Joint Gaussian Graphical Model Review Series – II

Gaussian Graphical Model Basics

Beilun Wang

[1]Department of Computer Science, University of Virginia
http://jointggm.org/

June 30th, 2017

Outline

Notation

Notation

Σ The covariance matrix.

Ω The precision matrix.

μ The mean vector.

x_i The i-th sample follows multivariate normal distribution.

Reviews

Reviews

- Probability basics
- Dependency vs. Correlation
- Conditional dependency vs. partial Correlation

Summary from last talk

- Partial correlation is easy to estimate the value while conditional independence is a relationship to infer.
- In the Gaussian Case, they are equivalent.
- From the structure learning angle, conditional dependence is about the causal relationship, while partial correlation is, more specifically, the linear relationship.

So the remaining question is why in the Gaussian case they are equivalent and how to infer this relationship.

Review: Gaussian Example

Suppose (X, Y) are uncorrelated. i.e.,$(X, Y) \sim N(0, diag(\sigma_X^2, \sigma_Y^2))$.

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_X\sigma_Y} \exp(-\frac{1}{2}(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2})) \\ &= \frac{1}{\sqrt{2\pi}\sigma_X} \exp(-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}) \frac{1}{\sqrt{2\pi}\sigma_Y} \exp(-\frac{1}{2}\frac{(y-\mu_Y)^2}{\sigma_Y^2}) \quad (2.1) \\ &= f(x)f(y) \end{aligned}$$

Therefore, if (X, Y) follows bivariate Gaussian, (X, Y) are uncorrelated if and only if (X, Y) are independent.

Why partial correlation and condition dependence are equivalent in the Gaussian case?

Multivariate Gaussian Distribution

Density function

Let $X \sim N(\mu, \Sigma)$. $f(x) = (2\pi)^{-\frac{p}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu))$

Partition X, μ, and Σ

Partition X, μ, Σ, Ω.

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

$$\Omega = \Sigma^{-1} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$$

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

Conditional Distribution of Multivariate Gaussian

If $X \sim N(\mu, \Sigma)$, it holds that $X_2 \sim N(\mu_2, \Sigma_{22})$.
If Σ_{22} is regular, it further holds that

$$X_1 | (X_2 = a) \sim N(\mu_{1|2}, \Sigma_{1|2})$$

where $\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(a - \mu_2)$, and
$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = (\Omega_{11})^{-1}$.

Partial correlation and condition dependence are equivalent in the Gaussian case

$X_1|X_2 = a \sim N(\mu_{1|2}, (\Omega_{11})^{-1})$,
If X_1 only contains x_i and x_j, then x_i and x_j are conditional independent on others iff $\Omega_{ij} = 0$.

Estimate the condition dependence graph/Partial correlation

Now the only thing left is to estimate $\Omega = \Sigma^{-1}$. There are three potential ways to do that. We call this problem as Gaussian Graphical model.

- Directly calculate the inverse of the sample covariance matrix $\widehat{\Sigma}$. However, we cannot do that when the sample covariance matrix is not invertible.
- Maximum Likelihood Method
- Regression method

For the first one, the sample covariance matrix $\widehat{\Sigma}$ may not be invertible.

Maximum Likelihood Method

The MLE of μ

$\mathcal{L}(\mu, \Omega) = (2\pi)^{-\frac{np}{2}} \prod_{i=1}^{n} \det(\Omega^{-1})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x_i - \mu)^T \Omega (x_i - \mu)\right)$.
After take a first derivative, it is easy to show that $\bar{x} = \frac{x_1 + \cdots + x_n}{n}$

The Likelihood of Ω

$\mathcal{L}(\bar{x}, \Omega) = (2\pi)^{-\frac{np}{2}} \prod_{i=1}^{n} \det(\Omega^{-1})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x_i - \bar{x})^T \Omega (x_i - \bar{x})\right)$.
Notice that $(x_i - \bar{x})^T \Omega (x_i - \bar{x})$ is a scalar. Therefore,
$(x_i - \bar{x})^T \Omega (x_i - \bar{x}) = \text{trace}((x_i - \bar{x})^T \Omega (x_i - \bar{x}))$.

The Likelihood of Ω

Since $\operatorname{tr}(A, B) = \operatorname{tr}(B, A)$.

$$\mathcal{L}(\bar{x}, \Omega) \propto \det(\Omega^{-1})^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \operatorname{tr}\left(\left(x_i - \bar{x}\right)^T \Omega \left(x_i - \bar{x}\right)\right)\right) \tag{4.1}$$

$$= \det(\Omega^{-1})^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \operatorname{tr}\left(\left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)^T \Omega\right)\right) \tag{4.2}$$

$$= \det(\Omega^{-1})^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left(\sum_{i=1}^{n} \left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)^T \Omega\right)\right) \tag{4.3}$$

$$= \det(\Omega^{-1})^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left(S\Omega\right)\right) \tag{4.4}$$

where, $S = \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T \in \mathbb{R}^{p \times p}$.

