Graphical model formalism, factorization properties and conditional independance.

Guillaume Obozinski

Swiss Data Science Center

African Masters of Machine Intelligence, 2018-2019, AIMS, Kigali

Outline

Independence concepts

Independence: $X \perp\!\!\!\perp Y$

We say that X et Y are independents and write $X \perp\!\!\!\perp Y$ ssi:

$$\forall x, y, \qquad P(X = x, Y = y) = P(X = x)\, P(Y = y)$$

Independence concepts

Independence: $X \perp\!\!\!\perp Y$

We say that X et Y are independents and write $X \perp\!\!\!\perp Y$ ssi:

$$\forall x, y, \qquad P(X = x, Y = y) = P(X = x)\, P(Y = y)$$

Conditional Independence: $X \perp\!\!\!\perp Y \mid Z$

- On says that X and Y are independent conditionally on Z and
- write $X \perp\!\!\!\perp Y \mid Z$ iff:

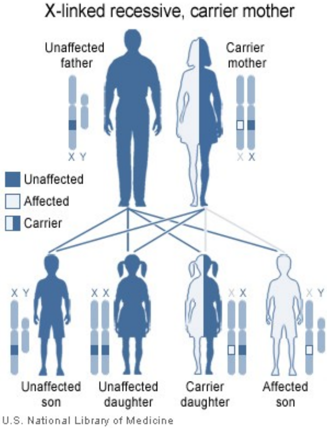
$\forall x, y, z,$

$$P(X = x, Y = y \mid Z = z) = P(X = x|Z = z)\, P(Y = y|Z = z)$$

Conditional Independence exemple

Example of
"X-linked recessive inheritance":

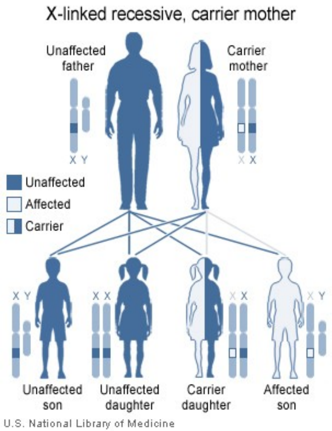
Transmission of the gene
responsible for hemophilia



Conditional Independence exemple

Example of
“X-linked recessive inheritance”:

Transmission of the gene
responsible for hemophilia



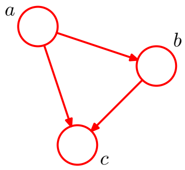
Risk for sons from an unaffected father:

- dependance between the situation of the two brothers.
- conditionally independent given that the mother is a carrier of the gene or not.

Outline

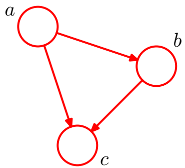
Directed graphical model or Bayesian network

$$p(a, b, c) = p(a)\, p(b|a)\, p(c|b, a)$$



Directed graphical model or Bayesian network

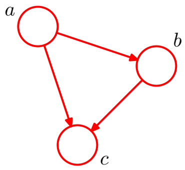
$$p(a, b, c) = p(a)\, p(b|a)\, p(c|b, a)$$



$$p(x_1, x_2) = p(x_1)p(x_2)$$

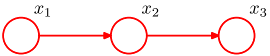
Directed graphical model or Bayesian network

$$p(a, b, c) = p(a)\, p(b|a)\, p(c|b, a)$$



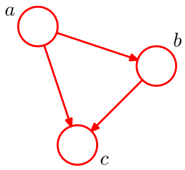
$$p(x_1, x_2) = p(x_1)p(x_2)$$

$$p(x_1, x_2, x_3) = p(x_1)p(x_2|x_1)p(x_3|x_2)$$



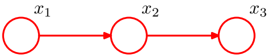
Directed graphical model or Bayesian network

$$p(a, b, c) = p(a)\, p(b|a)\, p(c|b, a)$$

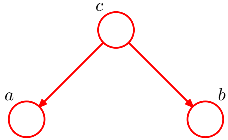


$$p(x_1, x_2) = p(x_1)p(x_2)$$

$$p(x_1, x_2, x_3) = p(x_1)p(x_2|x_1)p(x_3|x_2)$$



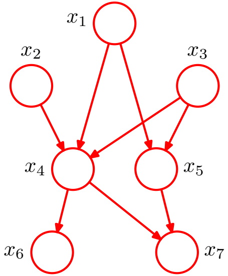
$$a \perp\!\!\!\perp b \mid c$$



Factorization according to a directed graph

Let Π_j denote the set of parents of node j.

