第8章　グラフィカルモデル

Pattern Recognition and Machine Learning

修士２年
浦田　淳司

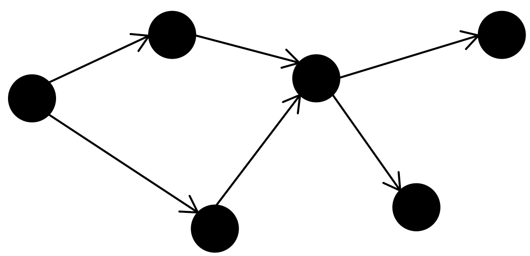
BehaviorInNetworks
都市生活学・ネットワーク行動学研究室

目次

グラフィカルモデルの特徴

A）確率モデルの構造を視覚化する簡単な方法の提供
→新しいモデルの設計方針
→条件つき独立性などのモデルの性質に関する知見

B）精巧なモデルにおける推論・学習には複雑な計算が必要
→数学的な表現をグラフ上の操作として表現

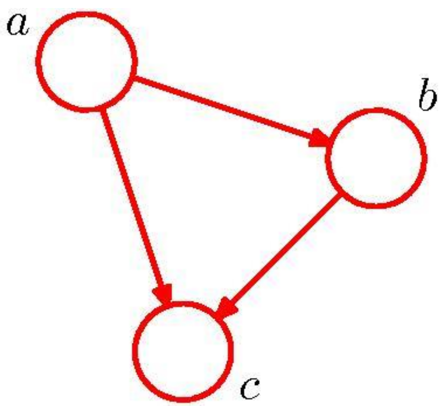


リンク：link, edge, arc
ノード：node, vertex

リンクが特定の方向性ー有向グラフィカルモデル（**ベイジアンネットワーク**）
→確率変数間の因果関係
リンクが方向性なし　ー無向グラフィカルモデル（**マルコフ確率場**）
→確率変数間の緩い束縛関係

※有向閉回路なし（非循環）

8.1 ベイジアンネットワーク



$$p(a,b,c) = p(c\,|\,a,b)\,p(b\,|\,a)\,p(a) \quad \cdots(8.2)$$

（aはbの親ノード⇔bはaの子ノード）

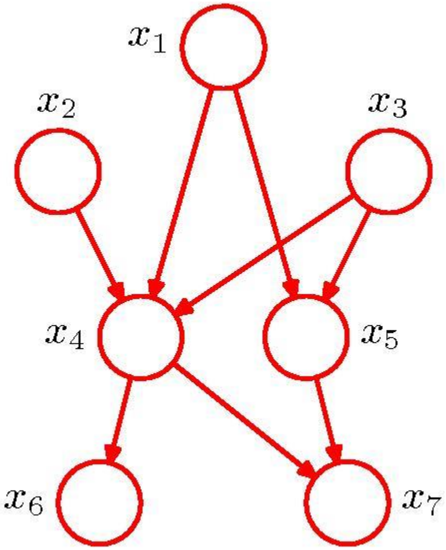
K変数の同時分布p(x_1,…,x_K）は，確率の乗法定理より

$$p(x_1,\cdots,x_K) = p(x_K\,|\,x_1,\cdots,x_{K-1})\cdots p(x_2\,|\,x_1)\,p(x_1) \quad \cdots(8.3)$$

自分より小さい番号を振られた全てのノードからのリンクを持つ

‥全結合

8.1 ベイジアンネットワーク



7変数全ての同時分布は

$$p(x_1)p(x_2)p(x_3)p(x_4 \mid x_1, x_2, x_3)p(x_5 \mid x_1, x_3)p(x_6 \mid x_4)p(x_7 \mid x_4, x_5) \quad \cdots(8.4)$$

同時分布の一般系

$$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k \mid pa_k) \quad \cdots(8.5)$$

pa_k : x_kの親ノードの集合

同時分布を，各ノードと対応する変数集合の形に分解

8.1.1 例）多項式フィッティング

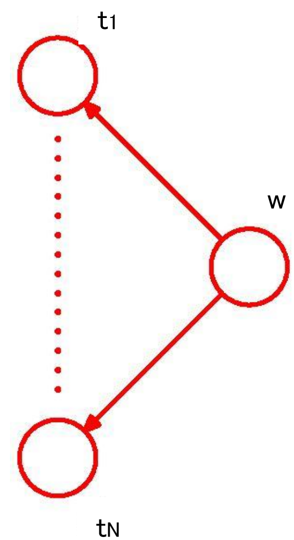
●有向グラフの利用方法

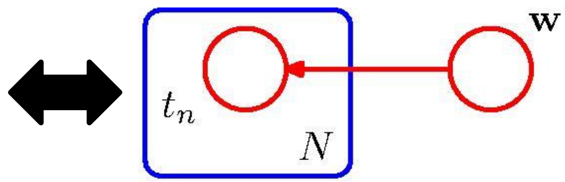
確率変数

・多項式係数ベクトル **w**

・観測データ **t**

$$p(\mathbf{t},\mathbf{w}) = p(\mathbf{w})\prod_{n=1}^{N} p(t_n \mid \mathbf{w}) \quad \cdots(8.6)$$





