PROBABILISTIC MODELS FOR STRUCTURED DATA

06: Factor Graph

Instructor: Yizhou Sun

yzsun@cs.ucla.edu

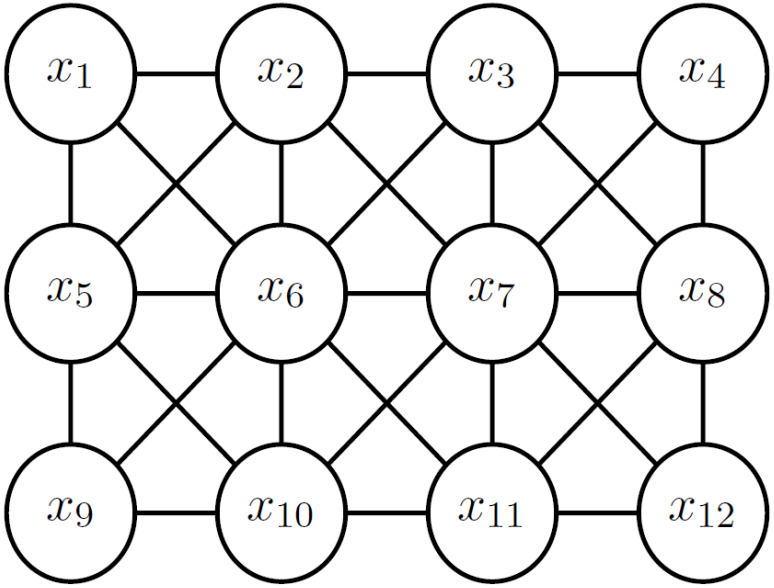
February 5, 2020

Content

- Unifying Everything: Factor Graph
- Inference
  - Sum-product message passing for factor graphs
- Learning
- Summary

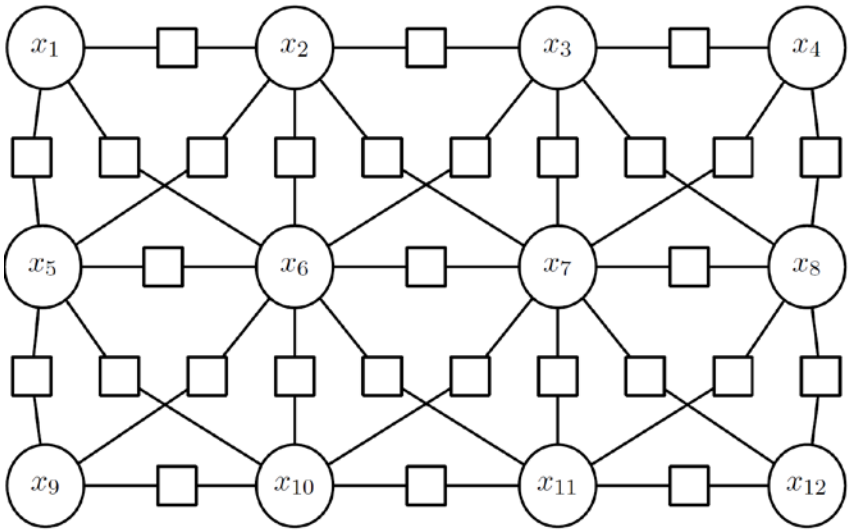
Limitations of Current Graph Representations

- Ambiguous
  - E.g., do we model pairwise interaction only or model size-4 clique?

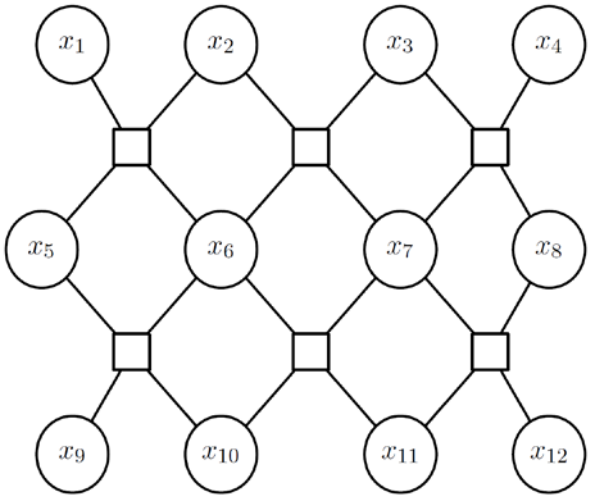


A Clearer Representation

- Factor graph
  - Use square nodes to denote factors, which explicitly tell which variables are involved



