

Learning Objectives

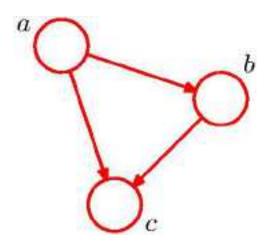
- 1、What are Bayesian Networks (BNs)?
- 2. How to use BNs to represent curve fitting, HMM, linear Gaussian models?
- 3. What is conditional independence?
- 4. What are Markov random fields?
- 5. What are directed, undirected and factor graphs?
- 6. How to perform sum-product algorithms within factor graphs?
- 7. How to perform max-product algorithms within factor graphs?

Outlines

- Bayesian Networks
- Bayesian Curve Fitting
- Discrete Variables and Linear Gaussian Models
- Conditional Independence
- Markov Random Fields
- Inference in Graphical Models

Bayesian Networks

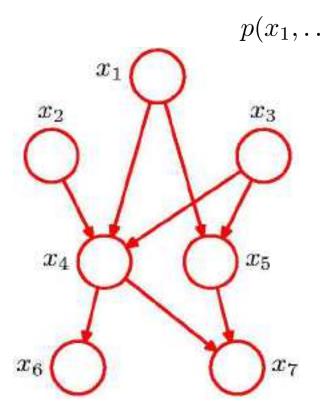
Directed Acyclic Graph (DAG)



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

$$p(x_1, \dots, x_K) = p(x_K | x_1, \dots, x_{K-1}) \dots p(x_2 | x_1) p(x_1)$$

Bayesian Networks



$$p(x_1, \dots, x_7) = p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)$$
$$p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)$$

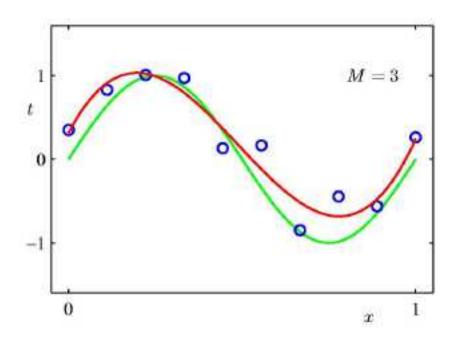
General Factorization

$$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k | \mathrm{pa}_k)$$

Outlines

- Bayesian Networks
- Bayesian Curve Fitting
- Discrete Variables and Linear Gaussian Models
- Conditional Independence
- Markov Random Fields
- Inference in Graphical Models

Bayesian Curve Fitting (1)



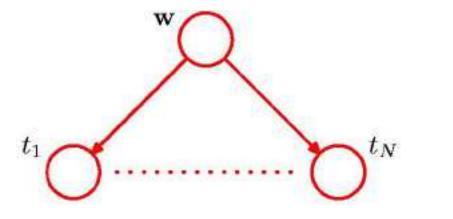
Polynomial

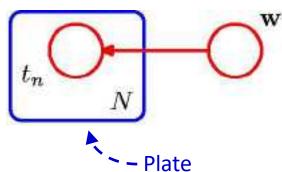
$$y(x, \mathbf{w}) = \sum_{j=0}^{M} w_j x^j$$

$$p(\mathbf{t}, \mathbf{w}) = p(\mathbf{w}) \prod_{n=1}^{N} p(t_n | y(\mathbf{w}, x_n))$$

Bayesian Curve Fitting (2)

$$p(\mathbf{t}, \mathbf{w}) = p(\mathbf{w}) \prod_{n=1}^{N} p(t_n | y(\mathbf{w}, x_n))$$

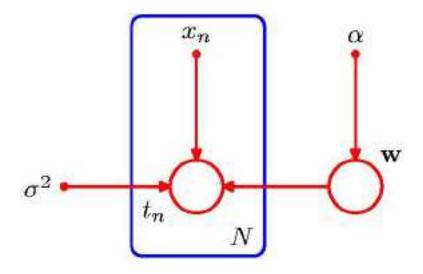




Bayesian Curve Fitting (3)

Input variables and explicit hyperparameters

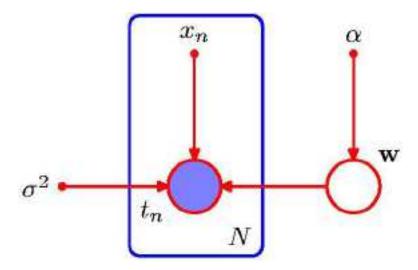
$$p(\mathbf{t}, \mathbf{w} | \mathbf{x}, \alpha, \sigma^2) = p(\mathbf{w} | \alpha) \prod_{n=1}^{N} p(t_n | \mathbf{w}, x_n, \sigma^2).$$



Bayesian Curve Fitting—Learning

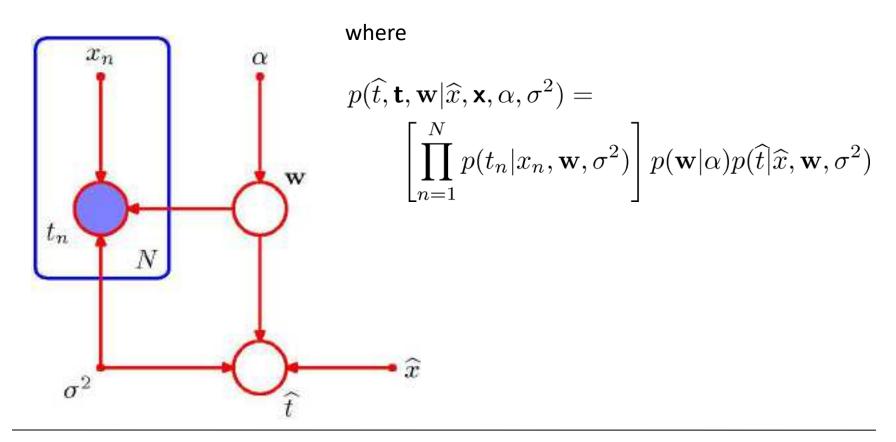
Condition on data

$$p(\mathbf{w}|\mathbf{t}) \propto p(\mathbf{w}) \prod_{n=1}^{N} p(t_n|\mathbf{w})$$



Bayesian Curve Fitting—Prediction

Predictive distribution: $p(\widehat{t}|\widehat{x}, \mathbf{x}, \mathbf{t}, \alpha, \sigma^2) \propto \int p(\widehat{t}, \mathbf{t}, \mathbf{w}|\widehat{x}, \mathbf{x}, \alpha, \sigma^2) d\mathbf{w}$

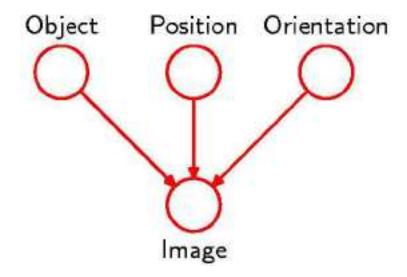


Outlines

- Bayesian Networks
- Bayesian Curve Fitting
- Discrete Variables and Linear Gaussian Models
- Conditional Independence
- Markov Random Fields
- Inference in Graphical Models