↑プレート
（N個のtがある）

モデルのパラメータ

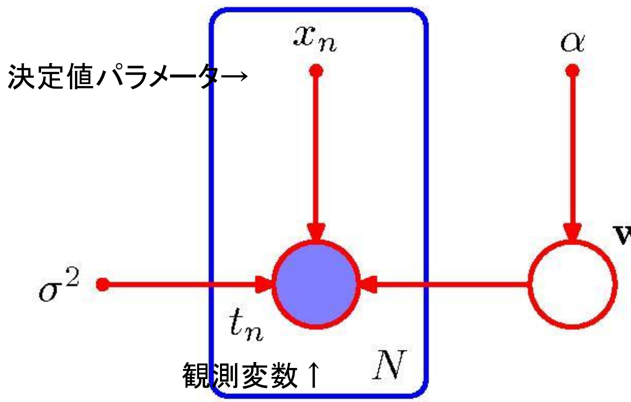
・入力データ **x**

・ノイズの分散 σ^2

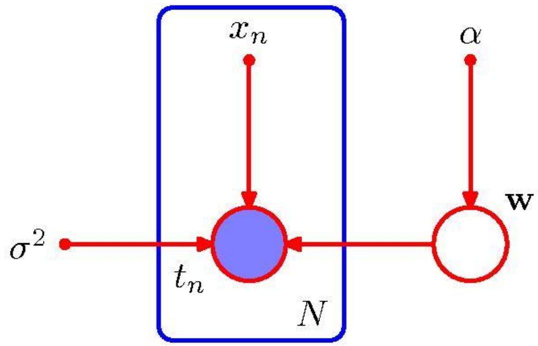
・w上のガウス事前分布の精度を表す超パラメータα

明示的に扱うと

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



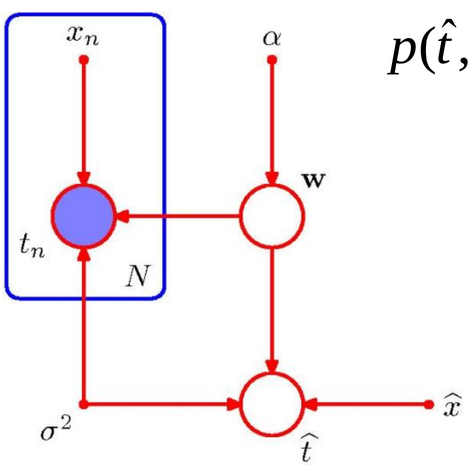
8.1.1 例）多項式フィッティング



wは観測されていない → 潜在変数
{t_n}の値を観測すると係数**w**の事後分布を求められる
(1.2.5より)

$$p(\mathbf{w}|\mathbf{t}) \propto p(\mathbf{w})\prod_{n=1}^{N} p(t_n|\mathbf{w}) \quad \cdots(8.7)$$

多項式フィッティングの最終目的:新しい入力値$\hat{x}$に対する$\hat{t}$の確率分布を求める



$$p(\hat{t},\mathbf{t},\mathbf{w}|\hat{x},\mathbf{x},\alpha,\sigma^2) = \left[\prod_{n=1}^{N} p(t_n|x_n,\mathbf{w},\sigma^2)\right] p(\mathbf{w}|\alpha)\,p(\hat{t}|\hat{x},\mathbf{w},\sigma^2) \quad \cdots(8.8)$$

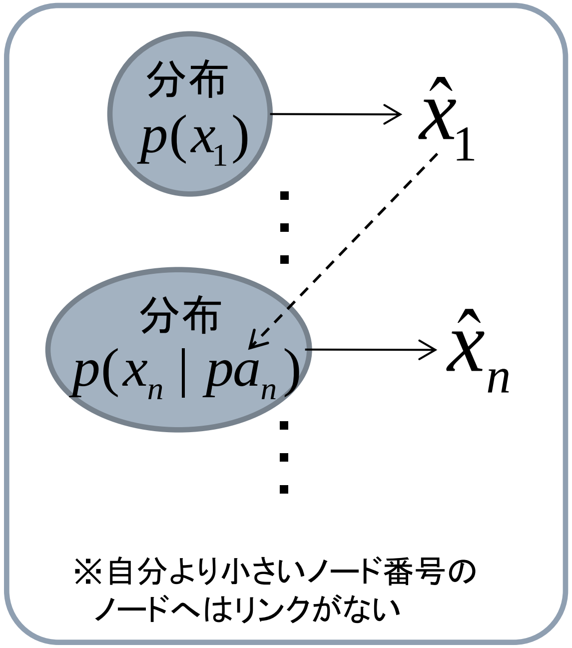
wを積分消去すると

$$p(\hat{t}|\hat{x},\mathbf{x},\mathbf{t},\alpha,\sigma^2) \propto \int p(\hat{t},\mathbf{t},\mathbf{w}|\hat{x},\mathbf{x},\alpha,\sigma^2)dw$$