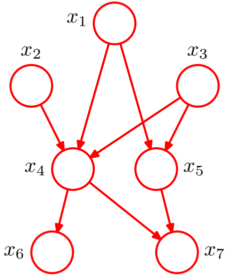
$$\prod_{j=1}^{p} p(x_j | x_{\Pi_j})$$



Factorization according to a directed graph

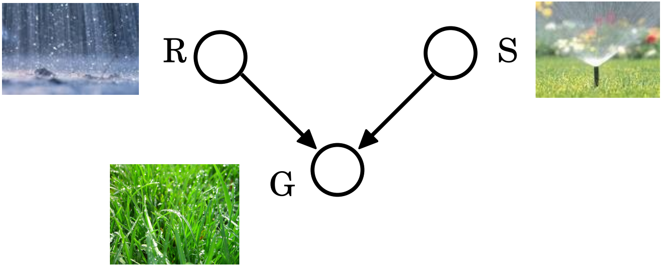
Let Π_j denote the set of parents of node j.

$$\prod_{j=1}^{p} p(x_j | x_{\Pi_j})$$



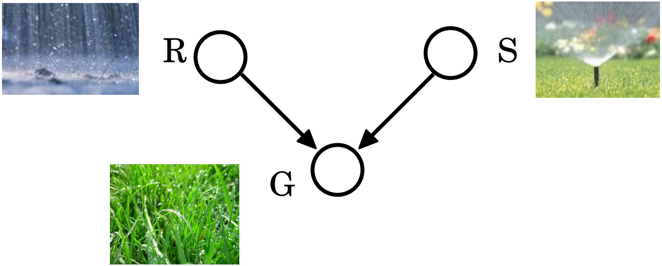
$$p(x_1) \prod_{j=2}^{M} p(x_j | x_{j-1})$$

The Sprinkler



- $R = 1$: it has rained
- $S = 1$: the sprinkler worked
- $G = 1$: the grass is wet

The Sprinkler



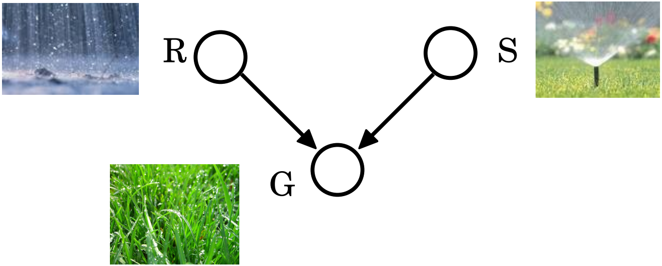
- $R = 1$: it has rained
- $S = 1$: the sprinkler worked
- $G = 1$: the grass is wet

$$P(S = 1) = 0.5$$

$$P(R = 1) = 0.2$$

$P(G = 1 \mid S, R)$	R=0	R=1
S=0	0.01	0.8
S=1	0.8	0.95

The Sprinkler



$$P(S = 1) = 0.5$$

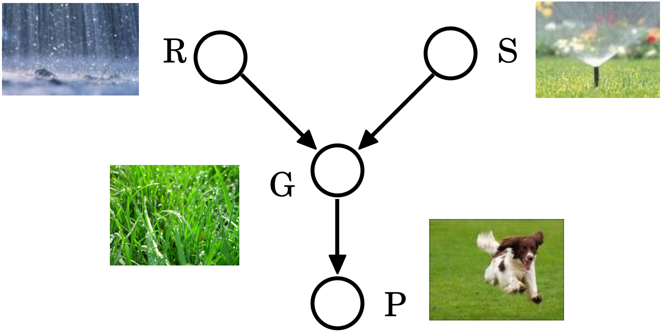
$$P(R = 1) = 0.2$$

- $R = 1$: it has rained
- $S = 1$: the sprinkler worked
- $G = 1$: the grass is wet

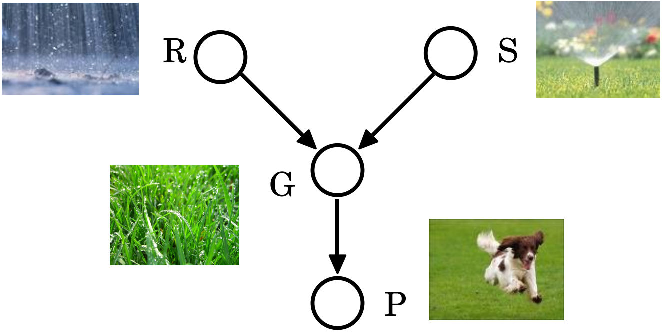
$P(G = 1 \mid S, R)$	R=0	R=1
S=0	0.01	0.8
S=1	0.8	0.95

- Given that we observe that the grass is wet, are R and S independent?

