メッセージ伝播に関する補足

- 講義資料は、09BayesianNetworks.pdf ではなく09BayesianNetworks-rev.pdf を参照してください。
- 課題は簡単すぎて、かえって、分かりにくいかもしれません。
- 定義に従うより、本資料中の「ファクター木：分離」以下３枚のスライド（別途用意）のように、変形していく方が分かりやすいと思います。

確率伝播/メッセージ伝播

確率推論

ランダム変数 $X_1, \dots, X_N$ のうち $X_{m+1}, \dots, X_N$ の観測値 $a_{m+1}, \dots, a_N$ が与えられたとき、非観測値の周辺事後確率

$$\Pr(X_i = a | X_{m+1} = a_{m+1}, \dots, X_N = a_N), i = 1, \dots, m$$

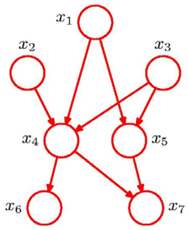
を計算すること

これを直接用いて、例えば $\Pr(X_1 = a)$ を求めようとすると

$$\Pr(X_i = a | X_{m+1} = a_{m+1}, \dots, X_N = a_N) = \alpha \sum_{x_2, \dots, x_m} p(a, x_2, \dots, x_m, a_{m+1}, \dots, a_N)$$

となるが、各 X_i が q 通りの値を取りうる場合およそ q^{m-1} 回の演算が必要となる

有向グラフ



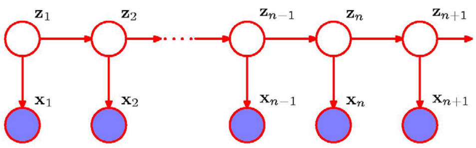
$$p(x_1, \dots, x_7) = p(x_1)p(x_2)p(x_3) \\ p(x_4|x_1, x_2, x_3)p(x_5|x_1, x_3) \\ p(x_6|x_4)p(x_7|x_4, x_5)$$

一般には

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

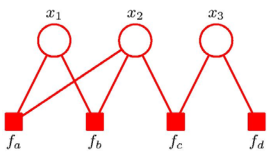
(C. Bishop 2009)

例: 時系列のモデリング



(C. Bishop 2009)

ファクターグラフ

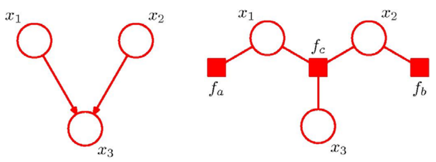


$$p(\mathbf{x}) = \prod_{s} f_s(\mathbf{x}_s)$$

(C. Bishop 2009)

有向グラフからファクターグラフへ

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



$$\begin{aligned} f_a(x_1) &= p(x_1) \\ f_b(x_2) &= p(x_2) \\ f_c(x_1, x_2, x_3) &= p(x_3|x_1, x_2) \end{aligned}$$

(C. Bishop 2009)

Sum-Product アルゴリズム (1)

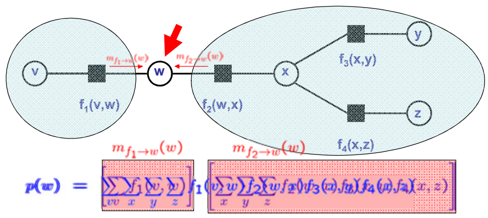
目的:

1. 周辺分布を見出す、効率的かつ（ある場合には）正確な推論アルゴリズムを得る
2. 複数個の周辺分布を求める時には、計算過程に共通部分があるようにし、その効率化が図れる.

アイデア: 分配側（ファクターへの分解）

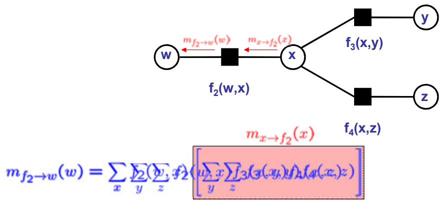
$$ab + ac = a(b + c)$$

ファクター木: 分離



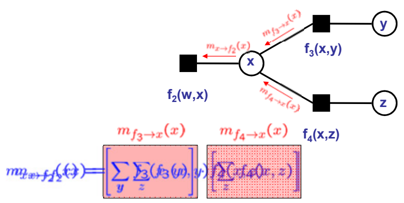
(C. Bishop 2009)