$\hat{t}$の予測分布

8.1.2 生成モデル

サンプリング法（11章）→伝承サンプリング



観測データが生成される因果過程を表現
→生成モデル

※多項式回帰モデル：入力変数xは確率分布ではない
→生成モデルではない

生成モデル→
観測データと同じ確率分布に従う「架空」データを発生できる

8.1.3 離散変数

グラフィカルモデル：構成要素の接続を表現

→有向グラフの親子対が共役関係になる分布であると，とくによい性質

→特に，離散変数，ガウス変数の場合は有効非循環グラフへ拡張可能

K個の状態をとりうる離散変数xの確率分布

$$p(x|\boldsymbol{\mu}) = \prod_{k=1}^{K} \mu_k^{x_k} \quad \cdots (8.9)$$

パラメータ　$\mu = (\mu_1, \cdots, \mu_K)^T$ により支配

規格化制約 $\sum_k \mu_k = 1$ により，パラメータはK−1個指定すればよい

2つのK状態離散変数x_1及びx_2がある場合

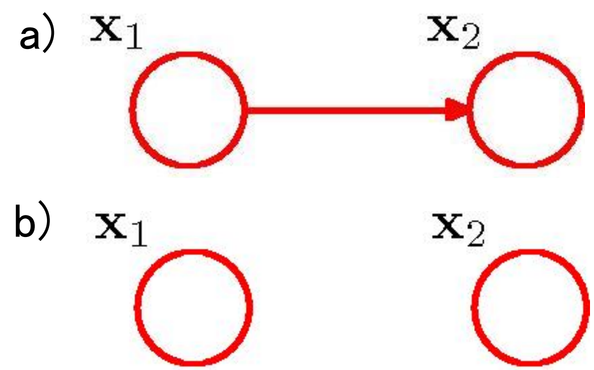
$$p(x_1, x_2 | \boldsymbol{\mu}) = \prod_{k=1}^{K} \prod_{l=1}^{K} \mu_{kl}^{x_{1k} x_{2k}}$$

規格化制約 $\sum_k \sum_l \mu_{kl} = 1$　K^2-1個のパラメータ

変数M個の時：K^M-1個のパラメータ→指数的に増大

8.1.3 離散変数

$$p(x_1, x_2 \mid \boldsymbol{\mu}) = \prod_{k=1}^{K}\prod_{l=1}^{K}\mu_{kl}^{x_{1k}x_{2k}}$$



a）乗法の定理　同時分布$p(x_1, x_2) = p(x_2 \mid x_1)p(x_1)$

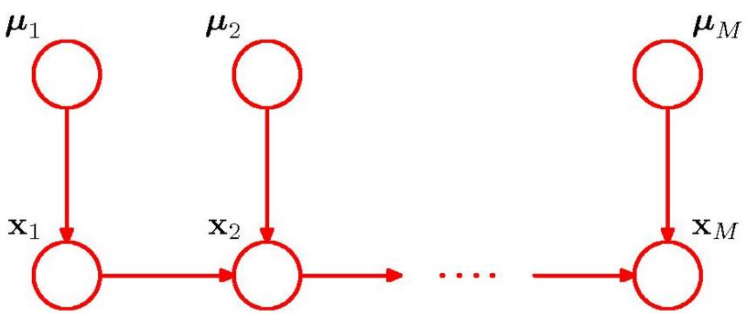
全パラメータ数は　K^2-1

b）変数x_1とx_2が独立　→各変数は別々の多項分布

全パラメータ数は　2(K−1)　‥線形に増加

→リンク除去によりパラメータ数減

リンクの数によって，パラメータ数の増え方が変わる



周辺分布	$p(x_1)$	K−1個
条件つき分布	$p(x_i \mid x_{i-1})$	K(K−1)個　×(M−1)
		線形増加

パラメータにディリクレ事前分布を導入

8.1.4 線形ガウスモデル

要素変数上の線形ガウスモデルに対応する有向グラフにより，
多変量ガウス分布を表現する方法

ノードi ーガウス分布に従う連続値確率変数x_i

$$p(x_i \mid pa_i) = \mathcal{N}\left(x_i \,\middle|\, \sum_{j \in pa_i} \omega_{ij} x_i + b_i , v_i \right) \quad \cdots (8.11)$$