The Sprinkler II



The Sprinkler II



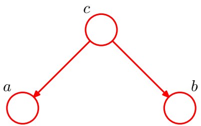
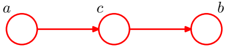
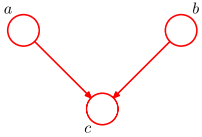
- $R = 1$: it has rained
- $S = 1$: the sprinkler worked
- $G = 1$: the grass is wet
- $P = 2$: the paws of the dog are wet

$P(S = 1) = 0.5 \quad P(R = 1) = 0.2$

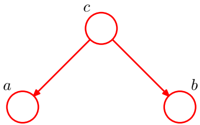
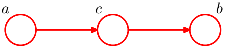
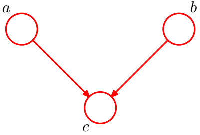
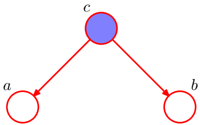
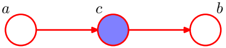
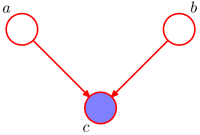
$P(G = 1 \mid S, R)$	R=0	R=1
S=0	0.01	0.8
S=1	0.8	0.95

$P(P = 1 \mid G)$	G=0	G=1
	0.2	0.7

Blocking nodes

diverging edges	head-to-tail	converging edges
		
		$\not\leftrightarrow$
$a \not\perp\!\!\!\perp b$	$a \not\perp\!\!\!\perp b$	$a \perp\!\!\!\perp b$

Blocking nodes

diverging edges	head-to-tail	converging edges
		
		$\not\leftrightarrow$
$a \not\perp\!\!\!\perp b$	$a \not\perp\!\!\!\perp b$	$a \perp\!\!\!\perp b$
		
$\not\leftrightarrow$	$\not\leftrightarrow$	
$a \perp\!\!\!\perp b \mid c$	$a \perp\!\!\!\perp b \mid c$	$a \not\perp\!\!\!\perp b \mid c$

The configuration with converging edges is called a v-structure

Factorization and Independence

A factorization imposes independence statements

Proposition

$$\forall x,\ p(x) = \prod_{j=1}^{p} p(x_j | x_{\Pi_j}) \quad \Leftrightarrow \quad \forall j,\ X_j \perp\!\!\!\perp X_{\{1,\ldots,j-1\}\backslash\Pi_j} \mid X_{\Pi_j}$$

Factorization and Independence

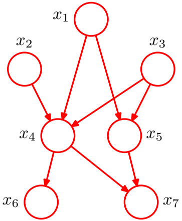
A factorization imposes independence statements

Proposition

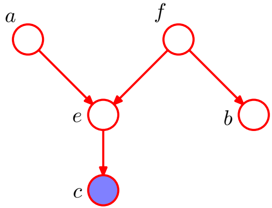
$$\forall x,\ p(x) = \prod_{j=1}^{p} p(x_j | x_{\Pi_j}) \quad \Leftrightarrow \quad \forall j,\ X_j \perp\!\!\!\perp X_{\{1,\ldots,j-1\}\setminus\Pi_j} \mid X_{\Pi_j}$$

Is it possible to read from the graph the (conditional) independence statements that hold given the factorization.

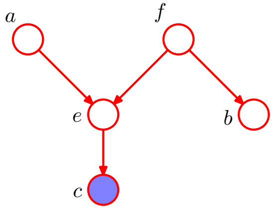
$$X_5 \overset{?}{\perp\!\!\!\perp} X_2 \mid X_4$$



d-separation



d-separation



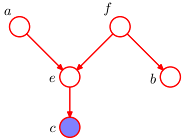
Theorem

If A, B and C are three disjoint sets of node, the statement $X_A \perp\!\!\!\perp X_B \mid X_S$ holds if all trails joining A to B go through at least one *blocking node*.

