Joint Gaussian Graphical Model Review Series – III

Markov Random Field and Log Linear Model

Beilun Wang

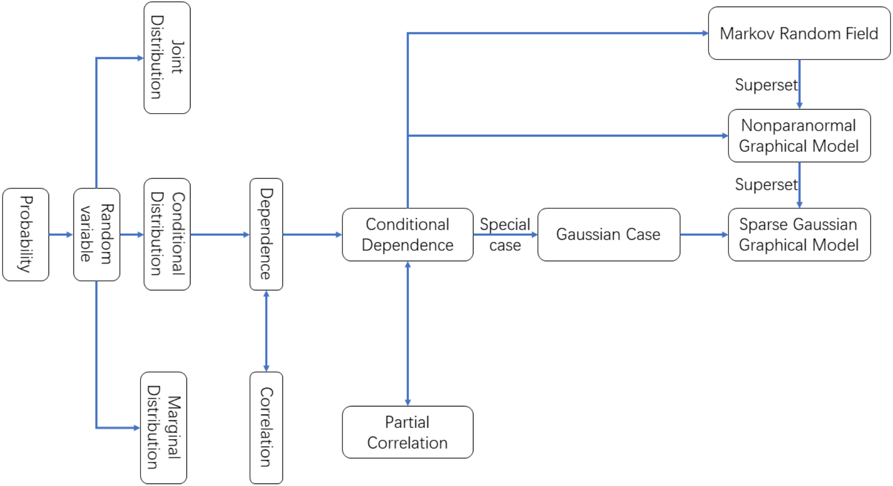
[1]Department of Computer Science, University of Virginia
http://jointggm.org/

July 7th, 2017

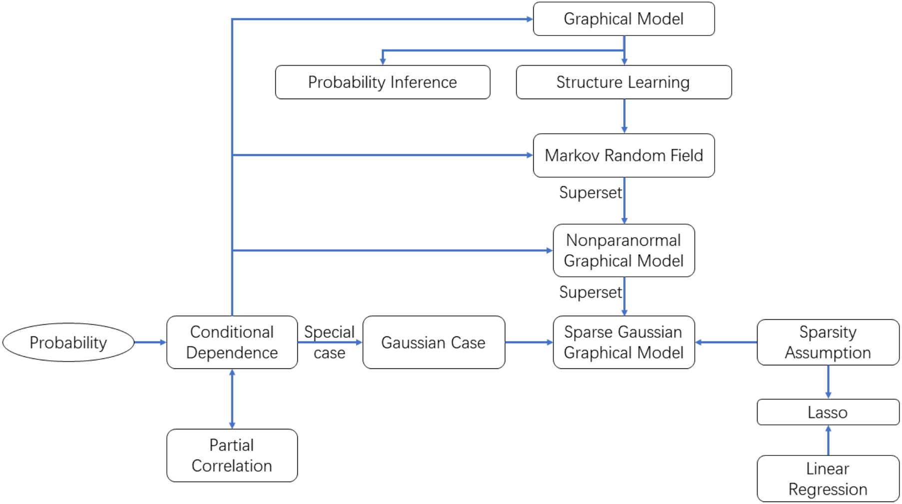
Outline

1. Why we need Graphical Model?
2. Graphical Model
3. Markov Random Field

Road Map


Probability
Random variable
Joint Distribution
Conditional Distribution
Marginal Distribution
Dependence
Correlation
Conditional Dependence
Partial Correlation
Special case
Gaussian Case
Markov Random Field
Superset
Nonparanormal Graphical Model
Superset
Sparse Gaussian Graphical Model

Road Map

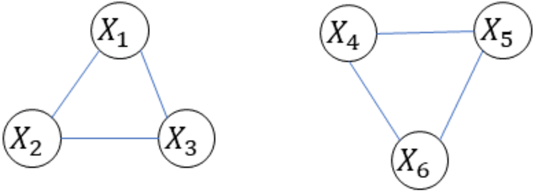

Graphical Model
Probability Inference
Structure Learning
Markov Random Field
Superset
Nonparanormal Graphical Model
Superset
Probability
Conditional Dependence
Special case
Gaussian Case
Sparse Gaussian Graphical Model
Sparsity Assumption
Lasso
Linear Regression
Partial Correlation