$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k \mid pa_k) \quad \cdots (8.5)$ より，同時分布の対数は，

$$\ln p(x) = \sum_{i=1}^{D} \ln p(x_i \mid pa_i)$$

$$= -\sum_{i=1}^{D} \frac{1}{2v_i} \left(x_i - \sum_{j \in pa_i} \omega_{ij} x_j - b_i \right)^2 + const \quad \cdots (8.13)$$

$$\left(\begin{aligned} &\mathcal{N}\left(x_i \,\middle|\, \sum_{j \in pa_i} \omega_{ij} x_i + b_i , v_i \right) \\ &= \frac{1}{(2\pi v_i)^{1/2}} \exp\left\{ -\frac{1}{2v_i} (x - (\sum_{j \in pa_i} \omega_{ij} x_i + b))^2 \right\} \end{aligned} \right)$$

これはxの成分に関する二次関数
→同時分布p(x)は多変量ガウス分布

8.1.4 線形ガウスモデル

●同時分布の平均

(8.11)に従うので $$x_i = \sum_{j \in pa_i} \omega_{ij} x_j + b_i + \sqrt{v_i}\,\varepsilon_i \quad \cdots(8.14)$$

※ε は平均0，分散1のガウス確率変数

期待値は

$$E[x_i] = \sum_{j \in pa_i} \omega_{ij} E[x_j] + b_i \quad \cdots(8.15)$$

有向非循環グラフなので，E(**x**)の全成分を再帰的に求められる

●同時分布の共分散行列

$$\mathrm{cov}[x_i, x_j] = E[(x_i - E[x_i])(x_j - E[x_j])] = E\left[(x_i - E[x_i]) \left\{ \sum_{k \in pa_j} \omega_{jk} (x_k - E[x_k]) + \sqrt{v_j}\,\varepsilon_j \right\} \right]$$

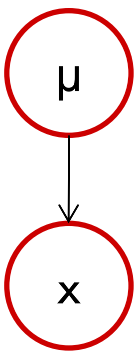
$$= E\left[\sum_{k \in pa_j} \omega_{jk} (x_i - E[x_i])(x_k - E[x_k]) + \left\{ \sum_{k \in pa_i} \omega_{ik} (x_k - E[x_k]) + \sqrt{v_i}\,\varepsilon_i \right\} \cdot \sqrt{v_j}\,\varepsilon_j \right] = \sum_{k \in pa_j} \omega_{jk}\, \mathrm{cov}[x_i, x_k] + I_{ij} v_j \quad \cdot\cdot(8.16)$$

共分散についても，再帰的に値を求められる

8.1.4 線形ガウスモデル

（2章より）

ガウス変数xの平均μに関する共役事前分布がガウス分布である場合，
xおよびμ上の同時分布はガウス分布になる



μ上の分布の平均は事前分布を制御するパラメータなので，
超パラメータとみなされる．

超パラメータの値自体が未知なので，
超パラメータにも事前分布を導入する．
（超事前分布）

これもガウス分布とすれば，ベイズ的取り扱いが可能
→階層ベイズモデルの一例

8.2 条件付き独立性

条件付き独立性

3変数a,b,cを考えたとき，aの条件付き分布がbの値に依存しない．つまり

$$p(a \mid b, c) = p(a \mid c) \quad \cdots (8.20)$$

このとき，cが与えられた下で，aはbに対して条件付き独立であるという

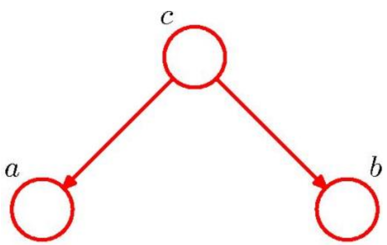
cが与えられたとき（cのとりうる全ての可能な値に対して），
(8.20)が成り立つとき，次のように示す．

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

条件付き独立：モデル構造の簡略化には重要

→グラフィカルモデルにより
同時分布の条件付き独立を直接グラフから読み取れる
→有向分離

8.2.1 3つのグラフの例①

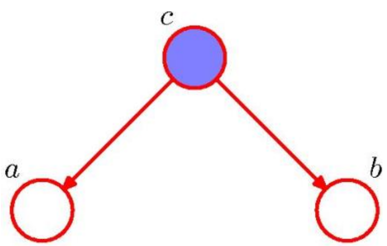


$$p(a,b,c) = p(a|c)p(b|c)p(c)$$

a,bが独立かを調べる（cの周辺化）

$$p(a,b) = \sum_c p(a|c)p(b|c)p(c) \quad \cdots(8.24)$$

p(a)p(b)の形に分解できない→ $a \not\perp\!\!\!\perp b \mid \emptyset$



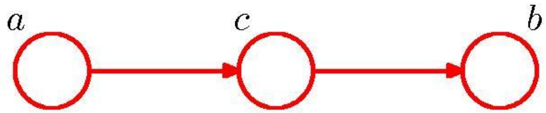
変数cで条件づけ

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

→ $a \perp\!\!\!\perp b \mid c$

cは経路に対して，tail–to–tailとなっている
二つのtailでつながれており，経路が存在し，非独立
cにより，a,b経路が遮断され（条件付き）独立となる