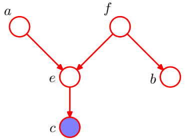
A node j is blocking a trail

- if the edges of the trails are diverging/following and $j \in S$
- if the edges of the trails are converging (i.e. form a v-structure) and neither j nor any of its descendants is in S

d-separation: Restatement in terms of observed node



d-separation: Restatement in terms of observed node



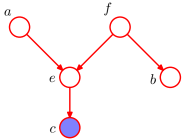
Theorem

If A, B and C are three disjoint sets of nodes, and we call C the set of *observed nodes*. Then the statement $X_A \perp\!\!\!\perp X_B \mid X_S$ holds if all trails joining A to B are blocked.
A trail is blocked if none of the regular nodes[a] are observed, and if all nodes with a v-structure on the trail are observed themselves or have a descendant which is observed.

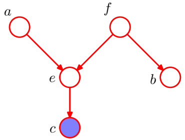
- observed themselves
- have a descendant which is observed.

[a] A "regular" node is a node without v-structure

Conditional independence for non-disjoint sets



Conditional independence for non-disjoint sets



Proposition

If A, B and C are three sets of nodes of a graph $G = (V, E)$. And if X_V satisfies the Markov Property w.r.t. G,

$$\text{then we have} \qquad X_A \perp\!\!\!\perp X_B \mid X_S$$

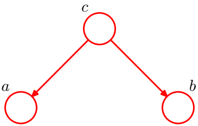
$$\text{if} \quad \begin{cases} A \cap B \subset S, \\ X_{A \setminus S} \perp\!\!\!\perp X_{B \setminus S} \mid X_S. \end{cases}$$

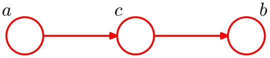
Factorization et Independence II

- Several graphs can induce the same set of conditional independences .

Factorization et Independence II

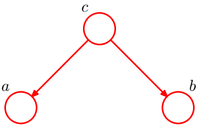
- Several graphs can induce the same set of conditional independences .

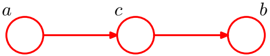




Factorization et Independence II

- Several graphs can induce the same set of conditional independences .

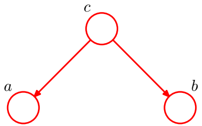


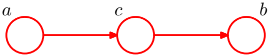


$$p(c)p(a|c)p(b|c)$$

Factorization et Independence II

- Several graphs can induce the same set of conditional independences .

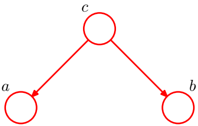


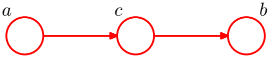


$$p(c)p(a|c)p(b|c) = p(a)p(c|a)p(b|c)$$

Factorization et Independence II

- Several graphs can induce the same set of conditional independences .



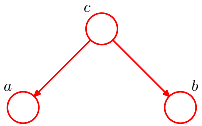


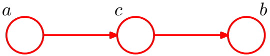
$$p(c)p(a|c)p(b|c) = p(a)p(c|a)p(b|c)$$

- Some combinations of conditional independences cannot be faithfully represented by a graphical model

Factorization et Independence II

- Several graphs can induce the same set of conditional independences .



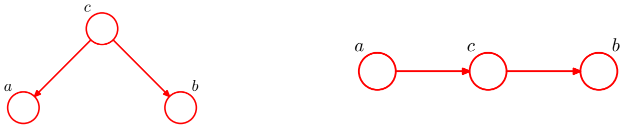


$$p(c)p(a|c)p(b|c) = p(a)p(c|a)p(b|c)$$

- Some combinations of conditional independences cannot be faithfully represented by a graphical model
  - Ex1: $X \sim \mathrm{Ber}\frac{1}{2} \qquad Y \sim \mathrm{Ber}\frac{1}{2} \qquad Z = X \oplus Y.$

Factorization et Independence II

- Several graphs can induce the same set of conditional independences .

