ECE 590.10: Graphical Models and Inference

Lecture: Marginalization and Factor Graphs

Henry D. Pfister

Department of Electrical and Computer Engineering
Duke University

Outline

The Importance of Marginalization (1)

- Consider a random vector $(X_1, X_2, \ldots, X_6) \in \mathcal{X}^6$ where
  - the vector is indirectly observed via the random variable Y
  - the posterior probability given $Y = y$ satisfies:

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_6 = x_6 | Y = y) \propto f(x_1, x_2, x_3, x_4, x_5, x_6)$$

The Importance of Marginalization (2)

- Marginalizing down to X_1 gives the a posteriori probability (APP)

$$\mathbb{P}(X_1 = x_1 | Y = y) \propto \sum_{(x_2,\dots,x_6)\in\mathcal{X}^5} f(x_1, x_2, x_3, x_4, x_5, x_6)$$

- The maximum a posteriori (MAP) decision for X_1 is given by

$$\hat{x}_1 = \arg\max_{x_1\in\mathcal{X}} \sum_{(x_2,\dots,x_6)\in\mathcal{X}^5} f(x_1, x_2, x_3, x_4, x_5, x_6)$$

- For many problems, computing the marginalization sum is a significant challenge

The Importance of Factorization (1)

- Simply summing requires $|\mathcal{X}|^5$ operations for each $x_1 \in \mathcal{X}$:

$$g_1(x_1) \triangleq \sum_{x_2^6 \in \mathcal{X}^5} f(x_1, x_2, x_3, x_4, x_5, x_6)$$

  - Thus, we need $|\mathcal{X}|^6$ operations

- If f factors as follows, then marginalization can be simplified:

$$f(x_1, x_2, x_3, x_4, x_5, x_6) = f_1(x_1, x_2, x_3) f_2(x_1, x_4, x_6) f_3(x_4) f_4(x_4, x_5)$$

The Importance of Factorization (2)

For example, we can write $g_1(x_1)$ as:

$$= \sum_{x_2^6} f_1(x_1, x_2, x_3) f_2(x_1, x_4, x_6) f_3(x_4) f_4(x_4, x_5)$$

The Importance of Factorization (2)

For example, we can write $g_1(x_1)$ as:

$$= \sum_{x_2^6} f_1(x_1, x_2, x_3) f_2(x_1, x_4, x_6) f_3(x_4) f_4(x_4, x_5)$$

$$= \sum_{x_2^5} f_1(x_1, x_2, x_3) f_3(x_4) f_4(x_4, x_5) \left[\sum_{x_6} f_2(x_1, x_4, x_6) \right]$$

The Importance of Factorization (2)

For example, we can write $g_1(x_1)$ as:

$$= \sum_{x_2^6} f_1(x_1, x_2, x_3) f_2(x_1, x_4, x_6) f_3(x_4) f_4(x_4, x_5)$$

$$= \sum_{x_2^5} f_1(x_1, x_2, x_3) f_3(x_4) f_4(x_4, x_5) \left[\sum_{x_6} f_2(x_1, x_4, x_6) \right]$$

$$= \sum_{x_2^4} f_1(x_1, x_2, x_3) f_3(x_4) \left[\sum_{x_5} f_4(x_4, x_5) \right] \left[\sum_{x_6} f_2(x_1, x_4, x_6) \right]$$

The Importance of Factorization (2)

For example, we can write $g_1(x_1)$ as:

$$= \sum_{x_2^6} f_1(x_1, x_2, x_3) f_2(x_1, x_4, x_6) f_3(x_4) f_4(x_4, x_5)$$

$$= \sum_{x_2^5} f_1(x_1, x_2, x_3) f_3(x_4) f_4(x_4, x_5) \left[\sum_{x_6} f_2(x_1, x_4, x_6) \right]$$

$$= \sum_{x_2^4} f_1(x_1, x_2, x_3) f_3(x_4) \left[\sum_{x_5} f_4(x_4, x_5) \right] \left[\sum_{x_6} f_2(x_1, x_4, x_6) \right]$$

$$= \sum_{x_2^3} f_1(x_1, x_2, x_3) \left[\sum_{x_4} f_3(x_4) \left[\sum_{x_5} f_4(x_4, x_5) \right] \left[\sum_{x_6} f_2(x_1, x_4, x_6) \right] \right]$$