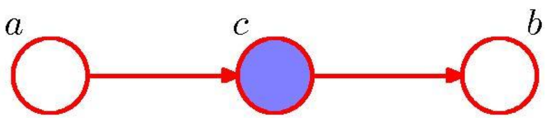
8.2.1 3つのグラフの例②



$$p(a,b,c) = p(a)p(c|a)p(b|c) \quad \cdots (8.26)$$

cについて周辺化 $$p(a,b) = p(a)\sum_c p(c|a)p(b|c) = p(a)p(b|a)$$

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



変数cで条件づけ

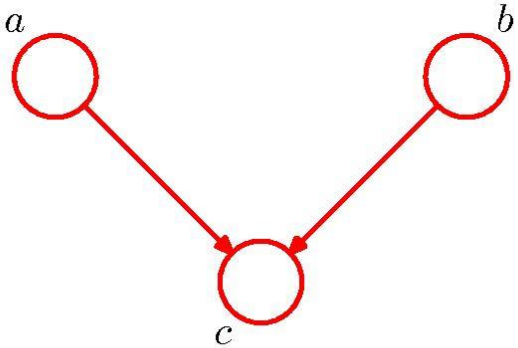
ベイズの定理と(8.26)式より

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

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

cは経路に対して，head-to-tailとなっている
cが観測されないときは，経路よりa,bは従属関係となる
cが観測されることでa→b経路を遮断し，条件付き独立

8.2.1 3つのグラフの例③

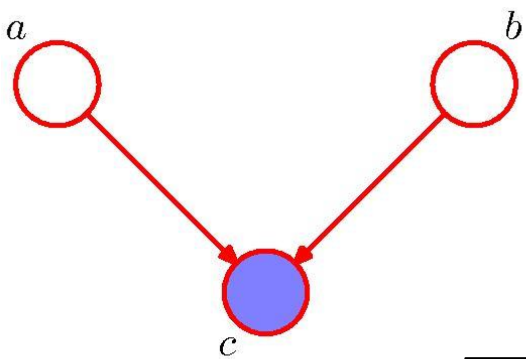


$$p(a,b,c) = p(a)p(b)p(c|a,b) \quad \cdots(8.28)$$

cについて周辺化

$$p(a,b) = p(a)p(b) \qquad \rightarrow a \perp\!\!\!\perp b \mid \emptyset$$

どの変数も観測されてないとき，a,bは独立である



変数cで条件づけ

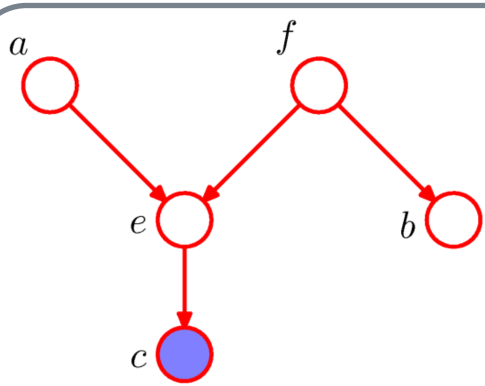
$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = \frac{p(a)p(b)p(c|a,b)}{p(c)}$$

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

cは経路に対して，head-to-headとなっている
cが観測されない時，a,bの関係は遮断されている
cが観測されると，遮断が解かれ，依存関係に
（cの子孫が観測されても遮断は解かれる）．

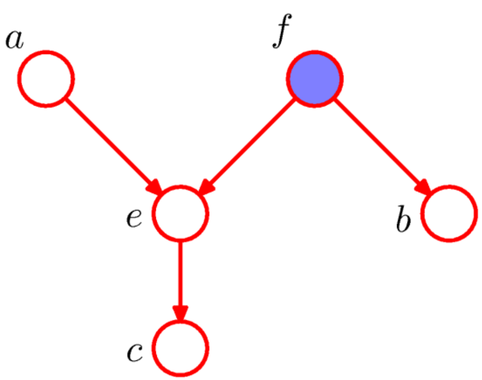
8.2.2 有向分離（D分離）

$a \perp\!\!\!\perp b \mid c$ について考える.