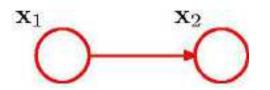
Generative Models

Causal process for generating images



Discrete Variables (1)

lacksquare General joint distribution: K^2-1 parameters



$$p(\mathbf{x}_1, \mathbf{x}_2 | \boldsymbol{\mu}) = \prod_{k=1}^K \prod_{l=1}^K \mu_{kl}^{x_{1k} x_{2l}}$$

lacksquare Independent joint distribution: 2(K-1) parameters

$$\bigcirc$$

$$\hat{p}(\mathbf{x}_1, \mathbf{x}_2 | \boldsymbol{\mu}) = \prod_{k=1}^K \mu_{1k}^{x_{1k}} \prod_{l=1}^K \mu_{2l}^{x_{2l}}$$

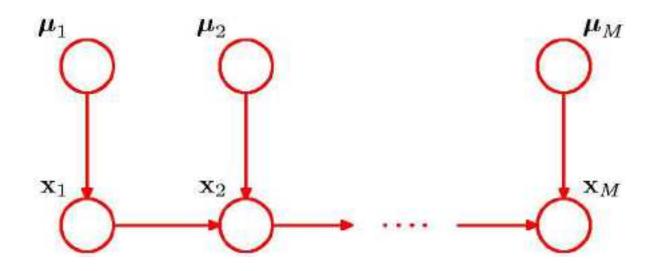
Discrete Variables (2)

lacksquare General joint distribution over M variables: K^M-1 parameters

lacksquare M-node Markov chain: K-1+(M-1)K(K-1) parameters



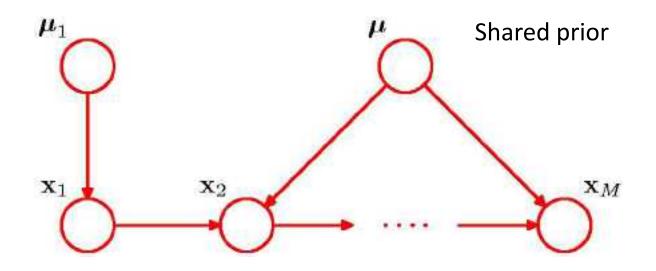
Discrete Variables: Bayesian Parameters (1)



$$p(\{\mathbf{x}_m, \boldsymbol{\mu}_m\}) = p(\mathbf{x}_1 | \boldsymbol{\mu}_1) p(\boldsymbol{\mu}_1) \prod_{m=2}^{M} p(\mathbf{x}_m | \mathbf{x}_{m-1}, \boldsymbol{\mu}_m) p(\boldsymbol{\mu}_m)$$

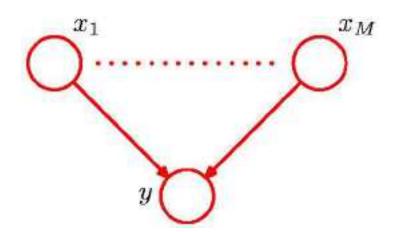
$$p(\boldsymbol{\mu}_m) = \operatorname{Dir}(\boldsymbol{\mu}_m | \boldsymbol{\alpha}_m)$$

Discrete Variables: Bayesian Parameters (2)



$$p(\{\mathbf{x}_m\}, \boldsymbol{\mu}_1, \boldsymbol{\mu}) = p(\mathbf{x}_1 | \boldsymbol{\mu}_1) p(\boldsymbol{\mu}_1) \prod_{m=2}^{M} p(\mathbf{x}_m | \mathbf{x}_{m-1}, \boldsymbol{\mu}) p(\boldsymbol{\mu})$$

Parameterized Conditional Distributions



If x_1,\ldots,x_M are discrete, K-state variables, $p(y=1|x_1,\ldots,x_M)$ in general has $O(K^M)$ parameters.

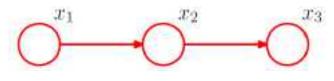
The parameterized form

$$p(y = 1 | x_1, \dots, x_M) = \sigma\left(w_0 + \sum_{i=1}^M w_i x_i\right) = \sigma(\mathbf{w}^T \mathbf{x})$$

requires only M+1 parameters

Linear-Gaussian Models

Directed Graph



$$p(x_i|pa_i) = \mathcal{N}\left(x_i \left| \sum_{j \in pa_i} w_{ij}x_j + b_i, v_i \right)\right)$$

Each node is Gaussian, the mean is a linear function of the parents.

Vector-valued Gaussian Nodes

$$p(\mathbf{x}_i|\mathrm{pa}_i) = \mathcal{N}\left(\mathbf{x}_i\left|\sum_{j\in\mathrm{pa}_i}\mathbf{W}_{ij}\mathbf{x}_j + \mathbf{b}_i, \mathbf{\Sigma}_i\right.
ight)$$

Outlines

- Bayesian Networks
- Bayesian Curve Fitting
- Discrete Variables and Linear Gaussian Models
- Conditional Independence
- Markov Random Fields
- Inference in Graphical Models

Conditional Independence

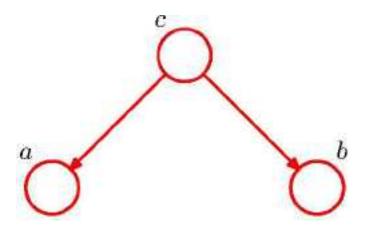
a is independent of b given c

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

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

Notation

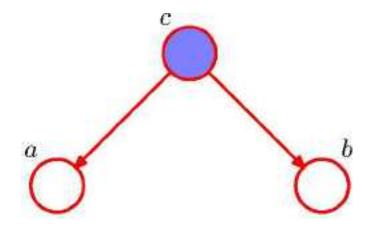
$$a \perp \!\!\!\perp b \mid c$$



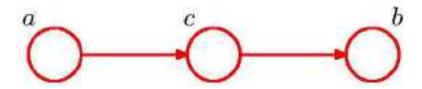
$$p(a, b, c) = p(a|c)p(b|c)p(c)$$

$$p(a,b) = \sum_{c} p(a|c)p(b|c)p(c)$$

$$a \not\perp \!\!\!\perp b \mid \emptyset$$



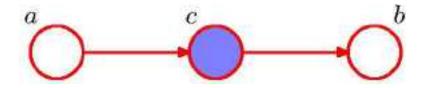
$$p(a,b|c) = \frac{p(a,b,c)}{p(c)}$$
$$= p(a|c)p(b|c)$$
$$a \perp \perp b \mid c$$



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

$$p(a,b) = p(a) \sum_{c} p(c|a)p(b|c) = p(a)p(b|a)$$

$$a \not\perp \!\!\!\perp b \mid \emptyset$$

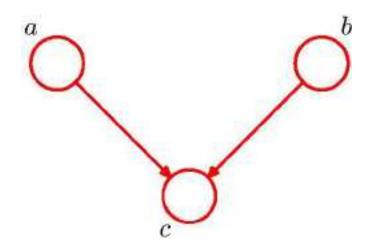


$$p(a,b|c) = \frac{p(a,b,c)}{p(c)}$$

$$= \frac{p(a)p(c|a)p(b|c)}{p(c)}$$

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

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

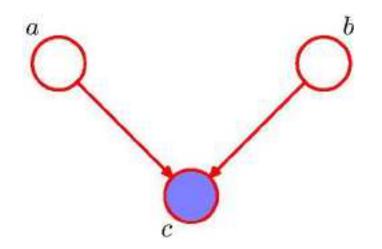


$$p(a,b,c) = p(a)p(b)p(c|a,b)$$

$$p(a,b) = p(a)p(b)$$

$$a \perp \!\!\!\perp b \mid \emptyset$$

Note: this is the opposite of Example 1, with c unobserved.



$$p(a,b|c) = \frac{p(a,b,c)}{p(c)}$$

$$= \frac{p(a)p(b)p(c|a,b)}{p(c)}$$

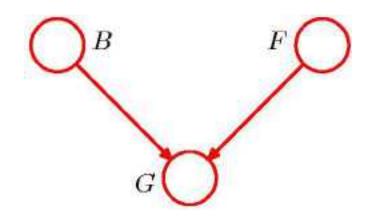
$$a \not\perp \!\!\!\perp b \mid c$$

Note: this is the opposite of Example 1, with c observed.