Size-2 clique



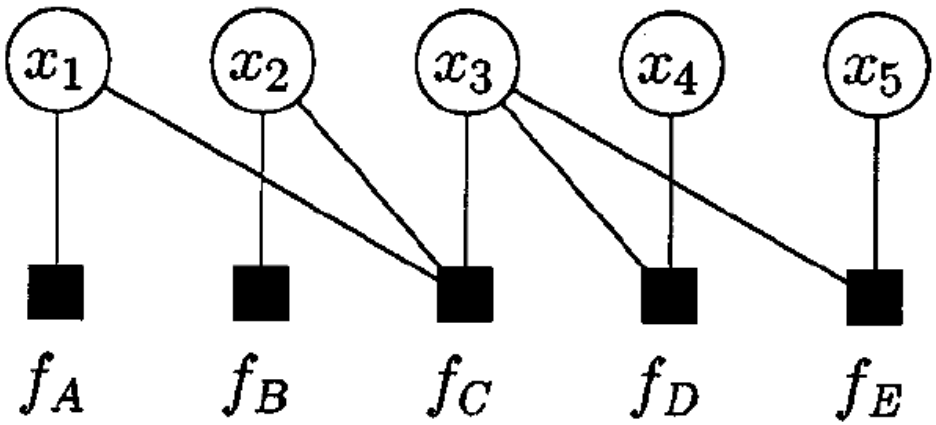
Size-4 clique

A Formal Definition of Factor Graph

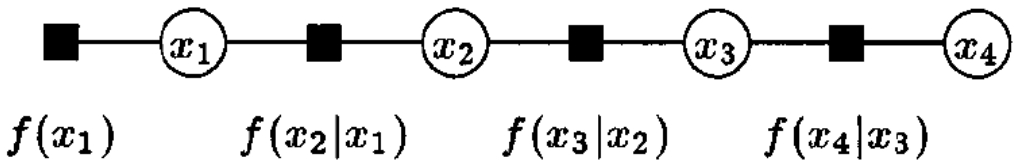
- A bipartite graph
  - Variable nodes: denote each variable x_i
  - Factor nodes: denote each factor f_j
  - Edges: an edge connects variable x_i and f_j if f_j takes x_i as an argument
- The distribution represented by a factor graph:
  - $p(x_1, x_2, \dots, x_n) \propto \prod_{j \in J} f_j(X_j)$
    - where X_j is a subset of $\{x_1, x_2, \dots, x_n\}$, and $f_j(X_j)$ is a function having elements of X_j as arguments

Example

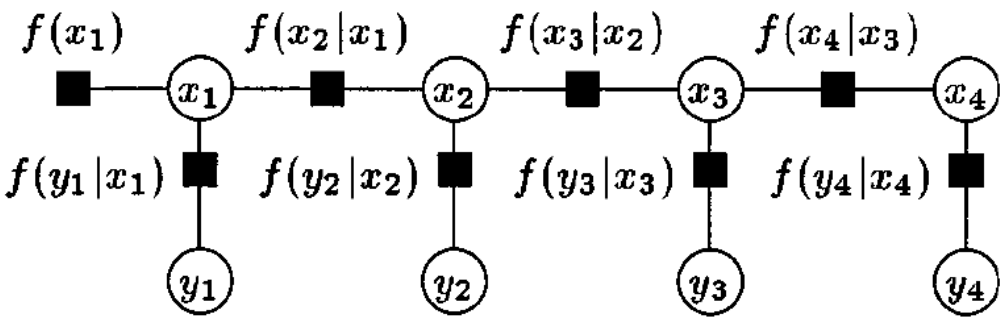
- A factor graph with 5 variables
  - $p(x_1, x_2, \ldots, x_5) \propto f_A(x_1) f_B(x_2) f_C(x_1, x_2, x_3) f_D(x_3, x_4) f_E(x_3, x_5)$



