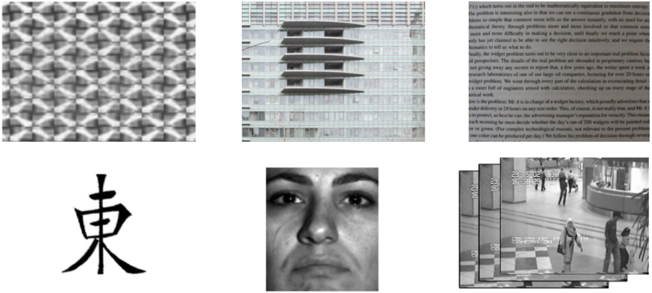
regard S^0 as a corruption pattern which is sparse.

$$\min \quad ||L||_* + \lambda ||S||_1 \quad s.t.\ M_{ij} = L_{ij} + S_{ij}$$

$$\min \quad ||b||_1 + \lambda ||e||_1 \quad s.t.\ y = Xb + e\ (robust\ sparse\ regression)$$

Outline

Low dimensional structures in visual data

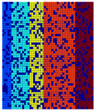


Visual data exhibit low-dimensional structures due to rich local regularities, global symmetries, repetitive patterns, or redundant sampling.

L+S Model

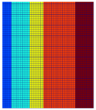
Video Y **=** **Low-rank appx.** X **+ Sparse error** E

Low Rank Matrix Recovery from corrupted observations



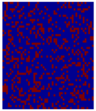
Matrix of corrupted observations

$\longrightarrow$



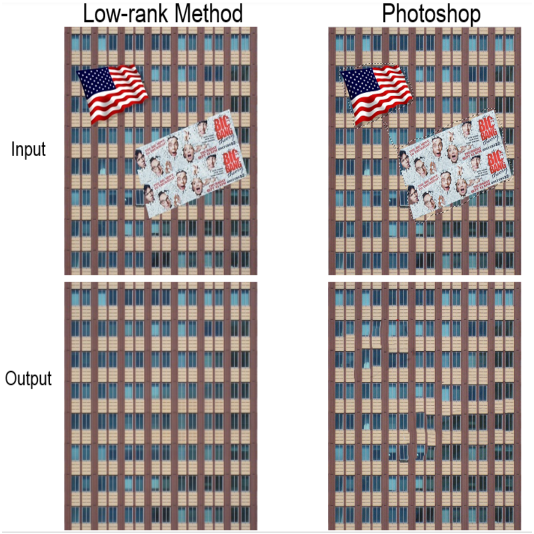
Underlying low-rank matrix

$+$

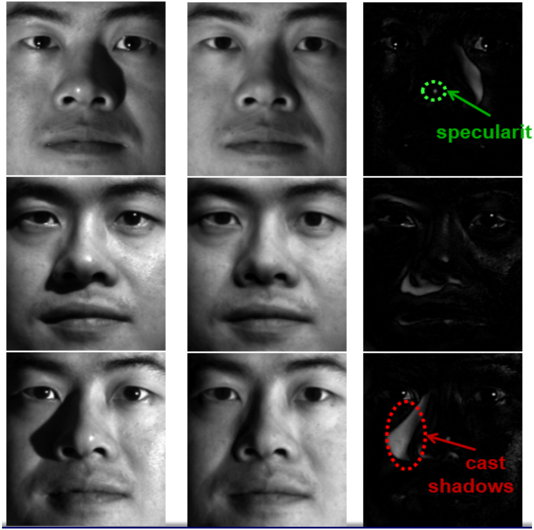


Sparse error matrix

Repairing Low-rank Textures



Repairing Multiple Correlated Images

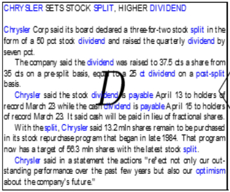


Web Document Corpus Analysis

Latent Semantic Indexing: the classical solution (PCA)

Documents

Words


CHRYSLER SETS STOCK SPLIT, HIGHER DIVIDEND

Chrysler Corp said its board declared a three-for-two stock split in the form of a 50 pct stock dividend and raised the quarterly dividend by seven pct.

The company said the dividend was raised to 37.5 cts a share from 35 cts on a pre-split basis, equal to a 25 ct dividend on a post-split basis.

Chrysler said the stock dividend is payable April 13 to holders of record March 23 while the cash dividend is payable April 15 to holders of record March 23. It said cash will be paid in lieu of fractional shares.

With the split, Chrysler said 13.2 mln shares remain to be purchased in its stock repurchase program that began in late 1984. That program now has a target of 56.3 mln shares with the latest stock split.

Chrysler said in a statement the actions "reflect not only our outstanding performance over the past few years but also our optimism about the company's future."

D

d_{ij} word frequency (or TF/IDF)

$$D = A + Z$$
$$= U_1\Sigma_1V_1^T + \underline{U_2\Sigma_2V_2^T}$$

Dense, difficult to interpret

a better model/solution?

$$D = A + \underline{E}$$

Low-rank "background" topic model

Informative, discriminative "keywords"

Outline

Discussion

Assumptions

- Gaussian distribution
- Conditionally independence (sparse)
- Not too may latent variables (low-rank)

Computation

- All we need is the sample cavariance matrix
- efficient convex programming

L+S model