Identifying the Irreducible Disjoint Factors of a Multivariate Probability Distribution

PGM 2016

Maxime Gasse, Alex Aussem

Laboratoire LIRIS, UMR 5205 CNRS
Université Lyon 1

sept. 6, 2016

Introduction

- Structure learning
- Probabilistic graphical models
- Disjoint factorization

Graphical characterization

- Irreducible disjoint factors
- Conditional independence properties
- Connected components

Quadratic pairwise algorithms

- Under the Intersection assumption
- Under the Composition assumption
- For any probability distribution

Discussion

Introduction

Structure learning

$p(\mathbf{v})$ a joint distribution over $\mathbf{V} = \{V_1, \ldots, V_n\}$, and $\perp\!\!\!\perp_p$ the independence model

$$\mathbf{X} \perp\!\!\!\perp_p \mathbf{Y} | \mathbf{Z} \iff p(\mathbf{x}, \mathbf{y}, \mathbf{z}) p(\mathbf{z}) = p(\mathbf{x}, \mathbf{z}) p(\mathbf{y}, \mathbf{z}).$$

Structure learning

$p(\mathbf{v})$ a joint distribution over $\mathbf{V} = \{V_1, \ldots, V_n\}$, and $\perp\!\!\!\perp_p$ the independence model

$$\mathbf{X} \perp\!\!\!\perp_p \mathbf{Y} | \mathbf{Z} \iff p(\mathbf{x}, \mathbf{y}, \mathbf{z}) p(\mathbf{z}) = p(\mathbf{x}, \mathbf{z}) p(\mathbf{y}, \mathbf{z}).$$

PGM: a graph $\mathcal{G} = (\mathbf{V}, \mathcal{E})$ and a set of parameters θ.

- $\mathcal{G}$ defines an independence model $\perp\!\!\!\perp_{\mathcal{G}}$;
- θ defines a probability distribution that respects $\perp\!\!\!\perp_{\mathcal{G}}$.

Structure learning

$p(\mathbf{v})$ a joint distribution over $\mathbf{V} = \{V_1, \ldots, V_n\}$, and $\perp\!\!\!\perp_p$ the independence model

$$\mathbf{X} \perp\!\!\!\perp_p \mathbf{Y} | \mathbf{Z} \iff p(\mathbf{x}, \mathbf{y}, \mathbf{z}) p(\mathbf{z}) = p(\mathbf{x}, \mathbf{z}) p(\mathbf{y}, \mathbf{z}).$$

PGM: a graph $\mathcal{G} = (\mathbf{V}, \mathcal{E})$ and a set of parameters θ.

- $\mathcal{G}$ defines an independence model $\perp\!\!\!\perp_{\mathcal{G}}$;
- θ defines a probability distribution that respects $\perp\!\!\!\perp_{\mathcal{G}}$.

Structure learning: find $\mathcal{G}$ that respects $\perp\!\!\!\perp_{\mathcal{G}} \subseteq \perp\!\!\!\perp_p$. Ideally we want the simplest model, i.e. $\mathcal{G}$ as sparse as possible, $\perp\!\!\!\perp_{\mathcal{G}}$ as large as possible, or θ as constrained as possible.

Classical PGMs

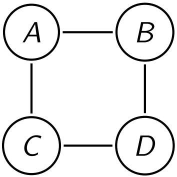
The graph $\mathcal{G} = (\mathbf{V}, \mathcal{E})$ encodes an independence model $\perp\!\!\!\perp_{\mathcal{G}} \subseteq \perp\!\!\!\perp_{p}$, induced by a factorization

$$p(\mathbf{v}) = \prod_{i=1}^{k} \phi_i(\mathbf{v}_i).$$

Classical PGMs

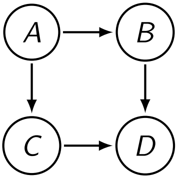
The graph $\mathcal{G} = (\mathbf{V}, \mathcal{E})$ encodes an independence model $\perp\!\!\!\perp_{\mathcal{G}} \subseteq \perp\!\!\!\perp_{p}$, induced by a factorization