"Am I out of fuel?"

$$p(G = 1|B = 1, F = 1) = 0.8$$

 $p(G = 1|B = 1, F = 0) = 0.2$
 $p(G = 1|B = 0, F = 1) = 0.2$
 $p(G = 1|B = 0, F = 0) = 0.1$



$$p(B=1) = 0.9$$

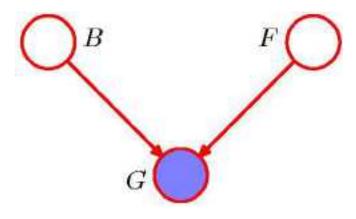
 $p(F=1) = 0.9$
and hence
 $p(F=0) = 0.1$

$$B = Battery$$
 (0=flat, 1=fully charged)
 $F = Fuel Tank$ (0=empty, 1=full)

G = Fuel Gauge Reading

(0=empty, 1=full)

"Am I out of fuel?"

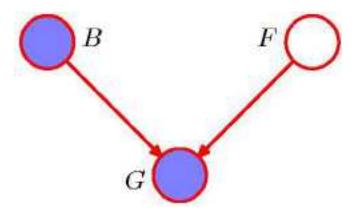


$$p(F = 0|G = 0) = \frac{p(G = 0|F = 0)p(F = 0)}{p(G = 0)}$$

\$\sim 0.257\$

Probability of an empty tank increased by observing G=0.

"Am I out of fuel?"



$$p(F = 0|G = 0, B = 0) = \frac{p(G = 0|B = 0, F = 0)p(F = 0)}{\sum_{F \in \{0,1\}} p(G = 0|B = 0, F)p(F)}$$

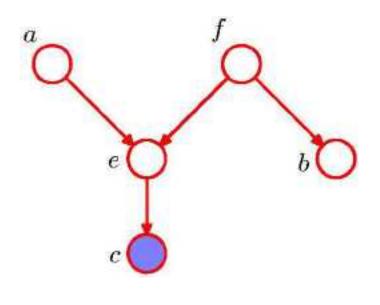
$$\simeq 0.111$$

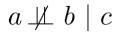
Probability of an empty tank reduced by observing B=0. This referred to as "explaining away".

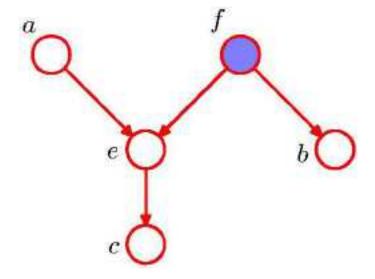
D-separation

- lacksquare A, B, and C are non-intersecting subsets of nodes in a directed graph.
- lacksquare A path from A to B is blocked if it contains a node such that either
 - \checkmark the arrows on the path meet either head-to-tail or tail-to-tail at the node, and the node is in the set C, or
 - ✓ the arrows meet head-to-head at the node, and neither the node, nor any of its descendants, are in the set C.
- lacksquare If all paths from A to B are blocked, A is said to be d-separated from B by C.
- lacksquare If A is d-separated from B by C, the joint distribution over all variables in the graph satisfies $A \perp\!\!\!\perp B \mid C$.

D-separation: Example

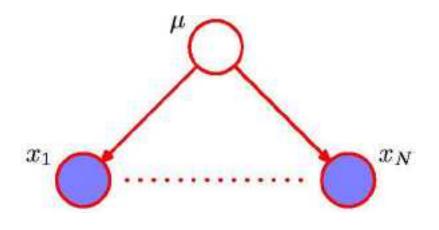






 $a \perp \!\!\! \perp b \mid f$

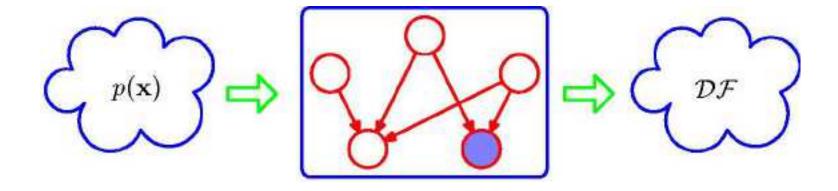
D-separation: I.I.D. Data



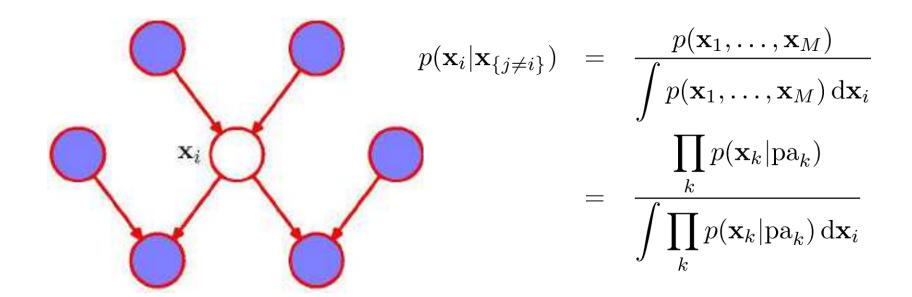
$$p(\mathcal{D}|\mu) = \prod_{n=1}^{N} p(x_n|\mu)$$

$$p(\mathcal{D}) = \int_{-\infty}^{\infty} p(\mathcal{D}|\mu) p(\mu) d\mu \neq \prod_{n=1}^{N} p(x_n)$$

Directed Graphs as Distribution Filters



The Markov Blanket

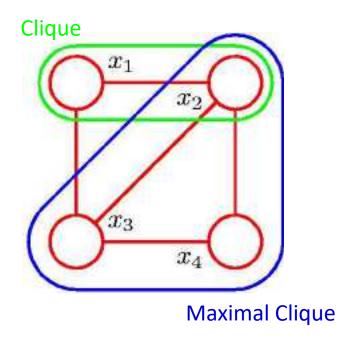


Factors independent of x_i cancel between numerator and denominator.

Outlines

- Bayesian Networks
- Bayesian Curve Fitting
- Discrete Variables and Linear Gaussian Models
- Conditional Independence
- Markov Random Fields
- Inference in Graphical Models

Cliques and Maximal Cliques



Joint Distribution

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{C} \psi_{C}(\mathbf{x}_{C})$$

where $\psi_C(\mathbf{x}_C)$ is the potential over clique C and

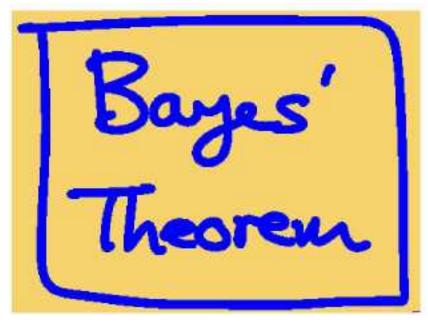
$$Z = \sum_{\mathbf{x}} \prod_{C} \psi_C(\mathbf{x}_C)$$

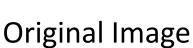
is the normalization coefficient; note: MK-state variables $\to K^M$ terms in Z.