Review: Gaussian Case

- In the Gaussian case, we know the conditional dependence and partial correlation are equivalent.
- This pairwise relationship can be naturally represented by a graph $G = (V, E)$.
- $|\Omega| > 0$ is a natural adjacency matrix.
- We call the pairwise conditional dependence relationship among variables as undirected Graphical Model.

Why we need Graphical Model?

A Toy Example



A Toy Example

Suppose $X = (X_1, X_2, X_3, X_4, X_5, X_6)$. Each variable only takes either 0 or 1. To estimate the joint probability $p(X)$, you need to estimate 2^6 values. However, if we know the conditional independence graph, $p(X) = p(X_1, X_2, X_3)p(X_4, X_5, X_6)$. You only need to estimate 2^4 values.

Proof of the decomposition

First, let's prove that if $X_1 \perp\!\!\!\perp X_3|X_2$, then $p(X_1|X_3, X_2) = p(X_1|X_2)$.
$p(X_1|X_2)p(X_3|X_2) = p(X_1, X_3|X_2) = p(X_1|X_3, X_2)p(X_3|X_2)$. Cancel out $p(X_3|X_2)$ in the both sides, we can have the conclusion.
It is easy to obtain the similar result under the local markov property: $p(X_v|X_{v\setminus N(v)}, X_{N(v)}) = p(X_v|X_{N(v)})$.

Proof of the decomposition

$$p(X_1, X_2, X_3, X_4, X_5, X_6) = p(X_1|X_2, X_3, X_4, X_5, X_6)p(X_2|X_3, X_4, X_5, X_6)p(X_3|$$

By the conclusion we have in the last page, the left equals to

$$p(X_1|X_2, X_3)p(X_2|X_3)p(X_3)p(X_4, X_5, X_6) \tag{1.1}$$

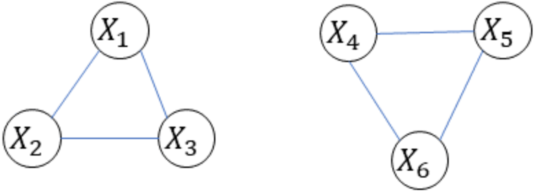
$$=p(X_1, X_2, X_3)p(X_4, X_5, X_6) \tag{1.2}$$

Graphical Model

Graphical Model

- **Probability Inference:** estimate joint probability, marginal probability, and conditional probability.
- **Structure learning:** Give dataset $\mathbf{X}$, learn the Graph structure from $\mathbf{X}$ (i.e., learn the edge patterns between variables).

A Toy Example



Probability Inference: Calculate the joint Probability

You know that $p(X) = p(X_1, X_2, X_3)p(X_4, X_5, X_6)$. Traditionally,
$p(X_1, X_2 = a) = \sum_{X_3, X_4, X_5, X_6} p(X_1, X_2 = a, X_3, X_4, X_5, X_6)$.

16 operators.

By the graph, we can have
$p(X_1, X_2 = a) = \sum_{X_3} p(X_1, X_2 = a, X_3) \sum_{X_4, X_5, X_6} p(X_4, X_5, X_6)$.

10 operators.

Probability Inference: Calculate the joint Probability

You know that $p(X) = p(X_1, X_2, X_3)p(X_4, X_5, X_6)$. Traditionally,
$p(X_1, X_2 = a) = \sum_{X_3, X_4, X_5, X_6} p(X_1, X_2 = a, X_3, X_4, X_5, X_6)$.

16 operators.

By the graph, we can have
$p(X_1, X_2 = a) = \sum_{X_3} p(X_1, X_2 = a, X_3) \sum_{X_4, X_5, X_6} p(X_4, X_5, X_6)$.

