732A96 Advanced Machine Learning

Graphical Models and Hidden Markov Models

Jose M. Peña
IDA, Linköping University, Sweden

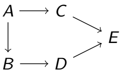
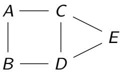
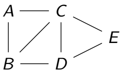
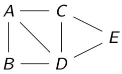
Lecture 2: Probabilistic Reasoning

Contents

Literature

- Main source
  - Lauritzen, S. L. and Spiegelhalter, D. J. Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems. *Journal of the Royal Statistical Society B* 50, 157-224, 1988.
- Additional source
  - Bishop, C. M. *Pattern Recognition and Machine Learning*. Springer, 2006.

Probabilistic Reasoning for BNs: Lauritzen-Spiegelhalter Algorithm

Original	Moralization	Triangulation	Triangulation
		 Ordering: E, D, A, B, C	 Ordering: E, B, A, C, D

Moralization

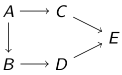
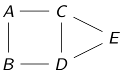
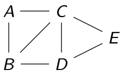
Repeat for every node in G
    Make its parents pairwise adjacent
Replace all the directed edges in G with undirected edges

- An UG is **triangulated** if every cycle in it contains a chord, i.e. an edge between two non-consecutive nodes in the cycle.

Triangulation

Repeat until all the nodes in G are marked
    If there is an unmarked node whose unmarked neighbours are a complete set then
        Mark it
    Else
        Mark any unmarked node and make its unmarked neighbours a complete set

Probabilistic Reasoning for BNs: Lauritzen-Spiegelhalter Algorithm

Original	Moralization	Triangulation
		 Ordering: E, D, A, B, C Candidate cliques: $\{C, D, E\}$, $\{B, C, D\}$, $\{A, B, C\}$, $\{B, C\}$, $\{C\}$ Cliques: $\{C, D, E\}$, $\{A, B, C\}$, $\{B, C, D\}$ RIP: $C_1 = \{A, B, C\}$, $C_2 = \{B, C, D\}$, $C_3 = \{C, D, E\}$

- The cliques of a triangulated graph can be ordered as $C_{1:k}$ to have the **running intersection property (RIP)**, i.e. $C_j \cap C_{1:j-1} \subset C_i$ with $i < j$.

RIP

Repeat until all the nodes in G are marked
    Mark an unmarked node whose unmarked neighbours are a complete set
    Record the union of the node and its unmarked neighbours as a candidate clique
Remove any candidate clique that is included in any other
Label every clique with the number of the first iteration that marked one of its nodes
Sort the cliques in descending order of their labels

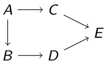
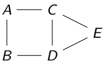
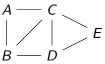
- Let $S_j = C_j \cap C_{1:j-1}$ and $R_j = C_j \setminus S_j$.
- **Note that** $R_j \perp_G C_{1:j-1} \setminus S_j | S_j$.

Probabilistic Reasoning for BNs: Lauritzen-Spiegelhalter Algorithm

- **Exercise**. Moralize, triangulate and order the resulting cliques to as to satisfy the RIP for the following DAG.

$$
\begin{tikzcd}
A \arrow[d] & B \arrow[dl] \arrow[dr] & C \arrow[d] \\
D \arrow[d] & & E \arrow[d] \\
F \arrow[d] \arrow[dr] & & G \arrow[dl] \arrow[d] \\
H & I & J
\end{tikzcd}
$$

Probabilistic Reasoning for BNs: Lauritzen-Spiegelhalter Algorithm

Original	Moralization	Triangulation
 $p(A,B,C,D,E) = q(A)q(B\|A)q(C\|A)q(D\|B)q(E\|C,D)$		 RIP: $C_1 = \{A,B,C\}$, $C_2 = \{B,C,D\}$, $C_3 = \{C,D,E\}$