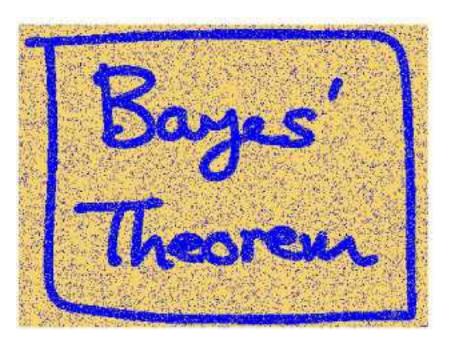
Energies and the Boltzmann distribution

$$\psi_C(\mathbf{x}_C) = \exp\left\{-E(\mathbf{x}_C)\right\}$$

Illustration: Image De-Noising (1)







Noisy Image

Illustration: Image De-Noising (2)

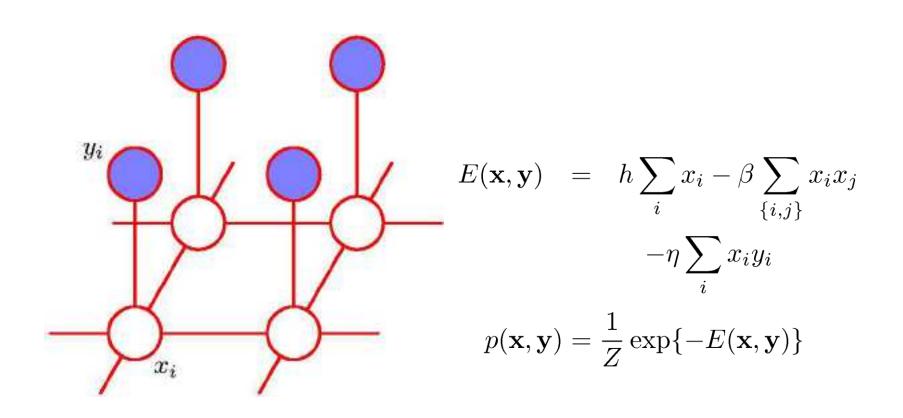
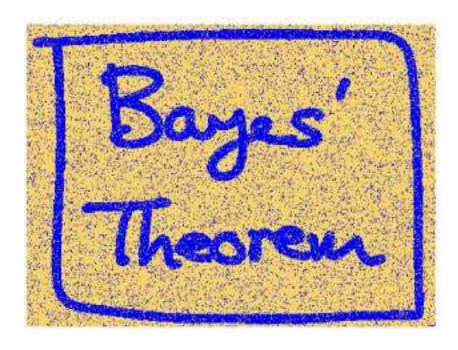


Illustration: Image De-Noising (3)

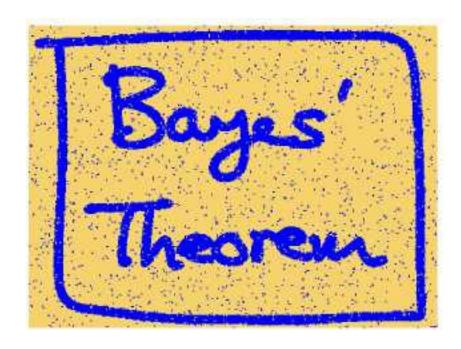




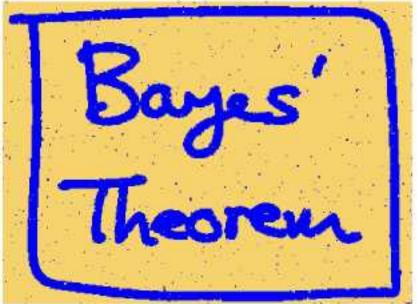


Restored Image (ICM)

Illustration: Image De-Noising (4)

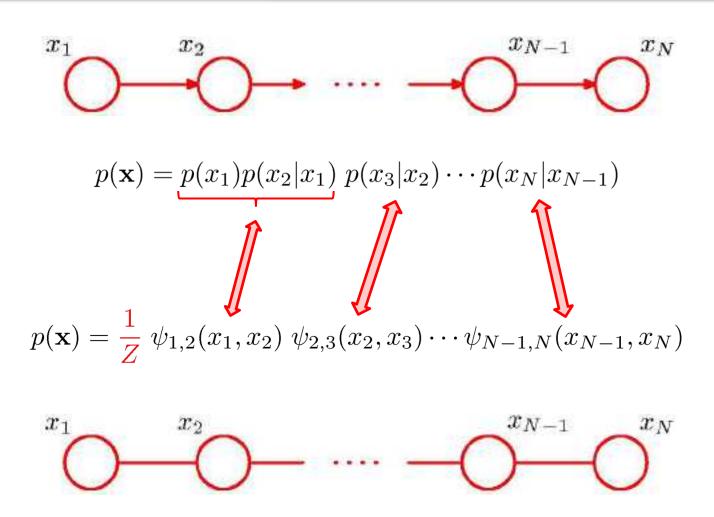


Restored Image (ICM)



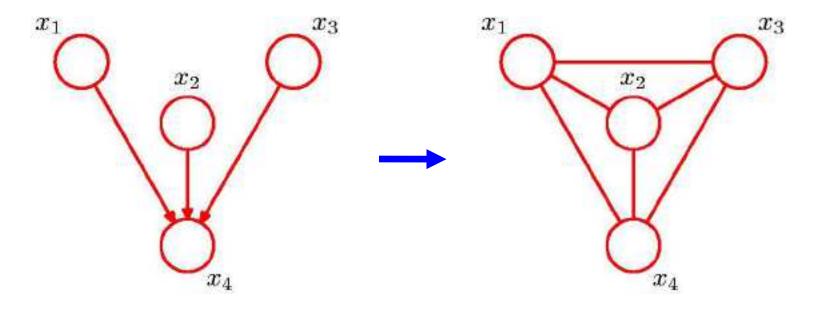
Restored Image (Graph cuts)

Converting Directed to Undirected Graphs (1)



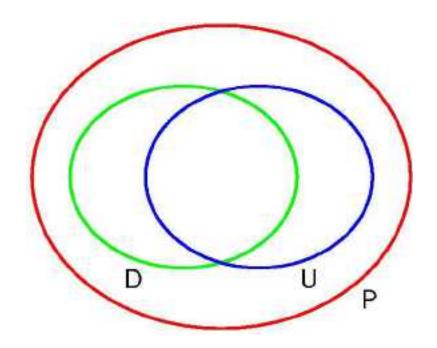
Converting Directed to Undirected Graphs (2)

Additional links

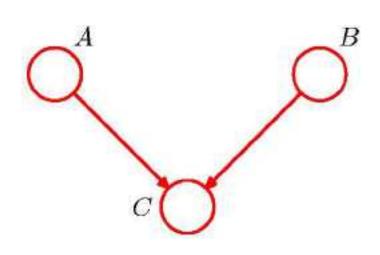


$$p(\mathbf{x}) = p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)$$
$$= \frac{1}{Z}\psi_A(x_1, x_2, x_3)\psi_B(x_2, x_3, x_4)\psi_C(x_1, x_2, x_4)$$