・aからbへの経路はfによって遮断されない
　→f：tail−to−tailで，観測されない
・ aからbへの経路はeによって遮断されない
　→e：head−to−headで，子孫cが条件づけ（観測）

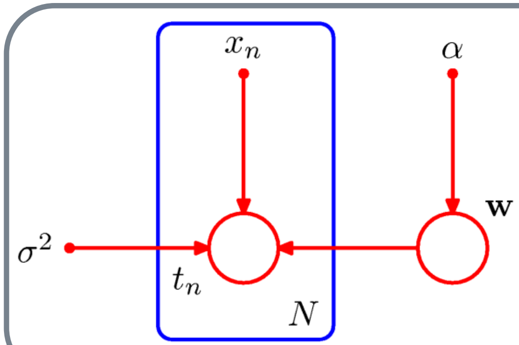
$a \perp\!\!\!\perp b \mid c$ は導けない



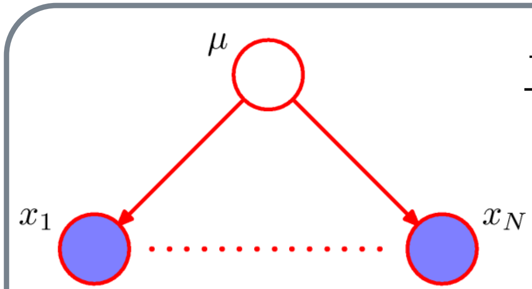
・aからbへの経路はfによって遮断される
　→f：tail−to−tailで，観測されている
・ aからbへの経路はeによって遮断される
　→e：head−to−headで，子孫も条件づけなし

$a \perp\!\!\!\perp b \mid c$ といえる

8.2.2 有向分離（D分離）



パラメータノード：観測済みノード
親ノードなし
すべての経路はtail-to-tail
→他ノードの有向分離性に影響なし



1変量ガウス分布の平均の事後分布について

$D=\{x_1,\cdots,x_N\}$ が観測されたもとでの μ の推論

μ を条件付け変数とみなし，観測変数の同時分布を考える．

$$p(D\mid\mu)=\prod_{n=1}^{N}p(x_n\mid\mu)\quad\cdots(8.34)$$

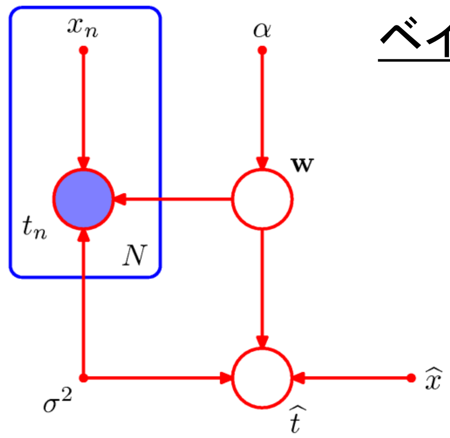
※観測値Dは互いに独立

μ を消去した場合は，観測値は一般に独立ではない．

$$p(D)=\int_{-\infty}^{\infty}p(D\mid\mu)p(\mu)d\mu\neq\prod_{n=1}^{N}p(x_n)\quad\cdots(8.35)$$

※ μ は観測されず，
潜在変数

8.2.2 有向分離（D分離）



ベイズ多項式回帰モデルについて

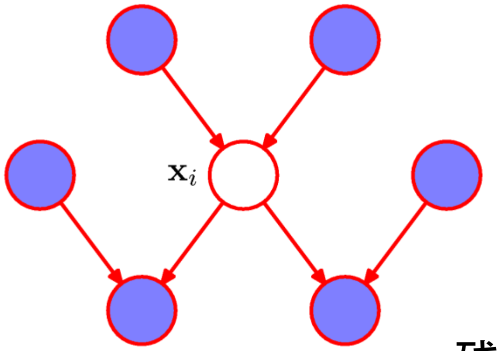
多項式変数wが条件付けられれば，wはtail-to-tailなので

$$\widehat{t} \perp\!\!\!\perp t_n \mid \mathbf{w}$$

$\hat{t}$の予測分布は訓練データt_nに対して独立

→訓練データよりwの事後分布を決定すれば，t_nはいらない

マルコフブランケット（マルコフ境界）



$$p(x_i \mid x_{\{j \neq i\}}) = \frac{p(x_1, \cdots, x_D)}{\int p(x_1, \cdots, x_D) dx_i} = \frac{\prod_k p(x_k \mid pa_k)}{\int \prod_k p(x_k \mid pa_k) dx_i}$$

x_iに関係ない項は分母分子でキャンセル

残る項は
- ・x_iの条件付き分布$p(x_i|pa_i)$ ―親
- ・x_iを条件付け変数集合に含む任意のx_kの条件付き分布$p(x_k|pa_k)$ ―子・共同親