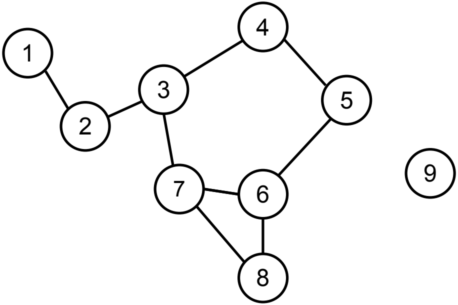


$$\color{blue}{p(c)p(a|c)}\color{black}{p(b|c)} = \color{blue}{p(a)p(c|a)}\color{black}{p(b|c)}$$

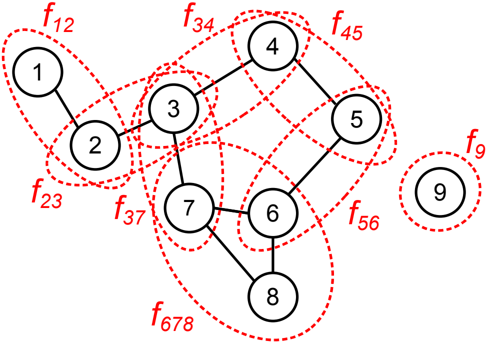
- Some combinations of conditional independences cannot be faithfully represented by a graphical model
  - Ex1: $X \sim \mathrm{Ber}\frac{1}{2} \qquad Y \sim \mathrm{Ber}\frac{1}{2} \qquad Z = X \oplus Y.$
  - Ex2: $X \perp\!\!\!\perp Y \mid Z = 1$ but $X \not\perp\!\!\!\perp Y \mid Z = 0$

Outline

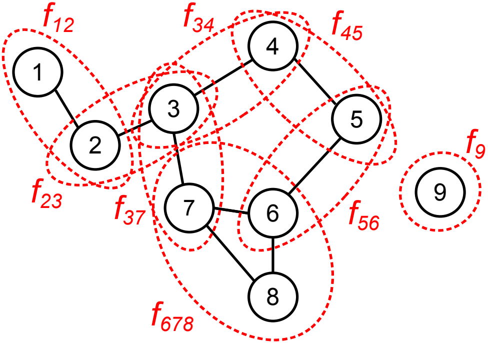
Undirected graphical model



Undirected graphical model



Undirected graphical model



$$p(x_1, x_2, \ldots, x_9) = f_{12}(x_1, x_2)\, f_{23}(x_2, x_3)\, f_{34}(x_3, x_4)\, f_{45}(x_4, x_5) \ldots$$
$$f_{56}(x_5, x_6)\, f_{37}(x_3, x_7)\, f_{678}(x_6, x_7, x_8)\, f_9(x_9)$$

Gibbs distribution

Clique Set of nodes that are all connected to one another.

Gibbs distribution

Clique Set of nodes that are all connected to one another.

Potential function The potential $\psi_C(x_C) \geq 0$ is associated to clique C.

Gibbs distribution

Clique Set of nodes that are all connected to one another.

Potential function The potential $\psi_C(x_C) \geq 0$ is associated to clique C.

Gibbs distribution

$$p(x) = \frac{1}{Z} \prod_C \psi_C(x_C)$$

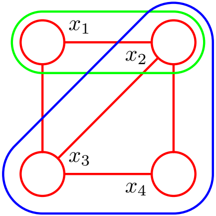
Gibbs distribution

Clique Set of nodes that are all connected to one another.

Potential function The potential $\psi_C(x_C) \geq 0$ is associated to clique C.

Gibbs distribution

$$p(x) = \frac{1}{Z} \prod_C \psi_C(x_C)$$



Partition function: Z

$$Z = \sum_x \prod_C \psi_C(x_C)$$

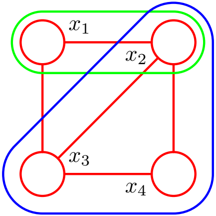
Gibbs distribution

Clique Set of nodes that are all connected to one another.

Potential function The potential $\psi_C(x_C) \geq 0$ is associated to clique C.

Gibbs distribution

$$p(x) = \frac{1}{Z} \prod_C \psi_C(x_C)$$



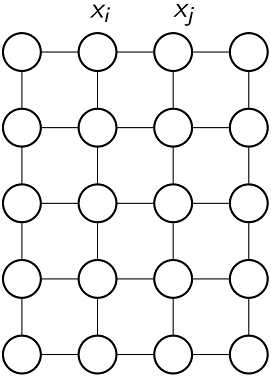
Partition function: Z

$$Z = \sum_x \prod_C \psi_C(x_C)$$

Writing potential in exponential form $\psi_C(x_C) = \exp\{-E(x_C)\}$.
$E(x_C)$ is an *energy*.
This a *Boltzmann distribution*.