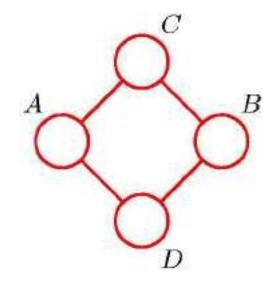
Directed vs. Undirected Graphs (1)



Directed vs. Undirected Graphs (2)



$$A \perp \!\!\!\perp B \mid \emptyset$$
$$A \perp \!\!\!\!\perp B \mid C$$



$$A \not\perp\!\!\!\perp B \mid \emptyset$$

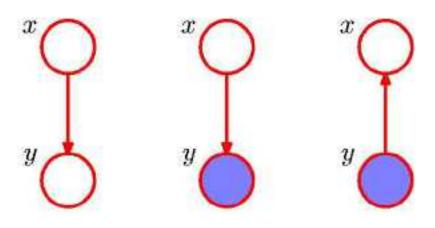
$$A \perp\!\!\!\perp B \mid C \cup D$$

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

Outlines

- Bayesian Networks
- Bayesian Curve Fitting
- Discrete Variables and Linear Gaussian Models
- Conditional Independence
- Markov Random Fields
- Inference in Graphical Models

Inference in Graphical Models

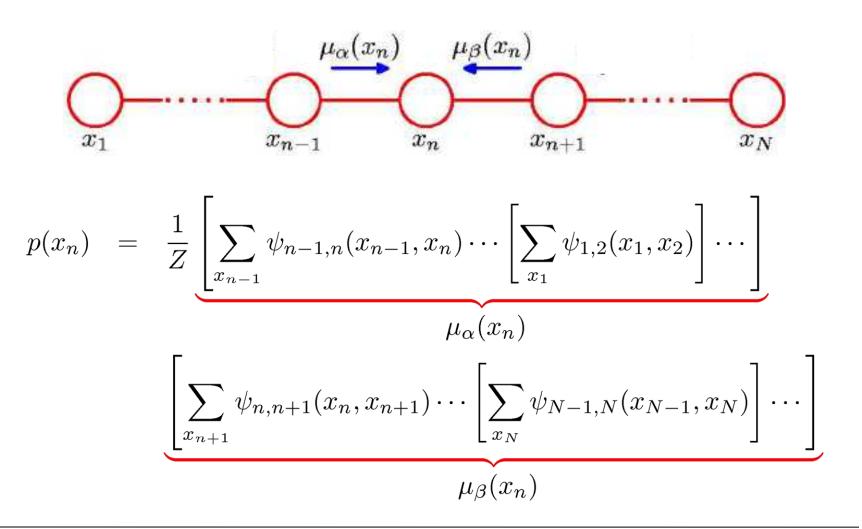


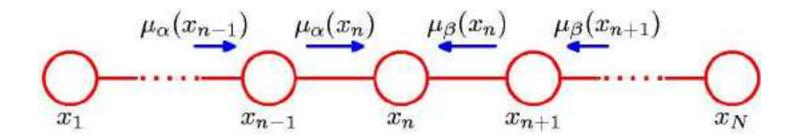
$$p(y) = \sum_{x'} p(y|x')p(x') \qquad p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$



$$p(\mathbf{x}) = \frac{1}{Z}\psi_{1,2}(x_1, x_2)\psi_{2,3}(x_2, x_3)\cdots\psi_{N-1,N}(x_{N-1}, x_N)$$

$$p(x_n) = \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_N} p(\mathbf{x})$$





$$\mu_{\alpha}(x_2) = \sum_{x_1} \psi_{1,2}(x_1, x_2) \qquad \mu_{\beta}(x_{N-1}) = \sum_{x_N} \psi_{N-1,N}(x_{N-1}, x_N)$$

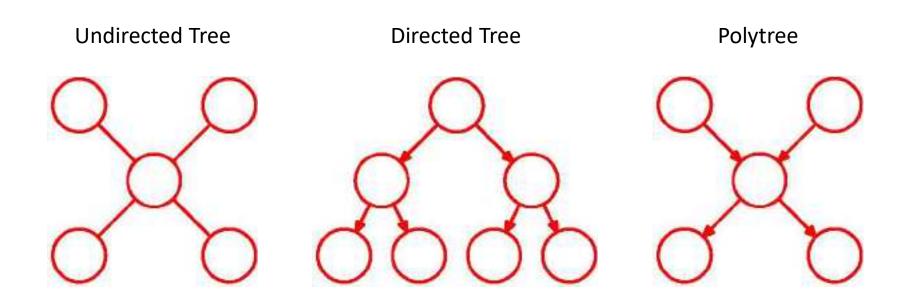
$$Z = \sum_{x_n} \mu_{\alpha}(x_n) \mu_{\beta}(x_n)$$

- ☐ To compute local marginals:
 - \checkmark Compute and store all forward messages, $\mu_{\alpha}(x_n)$
 - \checkmark Compute and store all backward messages, $\mu_{\beta}(x_n)$
 - \checkmark Compute Z at any node x_m
 - ✓ Compute

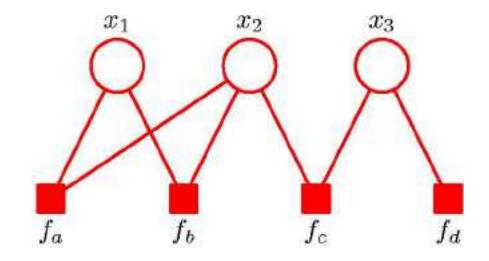
$$p(x_n) = \frac{1}{Z} \mu_{\alpha}(x_n) \mu_{\beta}(x_n)$$

for all variables required.

Trees



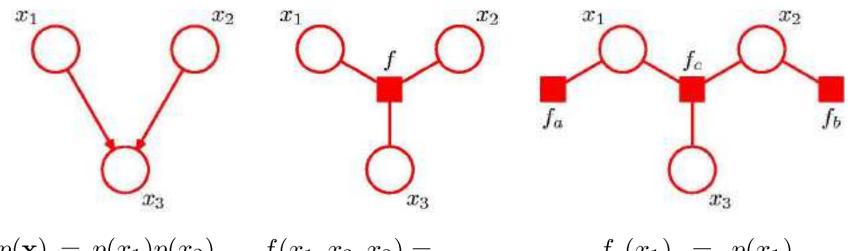
Factor Graphs



$$p(\mathbf{x}) = f_a(x_1, x_2) f_b(x_1, x_2) f_c(x_2, x_3) f_d(x_3)$$

$$p(\mathbf{x}) = \prod_{s} f_s(\mathbf{x}_s)$$

Factor Graphs from Directed Graphs

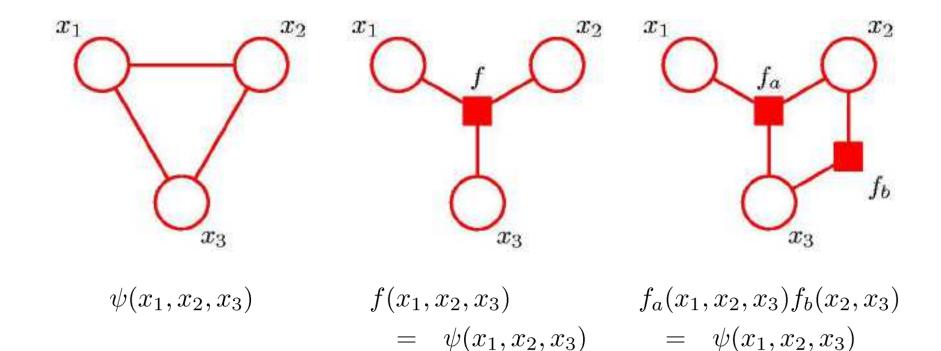


$$p(\mathbf{x}) = p(x_1)p(x_2) \qquad f(x_1, x_2, x_3) = \qquad \qquad f_a(x_1) = p(x_1)$$