10 operators.

Markov Random Field

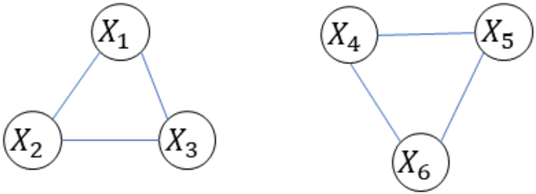
Markov Random Field

Markov Random Field

Given an undirected graph $G = (V, E)$, a set of random variables $X = (X_v)_{v \in V}$ indexed by V form a Markov random field with respect to G if they satisfy the local Markov property:
A variable is conditionally independent of all other variables given its neighbors: $X_v \perp\!\!\!\perp X_{V \setminus N(v)} | X_{N(v)}$

This property is stronger than the pairwise Markov property:
Any two non-adjacent variables are conditionally independent given all other variables: $X_u \perp\!\!\!\perp X_v \mid X_{V \setminus \{u,v\}} \quad \text{if } \{u, v\} \notin E.$

A Toy Example



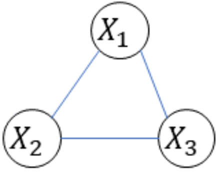
Clique factorization

If this joint density can be factorized over the cliques of G:

$$p(X = x) = \prod_{C \in \text{cl}(G)} \phi_C(x_C)$$

then X forms a Markov random field with respect to G. Here, $\text{cl}(G)$ is the set of cliques of G.

A Toy Example



Log-linear Model

Any Markov random field can be written as log-linear model with feature functions f_k such that the full-joint distribution can be written as:

$$P(X = x) = \frac{1}{Z} \exp\left(\sum_k w_k^\top f_k(X)\right)$$

. Notice that the reverse doesn't hold.

Example I: Pairwise Model

Pairwise Model

$$P(X = x) = \frac{1}{Z(\Theta)} \exp\left(\sum_{s\in V} \theta_s^\top x_s^2 + \sum_{(s,t)\in E} \theta_{st}^\top x_s x_t\right)$$

.

Examples:

- Gaussian Graphical Model
- Ising Model

These two models have good estimators to infer the MRF. Generally, estimate Θ is difficult. Since it involves computing $Z(\Theta)$ or its derivatives.

Example I: Pairwise Model – Gaussian Case

Gaussian Case

$$f(x_1, \ldots, x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right)}{\sqrt{(2\pi)^k|\Sigma|}}$$

.

Solution:

$$\ln \mathcal{L}(\bar{x}, \Omega) \propto \ln \det(\Omega) - \operatorname{tr}\left(\Omega \frac{1}{n} \sum_{i=1}^{n} (\bar{x}-\mu)(\bar{x}-\mu)^T\right) \quad (3.1)$$

$$= \ln \det(\Omega) - \operatorname{tr}\left(\Omega \widehat{S}\right) \quad (3.2)$$

where $\widehat{S}$ is the sample covariance matrix.
For the Ising model, we use generalized covariance matrix to avoid the normalization term.

Example II: Non-pairwise model – Nonparanormal Graphical Model

Are there any non-pairwise model which is easy to estimate?

Nonparanormal Graphical Model

$$P(X = x) = \frac{1}{Z} \exp\left(-\frac{1}{2}(f(x) - \mu)^T \Sigma^{-1} (f(x) - \mu)\right)$$

.

where $f(X) = (f_1(X_1), f_2(X_2), \ldots f_p(X_p))$ and each f_i is a univariate monotone function. $f(X) \sim N(\mu, \Sigma)$.

Summary

- The formal definition of Markov Random Field (undirected Graphical Model)
- General formulation: Clique factorization
- log-linear Model
- Two examples: pairwise model and nonparanormal Graphical Model.
- In the next talk, let's introduce the solutions of these two estimators for sGGM.