メッセージ: ファクターから変数へ



(C. Bishop 2009)

メッセージ: 変数からファクターへ



(C. Bishop 2009)

もう少し詳しく

有向グラフから無向グラフへ

$$p(\mathbf{x}) = \underbrace{p(x_1)p(x_2|x_1)}\ p(x_3|x_2)\cdots p(x_N|x_{N-1})$$

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

(C. Bishop PRML)

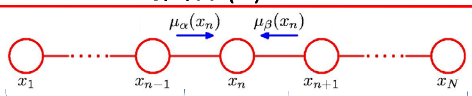
チェイン上の推論 (1)

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

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

(C. Bishop PRML)

チェイン上の推論 (2)



$$p(x_n) = \frac{1}{Z}\underbrace{\left[\sum_{x_{n-1}}\psi_{n-1,n}(x_{n-1},x_n)\cdots\left[\sum_{x_1}\psi_{1,2}(x_1,x_2)\right]\cdots\right]}_{\mu_\alpha(x_n)}$$

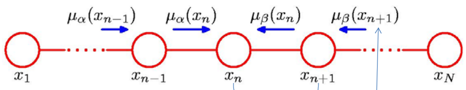
$$\underbrace{\left[\sum_{x_{n+1}}\psi_{n,n+1}(x_n,x_{n+1})\cdots\left[\sum_{x_N}\psi_{N-1,N}(x_{N-1},x_N)\right]\cdots\right]}_{\mu_\beta(x_n)}$$

注

$$\sum_x\sum_y\sum_z f(x)f(y)f(z) = \left(\sum_x f(x)\right)\left(\sum_y f(y)\right)\left(\sum_z f(z)\right)$$

(C. Bishop PRML)

チェイン上の推論 (3)



$$\mu_\alpha(x_n) = \sum_{x_{n-1}}\psi_{n-1,n}(x_{n-1},x_n)\left[\sum_{x_{n-2}}\cdots\right]$$

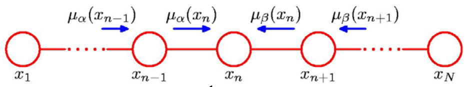
$$= \sum_{x_{n-1}}\psi_{n-1,n}(x_{n-1},x_n)\mu_\alpha(x_{n-1}).$$

$$\mu_\beta(x_n) = \sum_{x_{n+1}}\psi_{n,n+1}(x_n,x_{n+1})\left[\sum_{x_{n+2}}\cdots\right]$$

$$= \sum_{x_{n+1}}\psi_{n,n+1}(x_n,x_{n+1})\mu_\beta(x_{n+1}).$$

(C. Bishop PRML)

チェイン上の推論 (4)



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

$$\mu_\alpha(x_n) = \sum_{x_{n-1}}\psi_{n-2,n-1}(x_{n-1},x_n)\mu_\alpha(x_{n-1})$$

$$\mu_\alpha(x_{n-1}) = \sum_{x_{n-2}}\psi_{n-2,n-1}(x_{n-2},x_{n-1})\mu_\alpha(x_{n-2})$$

$$\mu_\beta(x_n) = \sum_{x_{n+1}}\psi_{n,n+1}(x_n,x_{n+1})\mu_\beta(x_{n+1})$$

$$\mu_\beta(x_{n+1}) = \sum_{x_{n+2}}\psi_{n+1,n+2}(x_{n+1},x_{n+})\mu_\beta(x_{n+2})$$

$$\mu_\alpha(x_2) = \sum_{x_1}\psi_{1,2}(x_1,x_2)$$

$$\mu_\beta(x_{N-1}) = \sum_{x_N}\psi_{N-1,N}(x_{N-1},x_N)$$

(C. Bishop PRML)

チェイン上の推論 (5)

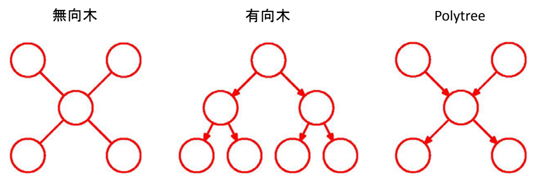
（局所的に）周辺分布を求めるには:

- 前進メッセージをすべて求め記憶する　$\mu_\alpha(x_n)$
- 後退メッセージをすべて求め記憶する　$\mu_\beta(x_n)$
- 各ノード x_m で規格化定数 Z を求める
- 必要なすべての変数につき、次を求める

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

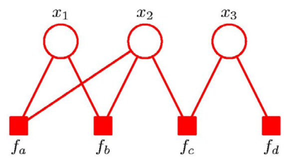
(C. Bishop PRML)

木



(C. Bishop PRML)

ファクターグラフ

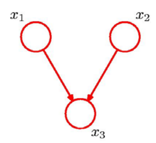


$$p(\mathbf{x}) = f_a(x_1, x_2) f_b(x_1, x_2) f_c(x_2, x_3) f_d(x_3)$$

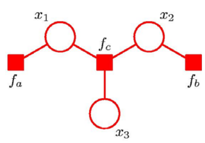
$$p(\mathbf{x}) = \prod_s f_s(\mathbf{x}_s)$$

(C. Bishop PRML)

有向グラフからファクターグラフへ

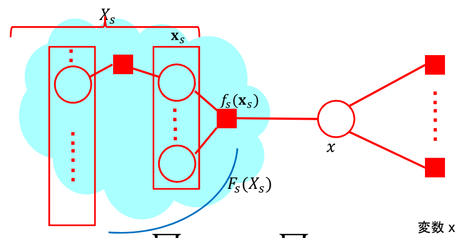


$$p(\mathbf{x}) = p(x_1)p(x_2) p(x_3|x_1, x_2)$$



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

Sum-Product アルゴリズム (2)

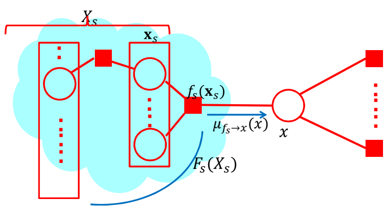


$$p(\mathbf{x}) = \prod_s f_s(\mathbf{x}_s) = \prod_{s \in N(x)} F_s(X_s)$$

$$p(x) = \sum_{\mathbf{x} \backslash x} p(\mathbf{x}) = \prod_{s \in N(x)} \sum_{X_s \backslash x} F_s(X_s)$$

変数 x を端として、変数・ファクタを排他的に分割することができる。その変数を Xs とし、該当するファクタの積を Fs とする

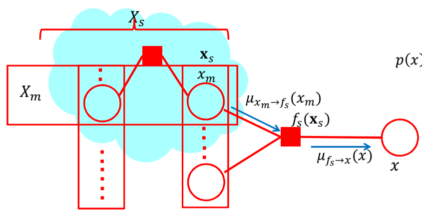
Sum-Product アルゴリズム (3)



$$p(x) = \prod_{s \in N(x)} \sum_{X_s \backslash x} F_s(X_s) = \prod_{s \in N(x)} \mu_{f_s \to x}(x)$$

$$\mu_{f_s \to x}(x) = \sum_{X_s \backslash x} F_s(X_s)$$

Sum-Product アルゴリズム (4)



$$p(x) = \prod_{s \in N(x)} \mu_{f_s \to x}(x)$$

$$\mu_{f_s \to x}(x) = \sum_{X_s \backslash x} F_s(X_s)$$
$$= \sum_{\mathbf{x}_s \backslash x} f_s(\mathbf{x}_s) \prod_{m \in N(f_s) \backslash x} \sum_{X_m} G_m(X_m)$$
$$= \sum_{\mathbf{x}_s \backslash x} f_s(\mathbf{x}_s) \prod_{m \in N(f_s) \backslash x} \mu_{s_m \to f_s}(x_m)$$

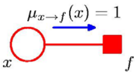
ファクタ fs を端として、Fs中のファクタ・Xs中の変数を排他的に分割することができる。その変数を Xm とし、該当するファクタの積を Gm とする

$$\mu_{x_m \to f_s}(x_m) = \sum_{X_m} G_m(X_m)$$

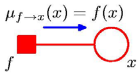
Sum-Product アルゴリズム (5)