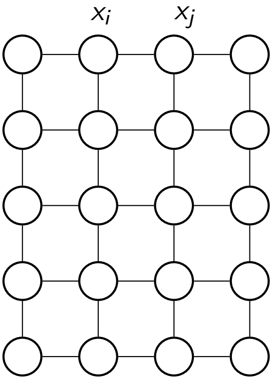
Example 1: Ising model

$X = (X_1, \dots, X_d)$ is a collection of binary variables.



Example 1: Ising model

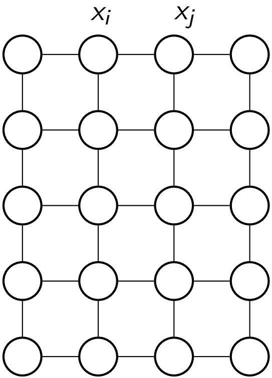
$X = (X_1, \ldots, X_d)$ is a collection of binary variables.



$$p(x_1, \ldots, x_d) = \frac{1}{Z(\eta)} \exp\Big(\sum_{i \in V} \eta_i x_i + \sum_{\{i,j\} \in E} \eta_{ij} x_i x_j\Big)$$

Example 1: Ising model

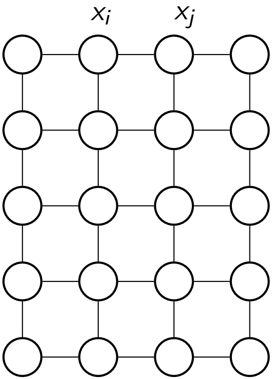
$X = (X_1, \ldots, X_d)$ is a collection of binary variables.



$$
\begin{aligned}
& p(x_1, \ldots, x_d) \\
& \quad = \frac{1}{Z(\eta)} \exp\Big(\sum_{i \in V} \eta_i x_i + \sum_{\{i,j\} \in E} \eta_{ij} x_i x_j\Big) \\
& \quad = \frac{1}{Z(\eta)} \prod_{i \in V} e^{\eta_i x_i} \prod_{\{i,j\} \in E} e^{\eta_{ij} x_i x_j}
\end{aligned}
$$

Example 1: Ising model

$X = (X_1, \ldots, X_d)$ is a collection of binary variables.



$$
\begin{aligned}
p(x_1, \ldots, x_d) & \\
&= \frac{1}{Z(\eta)} \exp\Big(\sum_{i \in V} \eta_i x_i + \sum_{\{i,j\} \in E} \eta_{ij} x_i x_j \Big) \\
&= \frac{1}{Z(\eta)} \prod_{i \in V} e^{\eta_i x_i} \prod_{\{i,j\} \in E} e^{\eta_{ij} x_i x_j} \\
&= \frac{1}{Z(\eta)} \prod_{i \in V} \psi_i(x_i) \prod_{\{i,j\} \in E} \psi_i(x_i, x_j)
\end{aligned}
$$

with $\psi_i(x_i) = e^{\eta_i x_i}$ and $\psi_{ij}(x_i, x_j) = e^{\eta_{ij} x_i x_j}$.

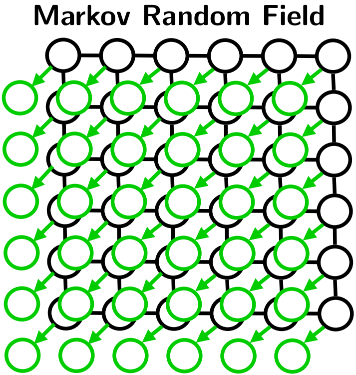
Example 2: Directed graphical model

Consider a distribution p that factorizes according to a directed graph $G = (V, E)$, then

$$
\begin{aligned}
p(x_1, \ldots, x_d) &= \prod_{i=1}^{d} p(x_i \mid x_{\pi_i}) \\
&= \prod_{i=1}^{d} \psi_{C_i}(x_{C_i}) \qquad \text{with} \quad C_i = \{i\} \cup \pi_i
\end{aligned}
$$

Consequence: A distribution that factorizes according to a directed model is a Gibbs distribution for the cliques $C_i = \{i\} \cup \pi_i$. As a consequence, it factorizes according to an undirected graph in which C_i are cliques.