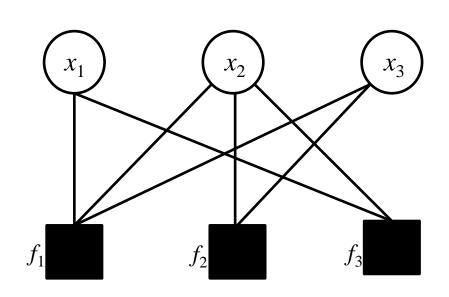
$$p(x_3|x_1, x_2) \qquad p(x_1)p(x_2)p(_3|x_1, x_2) \qquad \qquad f_b(x_2) = p(x_2)$$

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

Factor Graphs from Undirected Graphs



Factor Graph for Solving Equations



$$x_1 + 2x_2 + x_3 = 4$$
 $x_2 + 2x_3 = 3$

$$x_1 + x_2 = 2$$

(1)
$$x_1 = 1$$
 $x_2 = 1$ $x_3 = 0$

(2)
$$f_1 \rightarrow x_1$$
: $x_1 = 4 - 2x_2 - x_3 = 2$
 $f_1 \rightarrow x_2$: $x_2 = (4 - x_1 - x_3)/2 = 1.5$
 $f_1 \rightarrow x_3$: $x_3 = 4 - 2x_2 - x_1 = 1$

$$f_2 \rightarrow x_2$$
: $x_2 = 3 - 2x_3 = 3$
 $f_2 \rightarrow x_3$: $x_3 = 3 - 2x_2 = 1$

$$f_3 \rightarrow x_1$$
: $x_1 = 2 - x_2 = 1$
 $f_3 \rightarrow x_2$: $x_2 = 2 - x_1 = 1$

(3)
$$x_1 = (1+2+1)/3 = 4/3$$

$$x_2 = (1+1.5+3+1)/4 = 6.5/4$$
 $x_3 = (0+1+1)/3 = 2/3$

$$x_3 = (0+1+1)/3 = 2/3$$

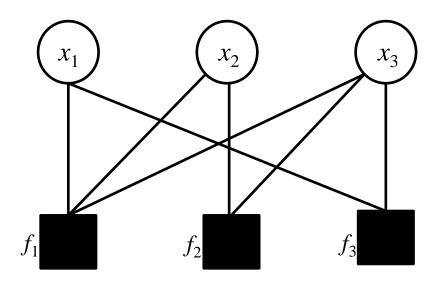
Factor Graph for Computing Means

$$S_i = x_i N_i$$

$$(1) S_1 = 11 N_1 = 10$$

$$S_1 = 11 \ N_1 = 10 \ S_2 = 10 \ N_2 = 10 \ S_3 = 18 \ N_3 = 20$$

$$S_3 = 18 \ N_3 = 20$$



$$x_1 = x_2 = x_3$$

$$x_2 = x_3$$

$$x_1 = x_3$$

(2)
$$f_1 \rightarrow x_1$$
: $S_1 = 28 N_1 = 30$

$$f_1 \rightarrow x_2$$
: $S_2 = 29 N_2 = 30$

$$f_1 \rightarrow x_3$$
: $S_3 = 21 N_3 = 20$

$$f_2 \rightarrow x_2$$
: $S_2 = 18 N_2 = 20$

$$f_2 \rightarrow x_3$$
: $S_3 = 10 N_3 = 10$

$$f_3 \rightarrow x_1$$
: $S_1 = 18 N_1 = 20$

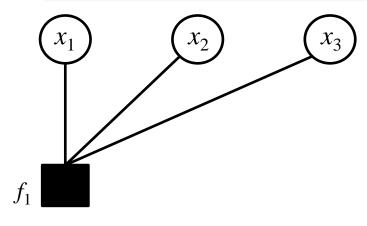
$$f_3 \rightarrow x_3$$
: $S_3 = 11 N_3 = 10$

(3)
$$S_1 = 57 \ N_1 = 60 \ S_2 = 57 \ N_2 = 60 \ S_3 = 60 \ N_3 = 60$$

$$S_2 = 57 \ N_2 = 60$$

$$S_3 = 60 \ N_3 = 60$$

Factor Graph for Belief Aggregation



$$\mathbf{x}_1 = x_2 = x_3$$

(1)
$$x_1 \sim \mathcal{N}(m_1, \Sigma_1)$$
 $x_2 \sim \mathcal{N}(m_2, \Sigma_2)$
(2) $f_1 \rightarrow x_1$: $x_3 \sim \mathcal{N}(m_3, \Sigma_3)$
 $\hat{\Sigma}_1^{-1} = \Sigma_2^{-1} + \Sigma_3^{-1}$ $\hat{\Sigma}_1^{-1} \hat{m}_1^{-1} = \Sigma_2^{-1} m_2^{-1} + \Sigma_3^{-1} m_3^{-1}$
 $f_1 \rightarrow x_2$:
 $\hat{\Sigma}_2^{-1} = \Sigma_1^{-1} + \Sigma_3^{-1}$ $\hat{\Sigma}_2^{-1} \hat{m}_2^{-1} = \Sigma_1^{-1} m_1^{-1} + \Sigma_3^{-1} m_3^{-1}$
 $f_1 \rightarrow x_3$:
 $\hat{\Sigma}_3^{-1} = \Sigma_1^{-1} + \Sigma_3^{-1}$ $\hat{\Sigma}_2^{-1} \hat{m}_2^{-1} = \Sigma_1^{-1} m_1^{-1} + \Sigma_3^{-1} m_3^{-1}$

$$\bar{\Sigma}_{1}^{-1} = \Sigma_{1}^{-1} + \hat{\Sigma}_{1}^{-1} \quad \bar{\Sigma}_{1}^{-1} \bar{m}_{1}^{-1} = \Sigma_{1}^{-1} m_{1}^{-1} + \hat{\Sigma}_{1}^{-1} \hat{m}_{1}^{-1}$$

$$\bar{\Sigma}_{2}^{-1} = \Sigma_{2}^{-1} + \hat{\Sigma}_{2}^{-1} \quad \bar{\Sigma}_{2}^{-1} \bar{m}_{2}^{-1} = \Sigma_{2}^{-1} m_{2}^{-1} + \hat{\Sigma}_{2}^{-1} \hat{m}_{2}^{-1}$$

$$\bar{\Sigma}_{3}^{-1} = \Sigma_{3}^{-1} + \hat{\Sigma}_{3}^{-1} \quad \bar{\Sigma}_{3}^{-1} \bar{m}_{3}^{-1} = \Sigma_{3}^{-1} m_{3}^{-1} + \hat{\Sigma}_{3}^{-1} \hat{m}_{3}^{-1}$$