$$p(\mathbf{v}) = \prod_{i=1}^{k} \phi_i(\mathbf{v}_i).$$



$A \perp\!\!\!\perp D \mid B, C$
$B \perp\!\!\!\perp C \mid A, D$

$\phi_1(a, b)\phi_2(b, c)\phi_3(c, d)\phi_4(c, a)$



$A \perp\!\!\!\perp D \mid B, C$
$B \perp\!\!\!\perp C \mid A$

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

Advanced PGMs

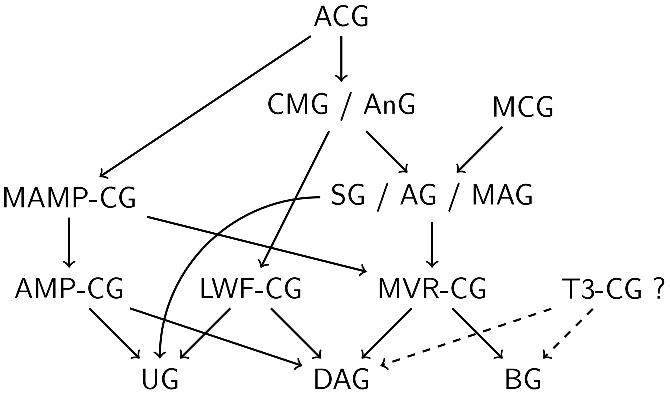


Figure: Hierarchy of PGM families by order of inclusion (in terms of expressive power as independence models).

[1] Chickering, Heckerman, and Meek, "Large-Sample Learning of Bayesian Networks is NP-Hard".

Advanced PGMs

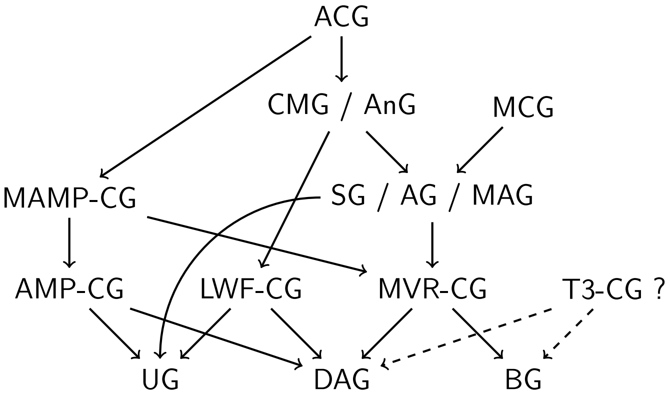


Figure: Hierarchy of PGM families by order of inclusion (in terms of expressive power as independence models).

Structure learning is hard.[1]

[1] Chickering, Heckerman, and Meek, "Large-Sample Learning of Bayesian Networks is NP-Hard".

Disjoint factorization

We consider a simpler problem: identifying disjoint factors of p,

$$p(\mathbf{v}) = \prod_{i=1}^{k} p(\mathbf{v}_i).$$

[2] Poon and Domingos, "Sum-Product Networks: A New Deep Architecture."
[3] Rudnicki, Wrzesien, and Paja, "All Relevant Feature Selection Methods and Applications."

Disjoint factorization

We consider a simpler problem: identifying disjoint factors of p,

$$p(\mathbf{v}) = \prod_{i=1}^{k} p(\mathbf{v}_i).$$

Greatly simplifies parameter learning and inference:

- each $p(\mathbf{v}_i)$ learned separately;
- decomposable MPE: $\arg\max_{\mathbf{v}} p(\mathbf{v}) = \bigcup_{i=1}^{k} \arg\max_{\mathbf{v}_i} p(\mathbf{v}_i)$.

[2] Poon and Domingos, "Sum-Product Networks: A New Deep Architecture."
[3] Rudnicki, Wrzesien, and Paja, "All Relevant Feature Selection Methods and Applications."

Disjoint factorization

We consider a simpler problem: identifying disjoint factors of p,

$$p(\mathbf{v}) = \prod_{i=1}^{k} p(\mathbf{v}_i).$$

Greatly simplifies parameter learning and inference:

- each $p(\mathbf{v}_i)$ learned separately;
- decomposable MPE: $\arg\max_{\mathbf{v}} p(\mathbf{v}) = \bigcup_{i=1}^{k} \arg\max_{\mathbf{v}_i} p(\mathbf{v}_i)$.

Relation to PGMs:

- Connected components of a PGM;
- Product nodes of a decomposable SPN;[2]
- Partial solution to all-relevant FSS.[3]

---

[2] Poon and Domingos, "Sum-Product Networks: A New Deep Architecture."
[3] Rudnicki, Wrzesien, and Paja, "All Relevant Feature Selection Methods and Applications."