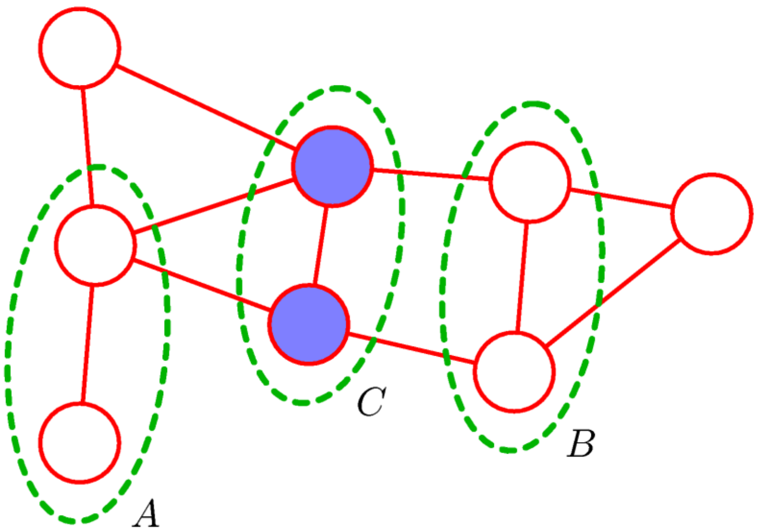
8.3 マルコフ確率場

有向グラフィカルモデル(**ベイジアンネットワーク**)
→確率変数間の因果関係，リンクが特定の方向性
同時分布を局所的な条件つき分布の積に因数分解
因数分解される分布の条件つき独立性の集合

無向グラフィカルモデル(**マルコフ確率場**)
→確率変数間の緩い束縛関係，リンクは方向性なし
ノード集合とリンク集合

8.3.1 条件付き独立性

有向グラフ：有向分離，head−to−headとtail−to−tailの混在
無向グラフ：親ノードと子ノードの非対称性なし



$A \perp\!\!\!\perp B \mid C$ を判断するには‥

集合Aと集合Bを結ぶ全ての経路
→集合Cのノードを少なくとも一つ含む
→全ての経路が遮断され，条件付き独立
（集合Cを除いた時，A−B経路の存在有無）

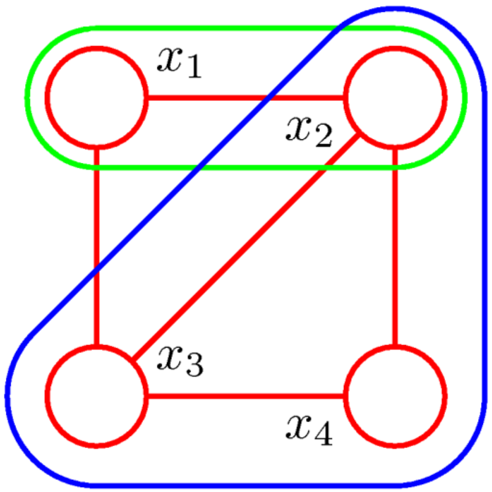
無向グラフのマルコフブランケット
→隣接ノード集合

8.3.2 分解特性

直接接続されない2つのノードx_i, x_jは条件付き独立

$$p(x_i, x_j \mid \mathbf{x}_{\setminus\{ij\}}) = p(x_i \mid \mathbf{x}_{\setminus\{ij\}})\, p(x_j \mid \mathbf{x}_{\setminus\{ij\}}) \quad \cdots (8.38)$$

x_i, x_jが**因子に含まれない**ように，因数分解される



クリーク：
全てのノードの組にリンクが存在するグラフの部分集合

$$\{x_1, x_2\}, \{x_2, x_3\}, \{x_3, x_4\}, \{x_4, x_2\}, \{x_1, x_3\},$$

$$\{x_1, x_2, x_3\}, \{x_2, x_3, x_4\}$$ ←極大クリーク

同時分布を因数分解したときの各因子を，
クリークが含む変数の集合の関数すれば良い

8.3.2 分解特性

クリークC，クリーク内の変数の集合x_cとする.
同時分布は，極大クリーク上のポテンシャル関数$\psi_C(x_C)$の積の形でかける.

$$p(x)=\frac{1}{Z}\prod_C \psi_C(\mathbf{x}_C) \quad \cdots(8.39)$$

Zは規格化定数(分配関数)であり，

$$Z=\sum_{\mathbf{x}}\prod_C \psi_C(\mathbf{x}_C) \quad \cdots(8.40)$$

ポテンシャル関数は，
周辺分布や条件付き分布のように確率的解釈が可能なものに**限定されない**
※規格化定数が必要. 積と和の計算により，モデルサイズに応じて計算量増