Modeling image structures

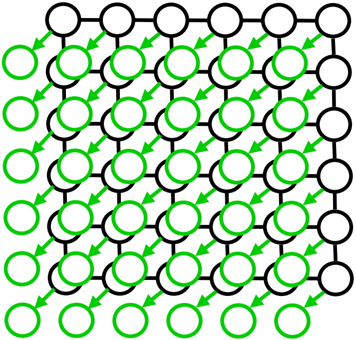


Original image

Segmentation

Modeling image structures

Markov Random Field



Original image

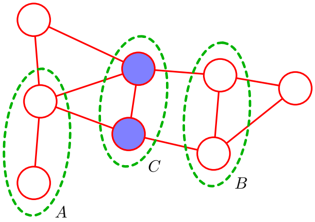
Segmentation

$\rightarrow$ *directed graphical model* vs *undirected*

Global Markov Property or *Undirected graphical model*

We say that a probability distribution p satisfies the *global Markov property* for the graph $G = (V, E)$, if for all $A, B, S \subset V$

$$S \text{ separates } A \text{ from } B \text{ in the graph} \Rightarrow X_A \perp\!\!\!\perp X_B \mid X_S$$



Theorem of Hammersley and Clifford (1971)

A distribution p, which is such that $p(x) > 0$ for all x satisfies the *global Markov property* for graph G if and only if it is a Gibbs distribution associated with G.

- Gibbs distribution: $\mathcal{P}_G : \; p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}_G} \psi_C(x_C)$
- Global Markov property:

$$\mathcal{P}_M : \; X_A \perp\!\!\!\perp X_B \mid X_C \quad \text{if} \quad C \text{ separated } A \text{ and } B \text{ in } G$$

Theorem

We have $\mathcal{P}_G \Rightarrow \mathcal{P}_M$ and (HC): if $\forall x, \; p(x) > 0$, then $\mathcal{P}_M \Rightarrow \mathcal{P}_G$

Markov Blanket in an undirected graph

Definition

The Markov Blanket B of a node i is the smallest set of nodes B such that

$$X_i \perp\!\!\!\perp X_R \mid X_B, \qquad \text{with} \quad R = V \backslash (B \cup \{i\})$$

Markov Blanket in an undirected graph

Definition

The Markov Blanket B of a node i is the smallest set of nodes B such that

$$X_i \perp\!\!\!\perp X_R \mid X_B, \qquad \text{with} \quad R = V \backslash (B \cup \{i\})$$

or equivalently such that

$$p(X_i \mid X_{-i}) = p(X_i \mid X_B).$$

Markov Blanket in an undirected graph

Definition

The Markov Blanket B of a node i is the smallest set of nodes B such that

$$X_i \perp\!\!\!\perp X_R \mid X_B, \qquad \text{with} \quad R = V \backslash (B \cup \{i\})$$

or equivalently such that

$$p(X_i \mid X_{-i}) = p(X_i \mid X_B).$$

- What is the Markov blanket of a node in an undirected graph?

Markov Blanket in an undirected graph

Definition

The Markov Blanket B of a node i is the smallest set of nodes B such that

$$X_i \perp\!\!\!\perp X_R \mid X_B, \qquad \text{with} \quad R = V \backslash (B \cup \{i\})$$

or equivalently such that

$$p(X_i \mid X_{-i}) = p(X_i \mid X_B).$$

- What is the Markov blanket of a node in an undirected graph?
- Answer:

Markov Blanket in an undirected graph

Definition

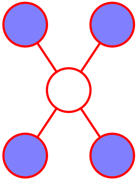
The Markov Blanket B of a node i is the smallest set of nodes B such that

$$X_i \perp\!\!\!\perp X_R \mid X_B, \qquad \text{with} \quad R = V \backslash (B \cup \{i\})$$

or equivalently such that

$$p(X_i \mid X_{-i}) = p(X_i \mid X_B).$$

- What is the Markov blanket of a node in an undirected graph?
- Answer:



Markov Blanket for a directed graph?

What is the Markov Blanket in a directed graph? By definition: the smallest set C of nodes such that conditionally on X_C, j is independent of all the other nodes in the graph?