The Importance of Factorization (2)

For example, we can write $g_1(x_1)$ as:

$$= \sum_{x_2^6} f_1(x_1, x_2, x_3) f_2(x_1, x_4, x_6) f_3(x_4) f_4(x_4, x_5)$$

$$= \sum_{x_2^5} f_1(x_1, x_2, x_3) f_3(x_4) f_4(x_4, x_5) \left[\sum_{x_6} f_2(x_1, x_4, x_6) \right]$$

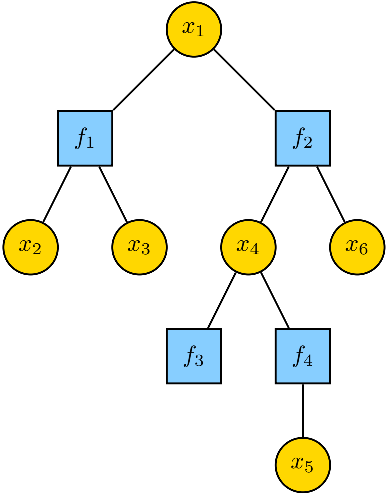
$$= \sum_{x_2^4} f_1(x_1, x_2, x_3) f_3(x_4) \left[\sum_{x_5} f_4(x_4, x_5) \right] \left[\sum_{x_6} f_2(x_1, x_4, x_6) \right]$$

$$= \sum_{x_2^3} f_1(x_1, x_2, x_3) \left[\sum_{x_4} f_3(x_4) \left[\sum_{x_5} f_4(x_4, x_5) \right] \left[\sum_{x_6} f_2(x_1, x_4, x_6) \right] \right]$$

$$= \sum_{x_2} \left[\sum_{x_3} f_1(x_1, x_2, x_3) \right] \left[\sum_{x_4} f_3(x_4) \left[\sum_{x_5} f_4(x_4, x_5) \right] \left[\sum_{x_6} f_2(x_1, x_4, x_6) \right] \right]$$

This implementation requires roughly $2\,|\mathcal{X}|^3 + 5\,|\mathcal{X}|^2$ operations

The Factor Graph and Variable-Leaf Removal



Graphical Marginalization:

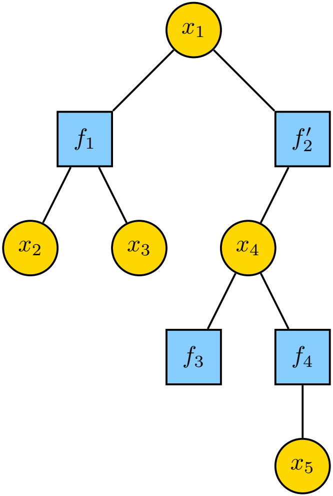
- Factor graph G shows variables involved in each factor
- If the factor graph is a tree, then marginalization by variable-leaf removal is efficient
- After marginalization of a variable leaf, the variable is removed and its effect is absorbed into the parent factor node.
- This costs $|\mathcal{X}|^d$, where d is the degree of updated factor node

$$g_1(x_1) = \sum_{x_2^5} f_1(x_1, x_2, x_3) f_3(x_4) f_4(x_4, x_5) \left[\sum_{x_6} f_2(x_1, x_4, x_6) \right]$$

The Factor Graph and Variable-Leaf Removal

Graphical Marginalization:

- Factor graph G shows variables involved in each factor
- If the factor graph is a tree, then marginalization by variable-leaf removal is efficient
- After marginalization of a variable leaf, the variable is removed and its effect is absorbed into the parent factor node.
- This costs $|\mathcal{X}|^d$, where d is the degree of updated factor node