Graphical characterization

Irreducible disjoint factors

Definition

A disjoint factor (DF) is a subset $\mathbf{V}_F \subseteq \mathbf{V}$ such that $\mathbf{V}_F \perp\!\!\!\perp \mathbf{V} \setminus \mathbf{V}_F$.
An irreducible disjoint factor (IDF) is non-empty and contains no other non-empty DF.

Irreducible disjoint factors

Definition

A disjoint factor (DF) is a subset $\mathbf{V}_F \subseteq \mathbf{V}$ such that $\mathbf{V}_F \perp\!\!\!\perp \mathbf{V} \setminus \mathbf{V}_F$.
An irreducible disjoint factor (IDF) is non-empty and contains no other non-empty DF.

The set of all IDFs forms a unique partition of $\mathbf{V}$.

Irreducible disjoint factors

Definition

A disjoint factor (DF) is a subset $\mathbf{V}_F \subseteq \mathbf{V}$ such that $\mathbf{V}_F \perp\!\!\!\perp \mathbf{V} \setminus \mathbf{V}_F$.
An irreducible disjoint factor (IDF) is non-empty and contains no other non-empty DF.

The set of all IDFs forms a unique partition of $\mathbf{V}$.

Our contribution: identify all IDFs with $O(n^2)$ pairwise CI tests

- for distributions satisfying the Intersection property;
- for distributions satisfying the Composition property;
- for any distribution.

Conditional independence properties

Every probability distribution satisfies the *semi-graphoid* axioms,[4]

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \iff \mathbf{Y} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z} \text{ (Symmetry)},$$

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \text{ (Decomposition)},$$

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \cup \mathbf{W} \text{ (Weak Union)},$$

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z} \cup \mathbf{Y} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \text{ (Contraction)}.$$

---

[4] Dawid, "Conditional Independence in Statistical Theory".

[5] Sadeghi and Lauritzen, "Unifying Markov properties in graphical models".

Conditional independence properties

Every probability distribution satisfies the *semi-graphoid* axioms,[4]

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \iff \mathbf{Y} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z} \text{ (Symmetry)},$$

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \text{ (Decomposition)},$$

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \cup \mathbf{W} \text{ (Weak Union)},$$

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z} \cup \mathbf{Y} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \text{ (Contraction)}.$$

A *graphoid* further satisfies

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \cup \mathbf{W} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z} \cup \mathbf{Y} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \text{ (Intersection)},$$

[4] Dawid, "Conditional Independence in Statistical Theory".

[5] Sadeghi and Lauritzen, "Unifying Markov properties in graphical models".

Conditional independence properties

Every probability distribution satisfies the *semi-graphoid* axioms,[4]

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \iff \mathbf{Y} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z} \text{ (Symmetry)},$$

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \text{ (Decomposition)},$$

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \cup \mathbf{W} \text{ (Weak Union)},$$

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z} \cup \mathbf{Y} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \text{ (Contraction)}.$$

A *graphoid* further satisfies

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \cup \mathbf{W} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z} \cup \mathbf{Y} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \text{ (Intersection)},$$

A *compositional graphoid* further satisfies

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \text{ (Composition)}.$$

[4] Dawid, "Conditional Independence in Statistical Theory".

[5] Sadeghi and Lauritzen, "Unifying Markov properties in graphical models".

Conditional independence properties

Every probability distribution satisfies the *semi-graphoid* axioms,[4]

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \iff \mathbf{Y} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z} \text{ (Symmetry)},$$

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \text{ (Decomposition)},$$

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \cup \mathbf{W} \text{ (Weak Union)},$$

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z} \cup \mathbf{Y} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \text{ (Contraction)}.$$

A *graphoid* further satisfies

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \cup \mathbf{W} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z} \cup \mathbf{Y} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \text{ (Intersection)},$$

A *compositional graphoid* further satisfies

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \text{ (Composition)}.$$

Every known PGM is a compositional graphoid.[5]

[4] Dawid, "Conditional Independence in Statistical Theory".
[5] Sadeghi and Lauritzen, "Unifying Markov properties in graphical models".

Connected components

Lemma

$\exists \mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$ *s.t.* $V_i \not\perp\!\!\!\perp V_j \mid \mathbf{Z} \implies V_i, V_j$ *in the same IDF.*

(Decomposition + Weak Union)

Connected components

Lemma

$\exists \mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$ *s.t.* $V_i \not\perp\!\!\!\perp V_j \mid \mathbf{Z} \implies V_i, V_j$ *in the same IDF.*

(Decomposition + Weak Union)

Converse does not hold.