The Log-Likelihood of Ω

$$\ln \mathcal{L}(\bar{x}, \Omega) = \text{const} - \tfrac{n}{2} \ln \det(\Omega^{-1}) - \tfrac{1}{2} \text{tr} \left(\Omega \sum_{i=1}^{n} (\bar{x} - \mu)(\bar{x} - \mu)^T \right).$$

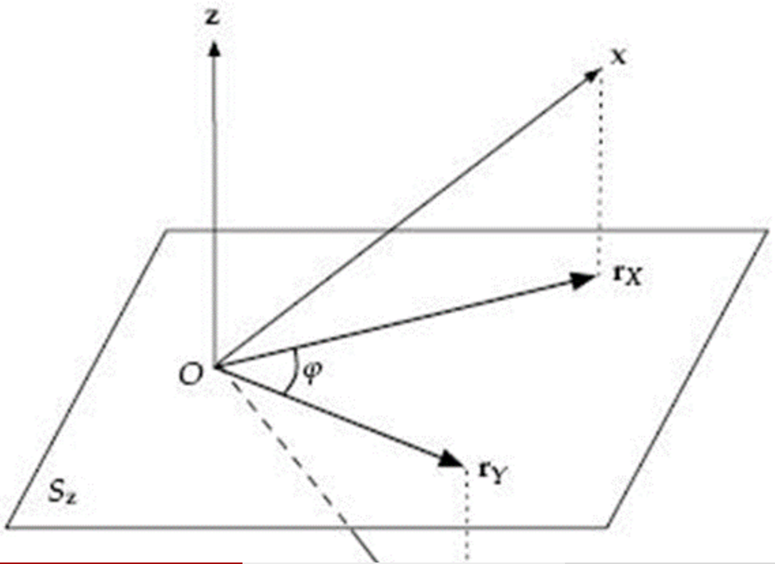
Since $\det(A^{-1}) = 1/\det(A)$,

$$\begin{aligned} \ln \mathcal{L}(\bar{x}, \Omega) &\propto \ln \det(\Omega) - \text{tr} \left(\Omega \frac{1}{n} \sum_{i=1}^{n} (\bar{x} - \mu)(\bar{x} - \mu)^T \right) && (4.5) \\ &= \ln \det(\Omega) - \text{tr} \left(\Omega \widehat{S} \right) && (4.6) \end{aligned}$$

where $\widehat{S}$ is the sample covariance matrix.

Regression Method

Partial Correlation



Partial correlation

- As we know, the partial correlation can also be solved by the linear regression.
- In the Gaussian case, we can use so-called neighborhood approach.

Conditional Distribution of Multivariate Gaussian

If $X \sim N(\mu, \Sigma)$, it holds that $X_2 \sim N(\mu_2, \Sigma_{22})$.
If Σ_{22} is regular, it further holds that

$$X_1 | X_2 = a \sim N(\mu_{1|2}, \Sigma_{1|2})$$

where $\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(a - \mu_2)$, and
$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = (\Omega_{11})^{-1}$.

Neighborhood approach

If $X \sim N(0, \Sigma)$ and let $X_1 = X_j$.
$X_j | X_{\setminus j}$ $N(\Sigma_{\setminus j, j}\Sigma^{-1}_{\setminus j, \setminus j}X_{\setminus j}, \Sigma_{jj} - \Sigma_{\setminus j, j}\Sigma^{-1}_{\setminus j, \setminus j}\Sigma_{\setminus j, j})$
Let $\alpha_j := \Sigma_{\setminus j, j}\Sigma^{-1}_{\setminus j, \setminus j}$ and $\sigma_j^2 := \Sigma_{jj} - \Sigma_{\setminus j, j}\Sigma^{-1}_{\setminus j, \setminus j}\Sigma_{\setminus j, j}$. We have that

$$X_j = \alpha_j^T X_{\setminus j} + \epsilon_j \tag{5.1}$$

where $\epsilon_j \sim N(0, \sigma_j^2)$ is independent of $X_{\setminus j}$.

Neighborhood approach

- We can estimate the α_j by solving p simple linear regression.
- if i-th entry of α_j equals to 0, it means that X_i and X_j are partial uncorrelated and conditional independent.
- Perhaps we want more assumption on α_j like sparsity.

Summary

- In Gaussian case, the partial correlation and the conditional dependence are equivalent
- We have two ways to estimate them. First, directly estimate the precision matrix by MLE. Second, solve p linear regression problem by neighborhood approach.
- None of them have any assumptions on the partial correlation coefficient.
- In the next talk, let's introduce the solutions of these two estimators.