初期化

$\mu_{x \to f}(x) = 1$



$\mu_{f \to x}(x) = f(x)$

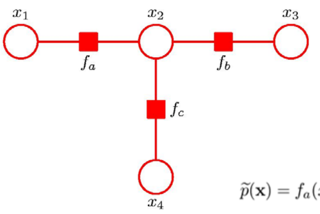


Sum-Product アルゴリズム (6)

（局所的な）周辺分布を計算するには:

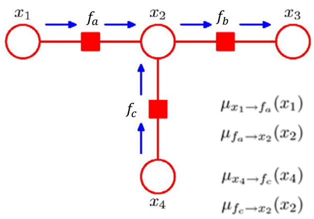
- 任意の変数を根とする木を考える
- 葉から根へとメッセージを計算し伝播させる. その過程で各ノードでメッセージを記憶する.
- 根から葉へとメッセージを計算し伝播させる. その過程で各ノードでメッセージを記憶する.
- 周辺分布を求めたいノード毎、メッセージの積を計算する。必要なら規格化（正規化）する.

例 (1)



$$\widetilde{p}(\mathbf{x}) = f_a(x_1, x_2) f_b(x_2, x_3) f_c(x_2, x_4)$$

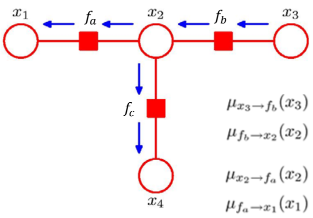
例 (2)



x3 を根としてみよう

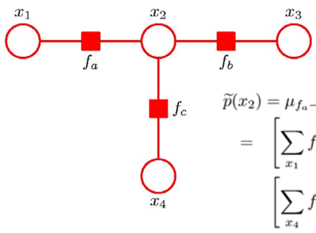
$$\begin{aligned}
\mu_{x_1 \to f_a}(x_1) &= 1 \\
\mu_{f_a \to x_2}(x_2) &= \sum_{x_1} f_a(x_1, x_2) \\
\mu_{x_4 \to f_c}(x_4) &= 1 \\
\mu_{f_c \to x_2}(x_2) &= \sum_{x_4} f_c(x_2, x_4) \\
\mu_{x_2 \to f_b}(x_2) &= \mu_{f_a \to x_2}(x_2) \mu_{f_c \to x_2}(x_2) \\
\mu_{f_b \to x_3}(x_3) &= \sum_{x_2} f_b(x_2, x_3) \mu_{x_2 \to f_b}(x_2)
\end{aligned}$$

例 (3)



$$\begin{aligned}
\mu_{x_3 \to f_b}(x_3) &= 1 \\
\mu_{f_b \to x_2}(x_2) &= \sum_{x_3} f_b(x_2, x_3) \\
\mu_{x_2 \to f_a}(x_2) &= \mu_{f_b \to x_2}(x_2) \mu_{f_c \to x_2}(x_2) \\
\mu_{f_a \to x_1}(x_1) &= \sum_{x_2} f_a(x_1, x_2) \mu_{x_2 \to f_a}(x_2) \\
\mu_{x_2 \to f_c}(x_2) &= \mu_{f_a \to x_2}(x_2) \mu_{f_b \to x_2}(x_2) \\
\mu_{f_c \to x_4}(x_4) &= \sum_{x_2} f_c(x_2, x_4) \mu_{x_2 \to f_c}(x_2)
\end{aligned}$$

例 (4)



$$\begin{aligned}
\widetilde{p}(x_2) &= \mu_{f_a \to x_2}(x_2) \mu_{f_b \to x_2}(x_2) \mu_{f_c \to x_2}(x_2) \\
&= \left[\sum_{x_1} f_a(x_1, x_2)\right] \left[\sum_{x_3} f_b(x_2, x_3)\right] \\
&\quad \left[\sum_{x_4} f_c(x_2, x_4)\right] \\
&= \sum_{x_1} \sum_{x_3} \sum_{x_4} f_a(x_1, x_2) f_b(x_2, x_3) f_c(x_2, x_4) \\
&= \sum_{x_1} \sum_{x_3} \sum_{x_4} \widetilde{p}(\mathbf{x})
\end{aligned}$$