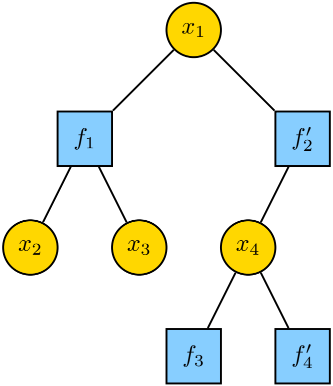


$$g_1(x_1) = \sum_{x_2^4} f_1(x_1, x_2, x_3) f_3(x_4) \left[\sum_{x_5} f_4(x_4, x_5) \right] f_2'(x_1, x_4)$$

The Factor Graph and Variable-Leaf Removal

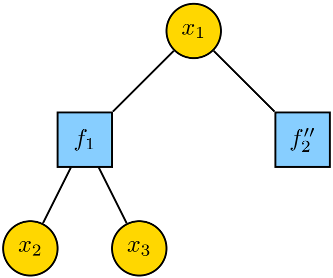
Graphical Marginalization:

- Factor graph G shows variables involved in each factor
- If the factor graph is a tree, then marginalization by variable-leaf removal is efficient
- After marginalization of a variable leaf, the variable is removed and its effect is absorbed into the parent factor node.
- This costs $|\mathcal{X}|^d$, where d is the degree of updated factor node



$$g_1(x_1) = \sum_{x_2^3} f_1(x_1, x_2, x_3) \left[\sum_{x_4} f_3(x_4) f_4'(x_4) f_2'(x_1, x_4) \right]$$

The Factor Graph and Variable-Leaf Removal

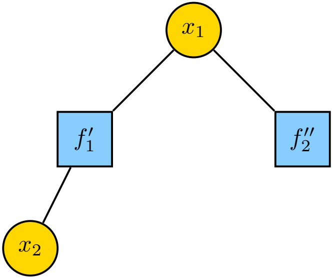


Graphical Marginalization:

- Factor graph G shows variables involved in each factor
- If the factor graph is a tree, then marginalization by variable-leaf removal is efficient
- After marginalization of a variable leaf, the variable is removed and its effect is absorbed into the parent factor node.
- This costs $|\mathcal{X}|^d$, where d is the degree of updated factor node

$$g_1(x_1) = \sum_{x_2} \left[\sum_{x_3} f_1(x_1, x_2, x_3) \right] f_2''(x_1)$$

The Factor Graph and Variable-Leaf Removal

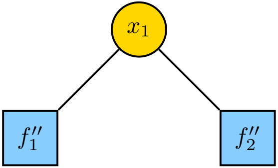


Graphical Marginalization:

- Factor graph G shows variables involved in each factor
- If the factor graph is a tree, then marginalization by variable-leaf removal is efficient
- After marginalization of a variable leaf, the variable is removed and its effect is absorbed into the parent factor node.
- This costs $|\mathcal{X}|^d$, where d is the degree of updated factor node

$$g_1(x_1) = \left[\sum_{x_2} f'_1(x_1, x_2) \right] f''_2(x_1)$$

The Factor Graph and Variable-Leaf Removal



Graphical Marginalization:

- Factor graph G shows variables involved in each factor
- If the factor graph is a tree, then marginalization by variable-leaf removal is efficient
- After marginalization of a variable leaf, the variable is removed and its effect is absorbed into the parent factor node.
- This costs $|\mathcal{X}|^d$, where d is the degree of updated factor node

$$g_1(x_1) = f_1''(x_1) f_2''(x_1)$$

Factor Graphs and Probability

- Factor graphs for PMF of random variables $X_1, \ldots, X_n$
  - Any non-negative $f(x_1, \ldots, x_n)$ can be normalized to be a PMF

$$\mathbb{P}\left(X_1^n = x_1^n\right) = \frac{1}{Z} f(x_1, \ldots, x_n)$$

  - Where $Z = \sum_{x_1^n} f(x_1, \ldots, x_n)$ is called the partition function
  - The complexity of computing Z is similar to marginalization