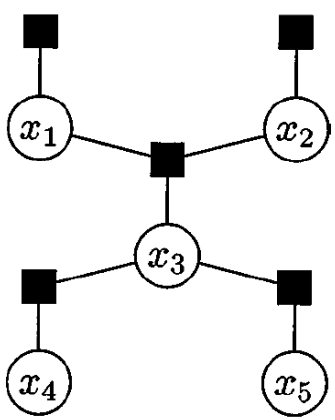
Other Models Represented as Factor Graph



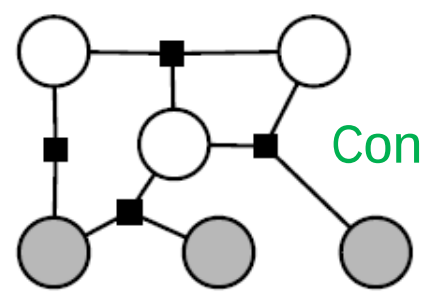
Markov Chain



HMM



MRF or Bayesian Network



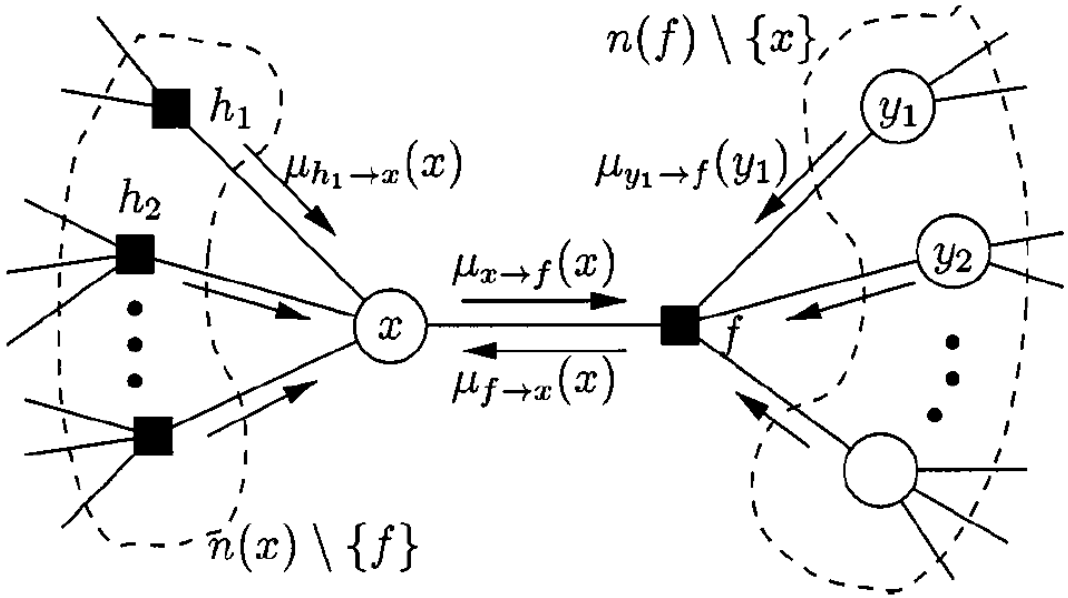
Conditional Random Field

Content

- Unifying Everything: Factor Graph
- Inference
  - Sum-product message passing for factor graphs
- Learning
- Summary

Sum-Product Message Passing

- Purpose: compute marginal distribution
- Two types of messages
  - Variable to function: $\mu_{x\to f}(x) = \prod_{h\in n(x)\setminus\{f\}} \mu_{h\to x}(x)$
  - Function to variable: $\mu_{f\to x}(x) = \sum_{\sim\{x\}} f(X) \prod_{y\in n(f)\setminus\{x\}} \mu_{y\to f}(y)$
    - $X = n(f)$
    - $\sim\{x\}$ means $X\setminus\{x\}$