V_1 in $\{00, 01, 10, 11\}$, $V_2 =$ first digit, $V_3 =$ second digit.

$$V_1 \not\perp\!\!\!\perp V_2 \mid \emptyset$$
$$V_1 \not\perp\!\!\!\perp V_3 \mid \emptyset$$
$$V_2 \perp\!\!\!\perp V_3 \mid \emptyset \text{ and } V_2 \perp\!\!\!\perp V_3 \mid V_1$$

Connected components

Lemma

$\exists \mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$ *s.t.* $V_i \not\perp\!\!\!\perp V_j \mid \mathbf{Z} \implies V_i, V_j$ *in the same IDF.*

(Decomposition + Weak Union)

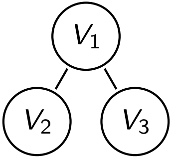
Converse does not hold.

V_1 in $\{00, 01, 10, 11\}$, $V_2 =$ first digit, $V_3 =$ second digit.

$V_1 \not\perp\!\!\!\perp V_2 \mid \emptyset$

$V_1 \not\perp\!\!\!\perp V_3 \mid \emptyset$

$V_2 \perp\!\!\!\perp V_3 \mid \emptyset$ and $V_2 \perp\!\!\!\perp V_3 \mid V_1$



Theorem

$\mathcal{G} = (\mathbf{V}, \mathcal{E})$ *an undirected graph,* $V_i - V_j$ *when* $\exists \mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$ *s.t.* $V_i \not\perp\!\!\!\perp V_j \mid \mathbf{Z} \implies$ *each connected component is an IDF.*

Connected components

Lemma

$\exists \mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$ s.t. $V_i \not\perp\!\!\!\perp V_j \mid \mathbf{Z} \implies V_i, V_j$ in the same IDF.

(Decomposition + Weak Union)

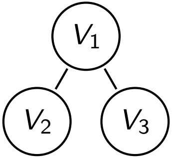
Converse does not hold.

V_1 in $\{00, 01, 10, 11\}$, $V_2 =$ first digit, $V_3 =$ second digit.

$V_1 \not\perp\!\!\!\perp V_2 \mid \emptyset$

$V_1 \not\perp\!\!\!\perp V_3 \mid \emptyset$

$V_2 \perp\!\!\!\perp V_3 \mid \emptyset$ and $V_2 \perp\!\!\!\perp V_3 \mid V_1$



Theorem

$\mathcal{G} = (\mathbf{V}, \mathcal{E})$ an undirected graph, $V_i - V_j$ when $\exists \mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$ s.t. $V_i \not\perp\!\!\!\perp V_j \mid \mathbf{Z} \implies$ each connected component is an IDF.

Complexity $O(2^n)$!

Quadratic pairwise algorithms

Under the Intersection assumption

Theorem

$\mathcal{G} = (\mathbf{V}, \mathcal{E})$ an undirected graph, $V_i - V_j$ when $V_i \not\perp\!\!\!\perp V_j \mid \mathbf{V} \setminus \{V_i, V_j\}$
$\implies$ each connected component is an IDF (under Intersection).

Under the Intersection assumption

Theorem

$\mathcal{G} = (\mathbf{V}, \mathcal{E})$ an undirected graph, $V_i - V_j$ when $V_i \not\perp\!\!\!\perp V_j \mid \mathbf{V} \setminus \{V_i, V_j\}$
$\implies$ each connected component is an IDF (under Intersection).

Reduces to a FSS problem.

Lemma

$V_i \not\perp\!\!\!\perp V_j \mid \mathbf{V} \setminus \{V_i, V_j\} \iff V_j \in \mathbf{MB}_i$ (under Intersection).

Under the Intersection assumption

Theorem

$\mathcal{G} = (\mathbf{V}, \mathcal{E})$ an undirected graph, $V_i - V_j$ when $V_i \not\perp\!\!\!\perp V_j \mid \mathbf{V} \setminus \{V_i, V_j\}$
$\implies$ each connected component is an IDF (under Intersection).

Reduces to a FSS problem.

Lemma

$V_i \not\perp\!\!\!\perp V_j \mid \mathbf{V} \setminus \{V_i, V_j\} \iff V_j \in \mathbf{MB}_i$ (under Intersection).

Counter-example: deterministic relations.

