Graphical Models

Franz Pernkopf

Signal Processing and Speech Communication Laboratory
Graz University of Technology
Austria

Outline

Graphical Models (GMs)

"Graphical models are a marriage between probability theory and graph theory. They provide a natural tool for dealing with two problems that occur throughout applied mathematics and engineering – uncertainty and complexity – ..." [Jordan, 1999]

Directed GMs

Directed GMs: Bayesian networks

- Represent a joint distribution P over some set of random variables $\boldsymbol{Z} = \{Z_1, \ldots, Z_N\}$.
- Explicit representation of P is hard.
- A Bayesian network is a directed acyclic graph $\mathcal{G} = (\boldsymbol{Z}, \boldsymbol{E})$ which represents factorization properties of the distribution.
- Each node Z_j is represented as conditional distribution given its parents Z_{Π_j}, i.e. $p(Z_j | Z_{\Pi_j})$.
- Joint distribution:

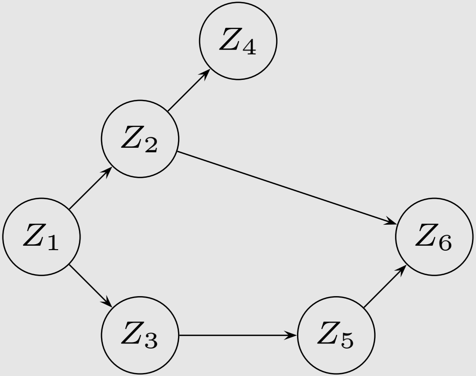
$$P(\boldsymbol{Z}) = \prod_{j=1}^{N} P(Z_j | Z_{\Pi_j})$$

- Application: Hidden Markov model, expert systems, ...

Example

- $P(Z_{1:6}) = P(Z_1)P(Z_2|Z_1)P(Z_3|Z_1)P(Z_4|Z_2)P(Z_5|Z_3)P(Z_6|Z_2, Z_5)$

Graph



Conditional independence

Definition: d-separation [Pearl, 1988]

Z_i and Z_j $(i \neq j)$ are d-separated if for all paths between Z_i and Z_j there is an intermediate variable Z_k $(i \neq j \neq k)$ such that

- the connection is serial or diverging and the state of Z_k is known.
- the connection is converging and neither the state of Z_k nor the state of any descendant of Z_k is known.

Canonical examples

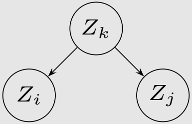
Serial connection

$$Z_i \longrightarrow Z_k \longrightarrow Z_j$$

- $Z_i \perp Z_j | Z_k$
- $P(Z_i, Z_k, Z_j) = P(Z_i)P(Z_k|Z_i)P(Z_j|Z_k)$
- $$P(Z_j|Z_i, Z_k) = \frac{P(Z_i, Z_k, Z_j)}{P(Z_i, Z_k)} = \frac{P(Z_i)P(Z_k|Z_i)P(Z_j|Z_k)}{P(Z_i)P(Z_k|Z_i)} = P(Z_j|Z_k)$$

Canonical examples

Diverging connection

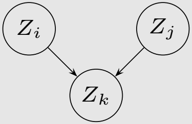


- $Z_i \perp Z_j | Z_k$
- $P(Z_i, Z_k, Z_j) = P(Z_k)P(Z_i|Z_k)P(Z_j|Z_k)$
-

$$P(Z_j, Z_i|Z_k) = \frac{P(Z_i, Z_k, Z_j)}{P(Z_k)} = \frac{P(Z_k)P(Z_i|Z_k)P(Z_j|Z_k)}{P(Z_k)} = P(Z_i|Z_k)P(Z_j|Z_k)$$

Canonical examples

Converging connection



- $Z_i \perp Z_j$
- $P(Z_i, Z_k, Z_j) = P(Z_k | Z_i, Z_j) P(Z_i) P(Z_j)$
- $$P(Z_j, Z_i) = \sum_{Z_k} P(Z_i, Z_k, Z_j) = P(Z_i) P(Z_j)$$
- $$P(Z_j, Z_i | Z_k) = \frac{P(Z_i, Z_k, Z_j)}{P(Z_k)} = \frac{P(Z_k | Z_i, Z_j) P(Z_i) P(Z_j)}{P(Z_k)}$$
- *Explaining away* phenomenon

Undirected GMs

Undirected GMs: Markov networks

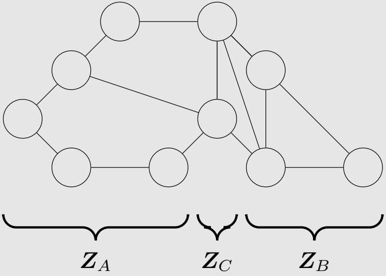
- Represent a joint distribution P over some set of random variables $\boldsymbol{Z} = \{Z_1, \ldots, Z_N\}$.
- A Markov network is an undirected graph $\mathcal{G} = (\boldsymbol{Z}, \boldsymbol{E})$ which represents factorization properties of the distribution.
- Application: Markov random field (image segmentation/denoising), Conditional random field, Ising model, ...