- Let $\varphi(C_1) := q(A)q(B|A)q(C|A)$, $\varphi(C_2) := q(D|B)$, $\varphi(C_3) := q(E|C,D)$.
- Then
$$p(R_3|S_3) = p(E|C,D) := \varphi(C_3)/\sum_{R_3}\varphi(C_3) = q(E|C,D)/\sum_{E} q(E|C,D) = q(E|C,D)$$
- Since $S_3 \subset C_2$ by RIP, let
$$\varphi(C_2) := \varphi(C_2)\sum_{R_3}\varphi(C_3) = q(D|B)\sum_{E} q(E|C,D) = q(D|B)$$
- Then
$$p(R_2|S_2) = p(D|B,C) := \varphi(C_2)/\sum_{R_2}\varphi(C_2) = q(D|B)/\sum_{D} q(D|B) = q(D|B)$$
- Since $S_2 \subset C_1$ by RIP, let
$$\varphi(C_1) := \varphi(C_1)\sum_{R_2}\varphi(C_2) = q(A)q(B|A)q(C|A)\sum_{D} q(D|B) = q(A)q(B|A)q(C|A)$$
- Then, $p(R_1|S_1) = p(A,B,C) := \varphi(C_1)/\sum_{R_1}\varphi(C_1)$
$$= q(A)q(B|A)q(C|A)/\sum_{A,B,C} q(A)q(B|A)q(C|A) = q(A)q(B|A)q(C|A)$$

Probabilistic Reasoning for BNs: Lauritzen-Spiegelhalter Algorithm

Original	Moralization	Triangulation
$\begin{tikzcd} A \arrow[r] \arrow[dd] & C \arrow[dr] & \\ & & E \\ B \arrow[r] & D \arrow[ur] & \end{tikzcd}$ $p(A,B,C,D,E) = q(A)q(B\mid A)q(C\mid A)q(D\mid B)q(E\mid C,D)$	$\begin{tikzcd} A \arrow[r, dash] \arrow[dd, dash] & C \arrow[dr, dash] \arrow[dd, dash] & \\ & & E \\ B \arrow[r, dash] & D \arrow[ur, dash] & \end{tikzcd}$	$\begin{tikzcd} A \arrow[r, dash] \arrow[dd, dash] & C \arrow[dr, dash] \arrow[dd, dash] & \\ & & E \\ B \arrow[r, dash] \arrow[uur, dash] & D \arrow[ur, dash] & \end{tikzcd}$ RIP: $C_1 = \{A,B,C\}$, $C_2 = \{B,C,D\}$, $C_3 = \{C,D,E\}$

- Note that $S_1 = \varnothing$ and, thus, $p(C_1) = p(R_1|S_1)$.
- Since $S_2 \subset C_1$ by RIP
$$p(S_2) = \sum_{C_1 \setminus S_2} p(C_1)$$
- Then
$$p(C_2) := p(R_2|S_2)p(S_2)$$
- Since $S_3 \subset C_2$ by RIP
$$p(S_3) = \sum_{C_2 \setminus S_3} p(C_2)$$
- Then
$$p(C_3) := p(R_3|S_3)p(S_3)$$
- **Output**: $p(C_1)$, $p(C_2)$ and $p(C_3)$.

Probabilistic Reasoning for BNs: Lauritzen-Spiegelhalter Algorithm

- In general, rewrite

$$p(X) = \prod_{i=1}^{n} q(X_i | pa_G(X_i)) = \prod_{i=1}^{k} \varphi(C_i)$$

- Note that

$$p(X) = f(C_{1:k-1} \setminus S_k, S_k) g(S_k, R_k)$$

  and thus

$$R_k \perp_p C_{1:k-1} \setminus S_k | S_k$$

- Then

$$p(X) = p(C_{1:k-1}) p(R_k | C_{1:k-1}) = p(C_{1:k-1}) p(R_k | S_k)$$

- Note that

$$p(C_{1:k-1}) = \sum_{R_k} p(X) = [\prod_{i=1}^{k-1} \varphi(C_i)] \sum_{R_k} \varphi(C_k)$$

  and thus

$$p(R_k | S_k) = \varphi(C_k) / \sum_{R_k} \varphi(C_k)$$

- Let $S_k \subset C_j$ by RIP. Then, replace $\varphi(C_j)$ with $\varphi(C_j) \sum_{R_k} \varphi(C_k)$.
- Then, $p(C_{1:k-1}) = \prod_{i=1}^{k-1} \varphi(C_i)$
- Repeat the above for $p(C_{1:k-1})$. **Output**: $p(R_i | S_i)$ for all i.