Markov Blanket for a directed graph?

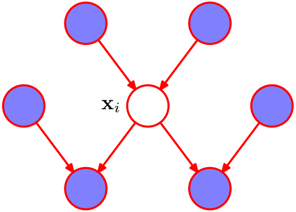
What is the Markov Blanket in a directed graph? By definition: the smallest set C of nodes such that conditionally on X_C, j is independent of all the other nodes in the graph?

- Answer:

Markov Blanket for a directed graph?

What is the Markov Blanket in a directed graph? By definition: the smallest set C of nodes such that conditionally on X_C, j is independent of all the other nodes in the graph?

- Answer:



Moralization

For a given oriented graphical model

- is there an unoriented graphical model which is equivalent?

Moralization

For a given oriented graphical model

- is there an unoriented graphical model which is equivalent?
- is there a smallest unoriented graphical which contains the oriented graphical model?

Moralization

For a given oriented graphical model

- is there an unoriented graphical model which is equivalent?
- is there a smallest unoriented graphical which contains the oriented graphical model?

$$p(x) = \frac{1}{Z} \prod_C \psi_C(x_C) \quad \text{vs} \quad \prod_{j=1}^{M} p(x_j | x_{\Pi_j})$$

Moralization

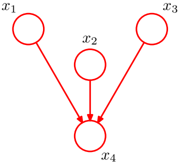
Given a directed graph G, its moralized graph G_M is obtained by

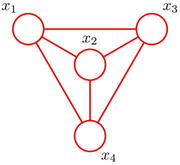
1. For any node i, add undirected edges between all its parents
2. Remove the orientation of all the oriented edges

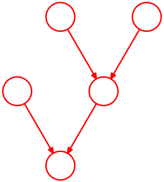
Moralization

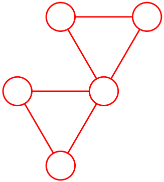
Given a directed graph G, its moralized graph G_M is obtained by

1. For any node i, add undirected edges between all its parents
2. Remove the orientation of all the oriented edges





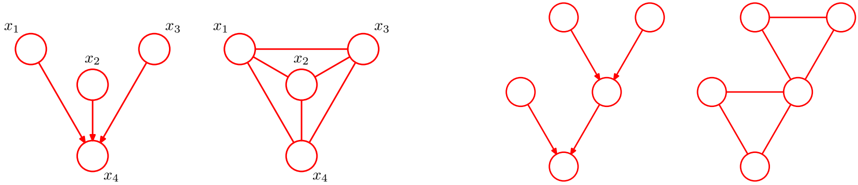




Moralization

Given a directed graph G, its moralized graph G_M is obtained by

1. For any node i, add undirected edges between all its parents
2. Remove the orientation of all the oriented edges



Proposition

If a probability distribution factorizes according to a directed graph G then it factorizes according to the undirected graph G_M.

Proof.

Write $p(x) := \prod_{i=1}^{n} p(x_i \mid x_{\pi_i}) = \prod_{i=1}^{n} \psi_{C_i}(x_{C_i})$ with $\begin{cases} C_i = \pi_i \cup \{i\} \\ \psi_{C_i}(x_{C_i}) = p(x_i \mid x_{\pi_i}). \end{cases}$

Directed vs undirected trees

Definition: directed tree

A directed tree is a DAG such that each node has at most one parent

Directed vs undirected trees

Definition: directed tree

A directed tree is a DAG such that each node has at most one parent

Remark: By definition a directed tree has no v-structure.

Directed vs undirected trees

Definition: directed tree

A directed tree is a DAG such that each node has at most one parent

Remark: By definition a directed tree has no v-structure.

Moralizing trees

- What is the moralized graph for a directed tree?
- The corresponding undirected tree!

Proposition (Equivalence between directed and undirected tree)

A distribution factorizes according to a directed tree if and only if it factorizes according to its undirected version.

Directed vs undirected trees

Definition: directed tree

A directed tree is a DAG such that each node has at most one parent

Remark: By definition a directed tree has no v-structure.

Moralizing trees

- What is the moralized graph for a directed tree?
- The corresponding undirected tree!

Proposition (Equivalence between directed and undirected tree)

A distribution factorizes according to a directed tree if and only if it factorizes according to its undirected version.

Corollary

All orientations of the edges of a tree that do not create v-structure are equivalent.