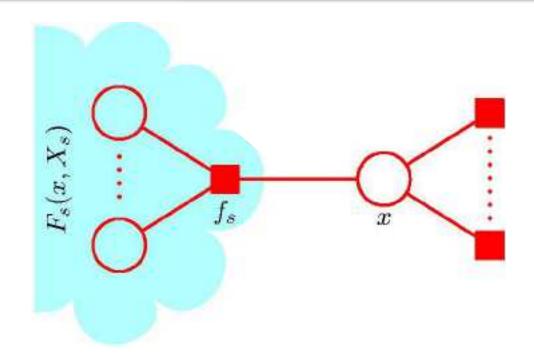
The Sum-Product Algorithm (1)

☐ Objective:

- i. to obtain an efficient, exact inference algorithm for finding marginals;
- ii. in situations where several marginals are required, to allow computations to be shared efficiently.
- Key idea: Distributive Law

$$ab + ac = a(b+c)$$

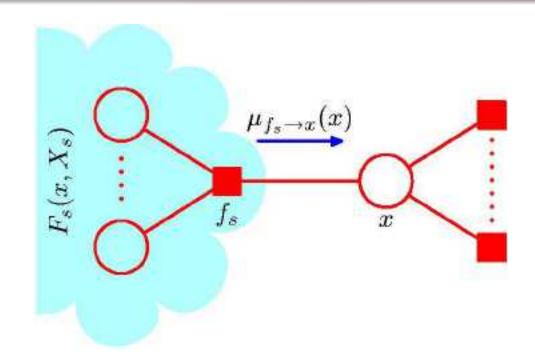
The Sum-Product Algorithm (2)



$$p(x) = \sum_{\mathbf{x} \setminus x} p(\mathbf{x})$$

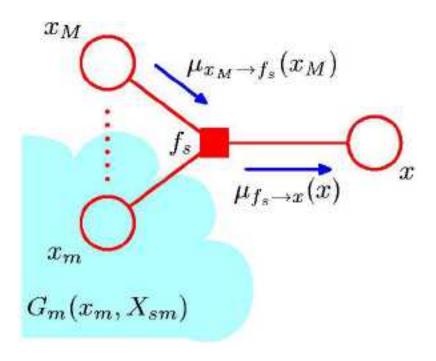
$$p(\mathbf{x}) = \prod_{s \in \text{ne}(x)} F_s(x, X_s)$$

The Sum-Product Algorithm (3)



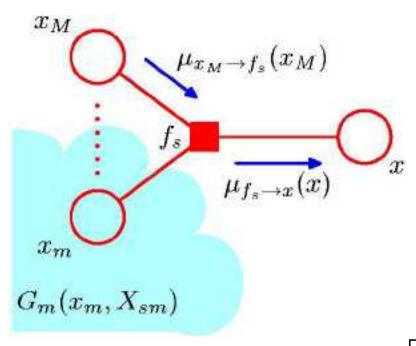
$$p(x) = \prod_{s \in ne(x)} \left[\sum_{X_s} F_s(x, X_s) \right]$$
$$= \prod_{s \in ne(x)} \mu_{f_s \to x}(x). \qquad \mu_{f_s \to x}(x) \equiv \sum_{X_s} F_s(x, X_s)$$

The Sum-Product Algorithm (4)



$$F_s(x, X_s) = f_s(x, x_1, \dots, x_M)G_1(x_1, X_{s1}) \dots G_M(x_M, X_{sM})$$

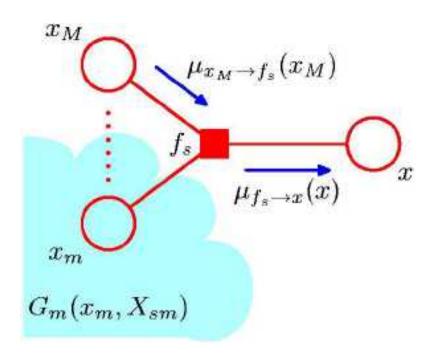
The Sum-Product Algorithm (5)



$$\mu_{f_s \to x}(x) = \sum_{x_1} \dots \sum_{x_M} f_s(x, x_1, \dots, x_M) \prod_{m \in \text{ne}(f_s) \setminus x} \left[\sum_{X_{sm}} G_m(x_m, X_{sm}) \right]$$

$$= \sum_{x_1} \dots \sum_{x_M} f_s(x, x_1, \dots, x_M) \prod_{m \in \text{ne}(f_s) \setminus x} \mu_{x_m \to f_s}(x_m)$$

The Sum-Product Algorithm (6)

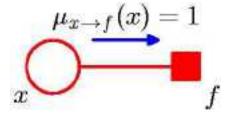


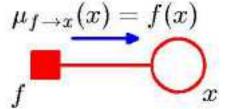
$$\mu_{x_m \to f_s}(x_m) \equiv \sum_{X_{sm}} G_m(x_m, X_{sm}) = \sum_{X_{sm}} \prod_{l \in \text{ne}(x_m) \setminus f_s} F_l(x_m, X_{ml})$$

$$= \prod_{l \in \text{ne}(x_m) \setminus f_s} \mu_{f_l \to x_m}(x_m)$$

The Sum-Product Algorithm (7)

Initialization

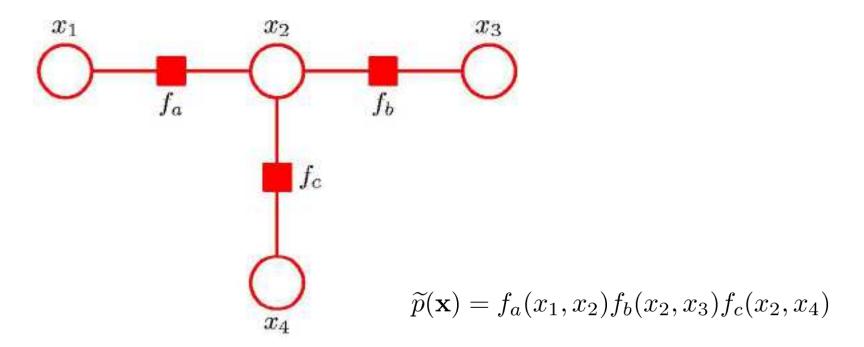




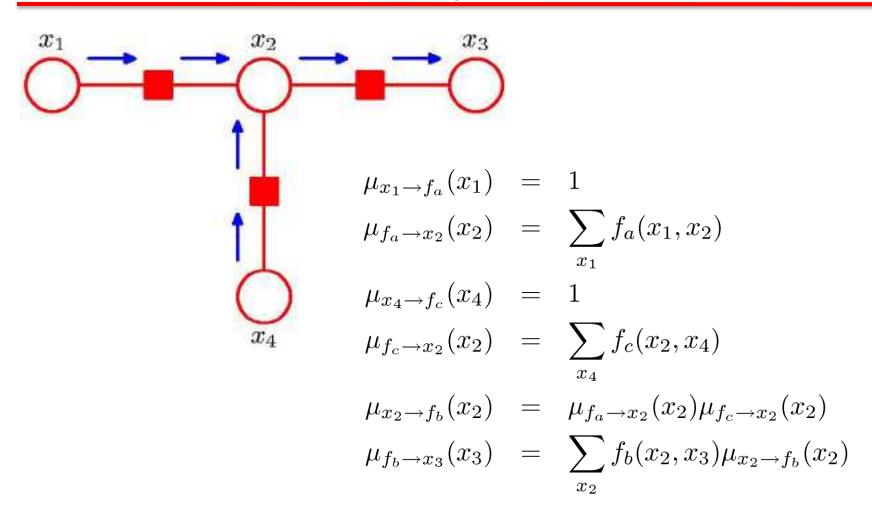
The Sum-Product Algorithm (8)

- ☐ To compute local marginals:
 - ✓ Pick an arbitrary node as root
 - ✓ Compute and propagate messages from the leaf nodes to the root, storing received messages at every node.
 - ✓ Compute and propagate messages from the root to the leaf nodes, storing received messages at every node.
 - ✓ Compute the product of received messages at each node for which the marginal is required, and normalize if necessary.