Probabilistic Reasoning for BNs: Lauritzen-Spiegelhalter Algorithm

- Note that $S_1 = \varnothing$ and, thus, $p(C_1) = p(R_1|S_1)$.
- Since $S_2 \subset C_1$ by RIP

$$p(S_2) = \sum_{C_1 \setminus S_2} p(C_1)$$

- Then

$$p(C_2) = p(R_2|S_2)p(S_2)$$

- Repeat the above for $C_{3:k}$. **Output**: $p(C_i)$ for all i.

LS algorithm

Moralize G
Triangulate G
Order the cliques of G so as to satisfy the RIP
For $i = k, \ldots, 1$ do
    Let $S_i \subset C_j$ by RIP
    $p(R_i|S_i) := \varphi(C_i) / \sum_{R_i} \varphi(C_i)$
    $\varphi(C_j) := \varphi(C_j) \sum_{R_i} \varphi(C_i)$
For $i = 1, \ldots, k$ do
    Let $S_i \subset C_j$ by RIP
    $p(S_i) := \sum_{C_j \setminus S_i} p(C_j)$
    $p(C_i) := p(R_i|S_i)p(S_i)$
Return $p(C_1), \ldots, p(C_k)$

Probabilistic Reasoning for BNs: Lauritzen-Spiegelhalter Algorithm

- Then, we know $p(Y)$ for all $Y \subseteq C_i$.
- What if we want $p(Y|o)$ with $O \subseteq X \setminus Y$?
  - Let $U = X \setminus O$. Note that $p(U|o)$ is a normalized version of $p(U, o)$, and the latter can be expressed as $P(U, o) = P(V) \prod_i 1_{O_i = o_i}$.
  - Then, just add the factors $1_{O_i = o_i}$ to the appropriate cliques, perform inference as before, and finally normalize.
- What if $Y \nsubseteq C_i$ for all i ?
  - Compute $p(z)$ where $z = \{y, o\}$ for all y, and normalize.
- **Exercise**. Apply the LS algorithm to compute $p(J|a)$ in the previous exercise.

Probabilistic Reasoning for BNs: Most Probable Configuration

- Note that

$$\max_X p(X) = \max_{R_1,\dots,R_k} \varphi(C_1)\dots\varphi(C_k) = \max_{R_1,\dots,R_{k-1}} [\varphi(C_1)\dots\varphi(C_{k-1}) \max_{R_k} \varphi(C_k)]$$

Probability of the most probable configuration

For $i = k, \dots, 1$ do
    Let $S_i \subset C_j$ by RIP
    $\phi(S_i) := \max_{R_i} \varphi(C_i)$
    $\varphi(C_j) := \varphi(C_j)\phi(C_i)$
    $\psi(S_i) := \arg\max_{R_i} \varphi(C_i)$
Return $\phi(S_1)$

Most probable configuration

For $i = 1, \dots, k$ do
    $R_i^{max} := \psi(S_i^{max})$
Return $R_1^{max}, \dots, R_k^{max}$

- After running the first algorithm and since $S_1 = S_1^{max} = \varnothing$, we have that

$$R_1^{max} = \arg\max_{R_1}[\max_{R_2,\dots,R_k} p(X)] = \arg\max_{R_1} \varphi(C_1) = \psi(S_1^{max})$$

- Note that

$$\arg\max_{R_1,R_2}[\max_{R_3,\dots,R_k} p(X)] = \arg\max_{R_1,R_2} \varphi(C_1)\varphi(C_2)$$

  and since it has to be consistent with R_1^{max}, then

$$R_2^{max} = \arg\max_{R_2} \varphi(S_1^{max}, R_1^{max})\varphi(S_2^{max}, R_2)$$

- And so on.

Probabilistic Reasoning for MNs

- The same as for BNs. Just skip the moralization step.
- No need to include Z. After inference, just normalize any $p(C_i)$ to obtain Z and use it to normalize the rest of the output.