$V_1 = V_2$, $p(V_2 = V_3) = \alpha$, $V_3 = V_4$.
Then, $\mathbf{MB}_1 = \{V_2\}$, $\mathbf{MB}_2 = \{V_1\}$, $\mathbf{MB}_3 = \{V_4\}$ and $\mathbf{MB}_4 = \{V_3\}$.

Under the Composition assumption

Theorem

$\mathcal{G} = (\mathbf{V}, \mathcal{E})$ an undirected graph, $V_i - V_j$ when $V_i \not\perp\!\!\!\perp V_j \mid \emptyset$
$\implies$ each connected component is an IDF (under Composition).

Under the Composition assumption

Theorem

$\mathcal{G} = (\mathbf{V}, \mathcal{E})$ *an undirected graph,* $V_i - V_j$ *when* $V_i \not\perp\!\!\!\perp V_j \mid \emptyset$
$\implies$ *each connected component is an IDF (under Composition).*

Very convenient test.

Under the Composition assumption

Theorem

$\mathcal{G} = (\mathbf{V}, \mathcal{E})$ an undirected graph, $V_i - V_j$ when $V_i \not\perp\!\!\!\perp V_j \mid \emptyset$
$\implies$ each connected component is an IDF (under Composition).

Very convenient test.

Counter-example: XOR relations.

$p(V_1 = V_2 \oplus V_3) = \alpha$ (binary variables).
$V_1 \perp\!\!\!\perp V_2$,
$V_1 \perp\!\!\!\perp V_3$,
yet $\{V_1\} \not\perp\!\!\!\perp \{V_2, V_3\}$.

For any probability distribution

Theorem

$<$ a strict total order of $\mathbf{V}$.

$\mathcal{G} \leftarrow (\mathbf{V}, \emptyset)$ *(empty graph)*
for all $V_i \in \mathbf{V}$ ***do***
    $\mathbf{V}^i_{ind} \leftarrow \emptyset$
    for all $V_j \in \{V | V > V_i\}$ *(processed in $<$ order)* ***do***
        if $V_i \perp\!\!\!\perp V_j \mid \{V | V < V_i\} \cup \mathbf{V}^i_{ind}$ ***then***
            $\mathbf{V}^i_{ind} \leftarrow \mathbf{V}^i_{ind} \cup \{V_j\}$
        else
            Insert a new edge (i, j) in $\mathcal{G}$

$\Longrightarrow$ *each connected component is an IDF.*

For any probability distribution

Theorem

$<$ a strict total order of $\mathbf{V}$.

$\mathcal{G} \leftarrow (\mathbf{V}, \emptyset)$ *(empty graph)*
for all $V_i \in \mathbf{V}$ **do**
  $\mathbf{V}^i_{ind} \leftarrow \emptyset$
  for all $V_j \in \{V | V > V_i\}$ *(processed in $<$ order)* **do**
    if $V_i \perp\!\!\!\perp V_j \mid \{V | V < V_i\} \cup \mathbf{V}^i_{ind}$ **then**
      $\mathbf{V}^i_{ind} \leftarrow \mathbf{V}^i_{ind} \cup \{V_j\}$
    else
      Insert a new edge (i, j) in $\mathcal{G}$

$\Longrightarrow$ *each connected component is an IDF.*

$O(n^2)$ without any assumption.

For any probability distribution

Theorem

$<$ a strict total order of $\mathbf{V}$.

$\mathcal{G} \leftarrow (\mathbf{V}, \emptyset)$ *(empty graph)*
for all $V_i \in \mathbf{V}$ **do**
    $\mathbf{V}^i_{ind} \leftarrow \emptyset$
    for all $V_j \in \{V | V > V_i\}$ *(processed in $<$ order)* **do**
        if $V_i \perp\!\!\!\perp V_j \mid \{V | V < V_i\} \cup \mathbf{V}^i_{ind}$ **then**
            $\mathbf{V}^i_{ind} \leftarrow \mathbf{V}^i_{ind} \cup \{V_j\}$
        else
            Insert a new edge (i, j) in $\mathcal{G}$

$\implies$ *each connected component is an IDF.*

$O(n^2)$ without any assumption.

Prone to cascading effects...

Discussion

Discussion

Three $O(n^2)$ procedures to factorize p into IDFs.

- $V_i \perp\!\!\!\perp V_j \mid \mathbf{V} \setminus \{V_i, V_j\}$, equivalent to FSS (under Intersection);
- $V_i \perp\!\!\!\perp V_j \mid \emptyset$ (under Composition);
- $V_i \perp\!\!\!\perp V_j \mid \mathbf{Z}$, sequential procedure with $\mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$.