Example

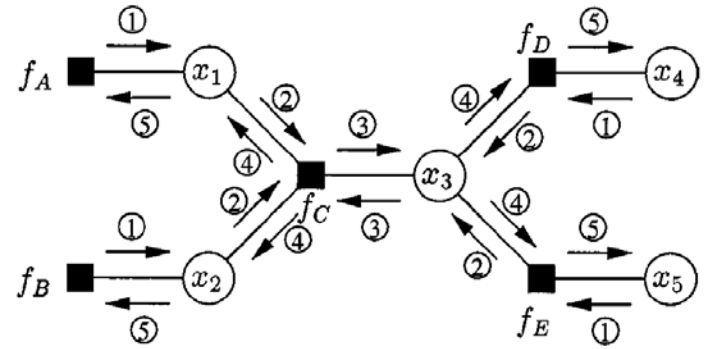
- $p(x_1, x_2, \dots, x_5) \propto f_A(x_1) f_B(x_2) f_C(x_1, x_2, x_3) f_D(x_3, x_4) f_E(x_3, x_5)$
- Step 1 (leaf nodes):

$$\mu_{f_A \to x_1}(x_1) = \sum_{\sim\{x_1\}} f_A(x_1) = f_A(x_1)$$

$$\mu_{f_B \to x_2}(x_2) = \sum_{\sim\{x_2\}} f_B(x_2) = f_B(x_2)$$

$$\mu_{x_4 \to f_D}(x_4) = 1$$

$$\mu_{x_5 \to f_E}(x_5) = 1.$$



- Step 2:

$$\mu_{x_1 \to f_C}(x_1) = \mu_{f_A \to x_1}(x_1)$$

$$\mu_{x_2 \to f_C}(x_2) = \mu_{f_B \to x_2}(x_2)$$

$$\mu_{f_D \to x_3}(x_3) = \sum_{\sim\{x_3\}} \mu_{x_4 \to f_D}(x_4) f_D(x_3, x_4)$$

$$\mu_{f_E \to x_3}(x_3) = \sum_{\sim\{x_3\}} \mu_{x_5 \to f_E}(x_5) f_E(x_3, x_5).$$

Example (Cont.)

- $p(x_1, x_2, \ldots, x_5) \propto f_A(x_1) f_B(x_2) f_C(x_1, x_2, x_3) f_D(x_3, x_4) f_E(x_3, x_5)$
- Step 3:

$$\mu_{f_C \to x_3}(x_3) = \sum_{\sim\{x_3\}} \mu_{x_1 \to f_C}(x_1) \mu_{x_2 \to f_C}(x_2) f_C(x_1, x_2, x_3)$$

$$\mu_{x_3 \to f_C}(x_3) = \mu_{f_D \to x_3}(x_3) \mu_{f_E \to x_3}(x_3).$$

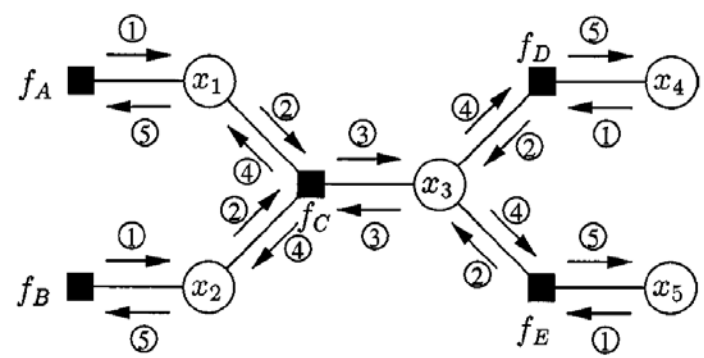
- Step 4:

$$\mu_{f_C \to x_1}(x_1) = \sum_{\sim\{x_1\}} \mu_{x_3 \to f_C}(x_3) \mu_{x_2 \to f_C}(x_2) f_C(x_1, x_2, x_3)$$

$$\mu_{f_C \to x_2}(x_2) = \sum_{\sim\{x_2\}} \mu_{x_3 \to f_C}(x_3) \mu_{x_1 \to f_C}(x_1) f_C(x_1, x_2, x_3)$$