Bayesian Networks

- Factor graphs are closely related to Bayesian Networks (BNs)
  - BN is a directed acyclic graph with nodes $V = \{X_1, \ldots, X_n\}$:

$$\mathbb{P}\left(X_1^n = x_1^n\right) = \prod_{i=1}^{n} \mathbb{P}\left(X_i = x_i \,|\, X_{\pi(i)} = x_{\pi(i)}\right),$$

  where $\pi(i)$ is the set of VNs with a directed edge to X_i

- For a tree factor graph, one can build a BN with similar structure
  - If the FG has cycles, then a similar BN may not exist

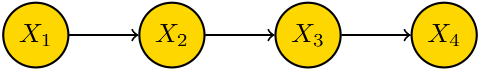
Markov Chain Example

- Factor graph of the Markov chain: $X_1 \to X_2 \to X_3 \to X_4$

$$f(x_1, x_2, x_3, x_4) = f_{12}(x_1, x_2) f_{23}(x_2, x_3) f_{34}(x_3, x_4)$$

- BN of the Markov chain: $X_1 \to X_2 \to X_3 \to X_4$

$$\mathbb{P}\left(X_1^n = x_1^n\right) = \mathbb{P}(X_1 = x_1) \prod_{i=2}^{4} \mathbb{P}\left(X_i = x_i \mid X_{i-1} = x_{i-1}\right),$$



Conditional Independence for Factor Graphs

- Let $A, B, S \subset [n]$ be disjoint subsets of VNs in factor graph G
  - If S separates A from B (i.e., there is no path in G from A to B that avoids S), then X_A cond. ind. of X_B given X_S

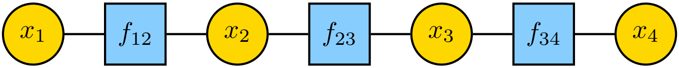
$$P(x_A, x_B | x_S) = P(x_A | x_S) P(x_B | x_S)$$

Conditional Independence for Factor Graphs

- Let $A, B, S \subset [n]$ be disjoint subsets of VNs in factor graph G
  - If S separates A from B (i.e., there is no path in G from A to B that avoids S), then X_A cond. ind. of X_B given X_S

$$P(x_A, x_B | x_S) = P(x_A | x_S) P(x_B | x_S)$$

  - Markov chain example: $A = \{x_1, x_2\}$, $B = \{x_4\}$, $S = \{x_3\}$



Conditional Independence for Factor Graphs

- Let $A, B, S \subset [n]$ be disjoint subsets of VNs in factor graph G
  - If S separates A from B (i.e., there is no path in G from A to B that avoids S), then X_A cond. ind. of X_B given X_S

$$P(x_A, x_B | x_S) = P(x_A | x_S) P(x_B | x_S)$$

  - Markov chain example: $A = \{x_1, x_2\}$, $B = \{x_4\}$, $S = \{x_3\}$

- Sketch of Proof:
  - Fixing $X_S = x_S$ separates the FG into disjoint components

Conditional Independence for Factor Graphs

- Let $A, B, S \subset [n]$ be disjoint subsets of VNs in factor graph G
  - If S separates A from B (i.e., there is no path in G from A to B that avoids S), then X_A cond. ind. of X_B given X_S

$$P(x_A, x_B | x_S) = P(x_A | x_S) P(x_B | x_S)$$

  - Markov chain example: $A = \{x_1, x_2\}$, $B = \{x_4\}$, $S = \{x_3\}$

- Sketch of Proof:
  - Fixing $X_S = x_S$ separates the FG into disjoint components
  - Distributive law shows VNs in disjoint components are ind.
  - X_A and X_B ind. because A and B in different components