---

[6] Gasse, Aussem, and Elghazel, "On the Optimality of Multi-Label Classification under Subset Zero-One Loss for Distributions Satisfying the Composition Property."

[7] Poon and Domingos, "Sum-Product Networks: A New Deep Architecture."

Discussion

Three $O(n^2)$ procedures to factorize p into IDFs.

- $V_i \perp\!\!\!\perp V_j \mid \mathbf{V} \setminus \{V_i, V_j\}$, equivalent to FSS (under Intersection);
- $V_i \perp\!\!\!\perp V_j \mid \emptyset$ (under Composition);
- $V_i \perp\!\!\!\perp V_j \mid \mathbf{Z}$, sequential procedure with $\mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$.

Apply also to conditional distributions $p(\mathbf{y} \mid \mathbf{x})$ (ICML 2015[6]).

[6] Gasse, Aussem, and Elghazel, "On the Optimality of Multi-Label Classification under Subset Zero-One Loss for Distributions Satisfying the Composition Property."

[7] Poon and Domingos, "Sum-Product Networks: A New Deep Architecture."

Discussion

Three $O(n^2)$ procedures to factorize p into IDFs.

- $V_i \perp\!\!\!\perp V_j \mid \mathbf{V} \setminus \{V_i, V_j\}$, equivalent to FSS (under Intersection);
- $V_i \perp\!\!\!\perp V_j \mid \emptyset$ (under Composition);
- $V_i \perp\!\!\!\perp V_j \mid \mathbf{Z}$, sequential procedure with $\mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$.

Apply also to conditional distributions $p(\mathbf{y} \mid \mathbf{x})$ (ICML 2015[6]).

Composition least restrictive in our view: assumed by linear models, greedy-FSS, greedy PGM structure learning...

[6] Gasse, Aussem, and Elghazel, "On the Optimality of Multi-Label Classification under Subset Zero-One Loss for Distributions Satisfying the Composition Property."

[7] Poon and Domingos, "Sum-Product Networks: A New Deep Architecture."

Discussion

Three $O(n^2)$ procedures to factorize p into IDFs.

- $V_i \perp\!\!\!\perp V_j \mid \mathbf{V} \setminus \{V_i, V_j\}$, equivalent to FSS (under Intersection);
- $V_i \perp\!\!\!\perp V_j \mid \emptyset$ (under Composition);
- $V_i \perp\!\!\!\perp V_j \mid \mathbf{Z}$, sequential procedure with $\mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$.

Apply also to conditional distributions $p(\mathbf{y} \mid \mathbf{x})$ (ICML 2015[6]).

Composition least restrictive in our view: assumed by linear models, greedy-FSS, greedy PGM structure learning...

The SPNlearn[7] product split criterion is optimal under Composition.

---

[6] Gasse, Aussem, and Elghazel, "On the Optimality of Multi-Label Classification under Subset Zero-One Loss for Distributions Satisfying the Composition Property."

[7] Poon and Domingos, "Sum-Product Networks: A New Deep Architecture."

Discussion

Three $O(n^2)$ procedures to factorize p into IDFs.

- $V_i \perp\!\!\!\perp V_j \mid \mathbf{V} \setminus \{V_i, V_j\}$, equivalent to FSS (under Intersection);
- $V_i \perp\!\!\!\perp V_j \mid \emptyset$ (under Composition);
- $V_i \perp\!\!\!\perp V_j \mid \mathbf{Z}$, sequential procedure with $\mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$.

Apply also to conditional distributions $p(\mathbf{y} \mid \mathbf{x})$ (ICML 2015[6]).

Composition least restrictive in our view: assumed by linear models, greedy-FSS, greedy PGM structure learning...

The SPNlearn[7] product split criterion is optimal under Composition.

The all-relevant FSS problem seems closely related

$$\{V_j \mid \exists \mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}, \text{ s.t. } V_i \not\perp\!\!\!\perp V_j \mid \mathbf{Z}\}.$$

---

[6] Gasse, Aussem, and Elghazel, "On the Optimality of Multi-Label Classification under Subset Zero-One Loss for Distributions Satisfying the Composition Property."

[7] Poon and Domingos, "Sum-Product Networks: A New Deep Architecture."

Identifying the Irreducible Disjoint Factors of a Multivariate Probability Distribution

PGM 2016

Maxime Gasse, Alex Aussem

Thank you!