$$\mu_{x_3 \to f_D}(x_3) = \mu_{f_C \to x_3}(x_3) \mu_{f_E \to x_3}(x_3)$$

$$\mu_{x_3 \to f_E}(x_3) = \mu_{f_C \to x_3}(x_3) \mu_{f_D \to x_3}(x_3).$$



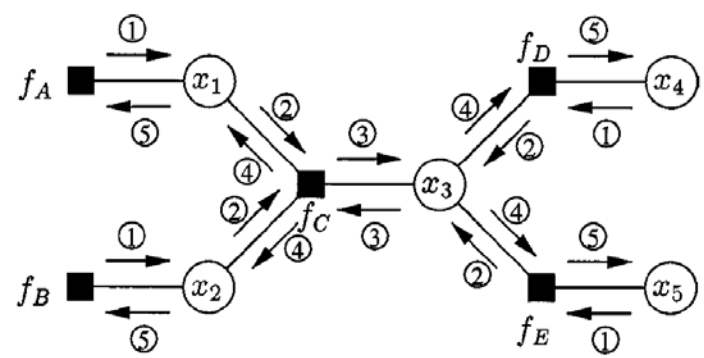
Example (Cont.)

- $p(x_1, x_2, \ldots, x_5) \propto f_A(x_1) f_B(x_2) f_C(x_1, x_2, x_3) f_D(x_3, x_4) f_E(x_3, x_5)$
- Step 5:

$$\mu_{x_1 \to f_A}(x_1) = \mu_{f_C \to x_1}(x_1)$$
$$\mu_{x_2 \to f_B}(x_2) = \mu_{f_C \to x_2}(x_2)$$
$$\mu_{f_D \to x_4}(x_4) = \sum_{\sim\{x_4\}} \mu_{x_3 \to f_D}(x_3) f_D(x_3, x_4)$$
$$\mu_{f_E \to x_5}(x_5) = \sum_{\sim\{x_5\}} \mu_{x_3 \to f_E}(x_3) f_E(x_3, x_5).$$

- Termination

$$g_1(x_1) = \mu_{f_A \to x_1}(x_1) \mu_{f_C \to x_1}(x_1)$$
$$g_2(x_2) = \mu_{f_B \to x_2}(x_2) \mu_{f_C \to x_2}(x_2)$$
$$g_3(x_3) = \mu_{f_C \to x_3}(x_3) \mu_{f_D \to x_3}(x_3) \mu_{f_E \to x_3}(x_3)$$
$$g_4(x_4) = \mu_{f_D \to x_4}(x_4)$$
$$g_5(x_5) = \mu_{f_E \to x_5}(x_5).$$



Max-Product Message Passing

- Purpose: compute maximal probability
- Two types of messages
  - Variable to function:

$$\mu_{x\to f}(x) = \prod_{h\in n(x)\setminus\{f\}} \mu_{h\to x}(x)$$

  - Function to variable:

$$\mu_{f\to x}(x) = \max_{\sim\{x\}} f(X) \prod_{y\in n(f)\setminus\{x\}} \mu_{y\to f}(y)$$

Content

- Unifying Everything: Factor Graph
- Inference
  - Sum-product message passing for factor graphs
- Learning
- Summary

Learning

- Follows solutions in MRF and CRF
- MLE on Exponential family
- Two cases:
  - Fully observed data
  - Partially observed data
- Need inference computation in the learning algorithm

Content

- Unifying Everything: Factor Graph
- Inference
  - Sum-product message passing for factor graphs
- Learning
- Summary

Summary

- Unifying Everything: Factor Graph
  - A bipartite representing a distribution
  - Variable node and function node
- Inference
  - Sum-product message passing
  - Max-product message passing
- Learning

References

- https://ermongroup.github.io/cs228-notes/inference/jt/
- Kschischang et al. Factor Graphs and the Sum-Product Algorithm. IEEE Transactions on Information Theory. 2001.

Guidelines For Presentation

- For presenters
  - Send slides and HW questions with answers to me one day ahead (post slides and HW questions on Piazza)
  - Presentation should
    1. With enough background knowledge
    2. Well structured and organized
    3. With enough details and the methodologies are clearly explained
    4. With running examples
    5. Summarize main messages delivered in the paper(s), and your thoughts
    6. With references
  - Handle questions
  - Give homework and provide solution in class
  - Grading homework

Guidelines For Presentation

- For other students
  - browse the paper before hand
  - Raise question and participate in the discussion
  - Finish homework in class
  - Rate the presentation