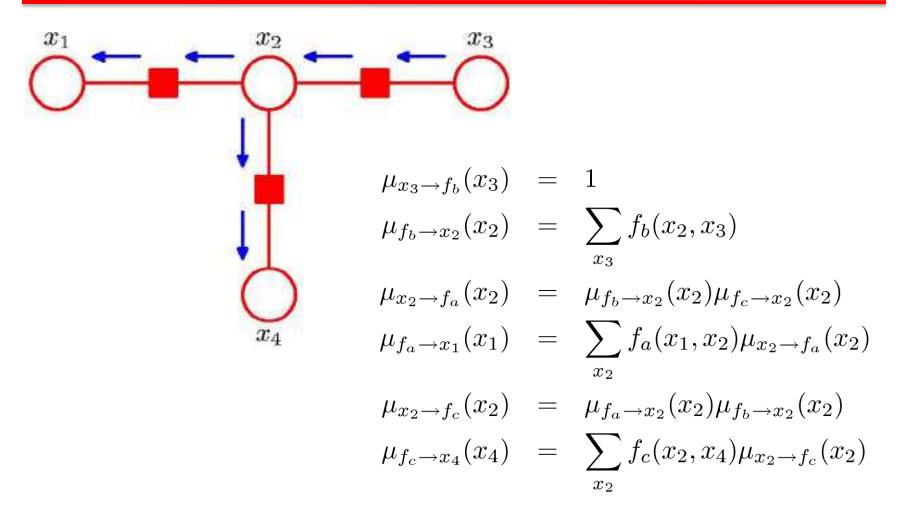
Sum-Product: Example (1)



Sum-Product: Example (2)



Sum-Product: Example (3)



Sum-Product: Example (4)

The Max-Sum Algorithm (1)

- ☐ Objective: an efficient algorithm for finding
 - i. the value \mathbf{x}^{\max} that maximizes $p(\mathbf{x})$;
 - ii. the value of $p(\mathbf{x}^{\text{max}})$.

In general, maximum marginals ≠ joint maximum.

$$\underset{x}{\arg\max} p(x,y) = 1 \qquad \underset{x}{\arg\max} p(x) = 0$$

The Max-Sum Algorithm (2)

■ Maximizing over a chain (max-product)



$$p(\mathbf{x}^{\max}) = \max_{\mathbf{x}} p(\mathbf{x}) = \max_{x_1} \dots \max_{x_M} p(\mathbf{x})$$

$$= \frac{1}{Z} \max_{x_1} \dots \max_{x_N} \left[\psi_{1,2}(x_1, x_2) \dots \psi_{N-1,N}(x_{N-1}, x_N) \right]$$

$$= \frac{1}{Z} \max_{x_1} \left[\max_{x_2} \left[\psi_{1,2}(x_1, x_2) \left[\dots \max_{x_N} \psi_{N-1,N}(x_{N-1}, x_N) \right] \dots \right] \right]$$

The Max-Sum Algorithm (3)

☐ Generalizes to tree-structured factor graph

$$\max_{\mathbf{x}} p(\mathbf{x}) = \max_{x_n} \prod_{f_s \in ne(x_n)} \max_{X_s} f_s(x_n, X_s)$$

maximizing as close to the leaf nodes as possible

The Max-Sum Algorithm (4)

- \square Max-Product \rightarrow Max-Sum
 - ✓ For numerical reasons, use

$$\ln\left(\max_{\mathbf{x}} p(\mathbf{x})\right) = \max_{\mathbf{x}} \ln p(\mathbf{x}).$$

✓ Again, use distributive law

$$\max(a+b, a+c) = a + \max(b, c).$$

The Max-Sum Algorithm (5)

■ Initialization (leaf nodes)

$$\mu_{x \to f}(x) = 0 \qquad \qquad \mu_{f \to x}(x) = \ln f(x)$$

■ Recursion

$$\mu_{f \to x}(x) = \max_{x_1, \dots, x_M} \left[\ln f(x, x_1, \dots, x_M) + \sum_{m \in \text{ne}(f_s) \setminus x} \mu_{x_m \to f}(x_m) \right]$$

$$\phi(x) = \arg \max_{x_1, \dots, x_M} \left[\ln f(x, x_1, \dots, x_M) + \sum_{m \in \text{ne}(f_s) \setminus x} \mu_{x_m \to f}(x_m) \right]$$

$$\mu_{x \to f}(x) = \sum_{l \in \text{ne}(x) \setminus f} \mu_{f_l \to x}(x)$$

The Max-Sum Algorithm (6)

□ Termination (root node)

$$p^{\max} = \max_{x} \left[\sum_{s \in \text{ne}(x)} \mu_{f_s \to x}(x) \right]$$

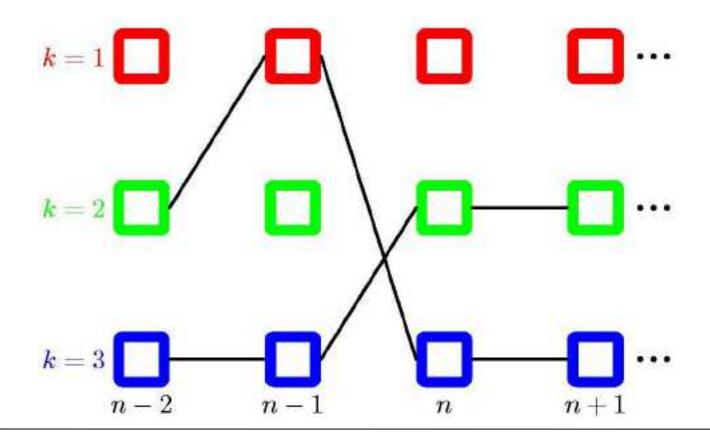
$$x^{\max} = \underset{x}{\operatorname{arg\,max}} \left[\sum_{s \in \operatorname{ne}(x)} \mu_{f_s \to x}(x) \right]$$

 \blacksquare Back-track, for all nodes i with l factor nodes to the root (l=0)

$$\mathbf{x}_l^{\max} = \phi(x_{i,l-1}^{\max})$$

The Max-Sum Algorithm (7)

Example: Markov chain



The Junction Tree Algorithm

- ☐ *Exact* inference on general graphs.
- Works by turning the initial graph into a *junction tree* and then running a sumproduct-like algorithm.
- ☐ *Intractable* on graphs with large cliques.

Loopy Belief Propagation

- Sum-Product on general graphs.
- Initial unit messages passed across all links, after which messages are passed around until convergence (not guaranteed!).
- ☐ Approximate but tractable for large graphs.
- ☐ Sometime works well, sometimes not at all.

Summary

- Bayesian Networks
- Bayesian Curve Fitting
- Discrete Variables and Linear Gaussian Models
- Conditional Independence
- Markov Random Fields
- Inference in Graphical Models