Conditional independence

Definition

Any two subsets of variables are conditionally independent given a separating subset: $\boldsymbol{Z}_A \perp \boldsymbol{Z}_B | \boldsymbol{Z}_C$, where every path from a node in set A to a node in set B passes through set C.

Example



Factorization of joint distribution

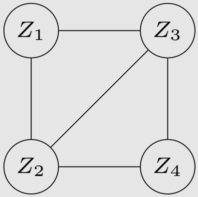
Definition: Clique

A clique C is a subset of nodes $\boldsymbol{Z}_C$ in $\mathcal{G}$ such that there exists an edge between all pairs of nodes in the subset.

Definition: Maximal clique

A maximal clique $\tilde{C}$ is a clique C such that adding any other node in the graph makes it no longer a clique.

Example



Factorization of joint distribution

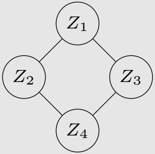
- Joint distribution is a product of potential functions $\Psi_{\tilde{C}}(\boldsymbol{Z}_{\tilde{C}})$ over maximal cliques of $\mathcal{G}$

$$P(\boldsymbol{Z}) = \frac{1}{W} \prod_{\tilde{C}} \Psi_{\tilde{C}}(\boldsymbol{Z}_{\tilde{C}}).$$

- Partition function: $W = \sum_{\boldsymbol{Z}} \prod_{\tilde{C}} \Psi_{\tilde{C}}(\boldsymbol{Z}_{\tilde{C}})$

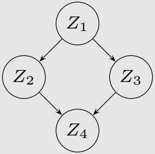
Examples

Some UGMs can not be represented by DGMs



$$Z_1 \perp Z_4 | \{Z_2, Z_3\}$$

$$Z_2 \perp Z_3 | \{Z_1, Z_4\}$$

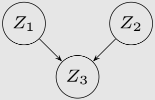


$$Z_1 \perp Z_4 | \{Z_2, Z_3\}$$

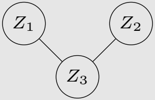
$$Z_2 \perp Z_3 | \{Z_1\}$$

$$Z_2 \not\perp Z_3 | \{Z_1, Z_4\}$$

Some DGMs can not be represented by UGMs



$$Z_1 \perp Z_2$$

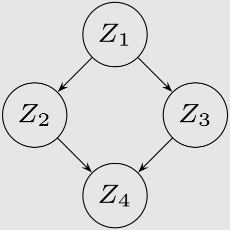


$$Z_1 \perp Z_2 | Z_3$$

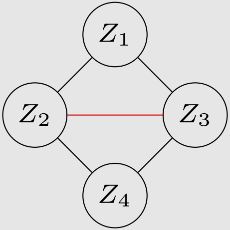
Convert DGM to UGM

- Add edges between all pairs of parents of each node Z_i (*moralization*).
- Drop the arrows on the original edges.
- Initialize each clique potential to 1.
- Multiply each conditional distribution of the DGM into one of the clique potentials.

Example



Directed GM



Undirected GM (*Moral Graph*)

Factor Graphs

- Undirected GMs and directed GMs can be formulated as *factor graph*.
- Aim: Capturing factorizations between variables.
- A factor graph is a bipartite undirected graph.
- Consists of a set of $\mathcal{F}$ factor nodes $F_j(\cdot) \in \mathcal{F}$ and variables nodes $\boldsymbol{Z} = \{Z_1, \ldots, Z_N\}$.
- Each factor $F_j(\cdot)$ depends on a subset of variable nodes $\boldsymbol{Z}_j \subseteq \boldsymbol{Z}$.
- Joint distribution is a product of factor nodes $F_j(\cdot)$

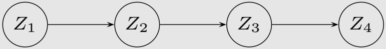
$$P(\boldsymbol{Z}) = \frac{1}{W} \prod_{j=1}^{|\mathcal{F}|} F_j(\boldsymbol{Z}_j).$$

- Normalization constant: $W = \sum_{\boldsymbol{Z}} \prod_{j=1}^{|\mathcal{F}|} F_j(\boldsymbol{Z}_j)$

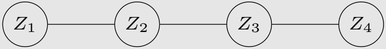
Example: Markov chain

- $P(Z_{1:4}) = P(Z_1)P(Z_2|Z_1)P(Z_3|Z_2)P(Z_4|Z_3)$

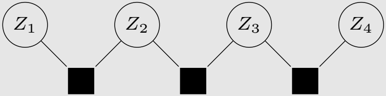
Directed GM



Undirected GM



Factor graph



$F_1(Z_1, Z_2)$ $F_2(Z_2, Z_3)$ $F_3(Z_3, Z_4)$

Research challenges for GMs:

- Learning
- Inference