<u>定式化</u>

ポテンシャル関数$\psi_C(x_C)$が狭義に正(因数分解と条件付き独立の関係から)
指数関数で表現

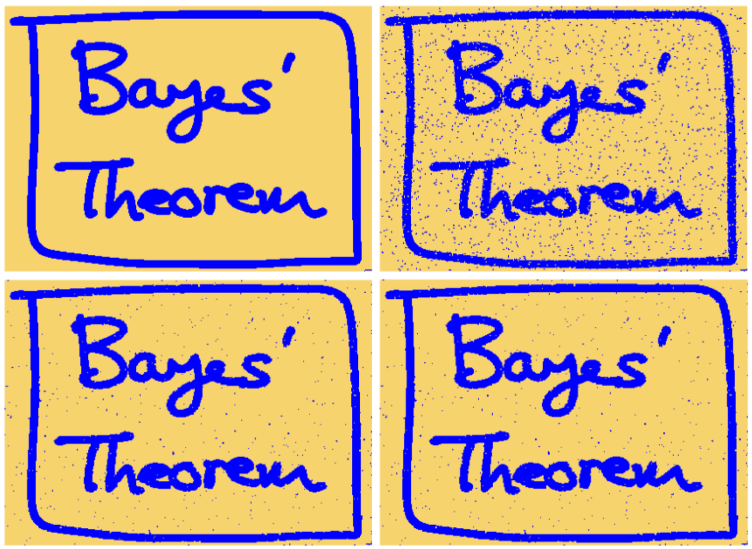
$$\psi_C(\mathbf{x}_C)=\exp\{-E(\mathbf{x}_C)\} \quad \cdots(8.41)$$

($E(x_C)$はエネルギー関数，この指数表現はボルツマン分布)

ポテンシャル関数は確率的解釈はないので，ポテンシャル関数は自由に選べる
→選び方は・・・局所的な変数がどのような形状を持てばいいのか

8.3.3 例：画像のノイズ除去

ノイズなし　　値y_i(10%反転)



ICM法（96%一致）　　グラフカットアルゴリズム（99%一致）

観測画像2値ピクセル値$y_i \in \{-1,1\}$の二次元配列
（ノイズのない2値画像$x_i \in \{-1,1\}$からランダムに反転）

ノイズレベルが低いために
x_iとy_iとの間に強い相関が残っているはず.

$$-\eta x_i y_i$$

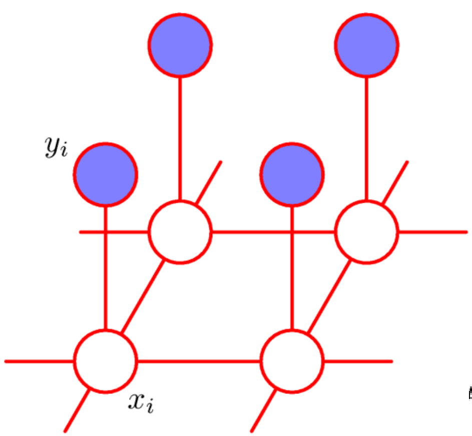
隣接ピクセルx_iとx_jとの間に強い相関があるはず.

$$-\beta x_i x_j$$

同符号の時，低いエネルギー（高い確率）
異符号の時，高いエネルギー（低い確率）

$$E(\mathbf{x},\mathbf{y}) = h\sum_i x_i - \beta \sum_{\{i,j\}} x_i x_j - \eta \sum_i x_i y_i \quad \cdots(8.42)$$

$$p(\mathbf{x},\mathbf{y}) = \frac{1}{Z}\exp\{-E(\mathbf{x},\mathbf{y})\} \quad \cdots(8.43)$$



h x_iは特定の符号を持ちやすくするためのバイアス効果

8.3.3 例：画像のノイズ除去

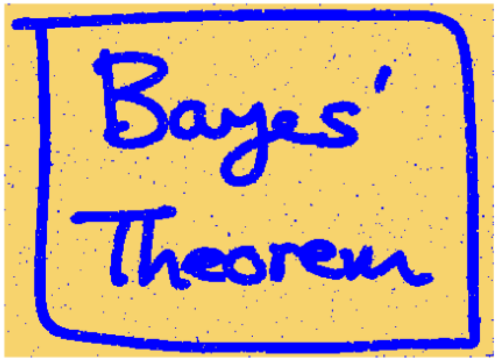
画像復元のために高い確率を持つ画像**x**を求めたい(イジングモデル)

●ICM法（反復条件付きモード）

1. 変数{xi}を初期化（$x_i=y_i$など）
2. あるノードx_jを選ぶ
3. $x_j=+1$と$x_j=-1$における全エネルギーを計算
4. エネルギーが小さくなる方にx_jを設定
5. 2に戻り，違う場所で計算
6. ある規準になるまで繰り返し

→全ての場所を少なくとも1回は通るシークエンスで，値が更新されず

→極大点の発見



8.3.4 有向グラフとの関係

有向グラフ

$$p(\mathbf{x}) = p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_2)\cdots p(x_N \mid x_{N-1}) \quad \cdots(8.44)$$

無向グラフ

$$p(\mathbf{x}) = \frac{1}{Z}\psi_{1,2}(x_1, x_2)\psi_{2,3}(x_2, x_3)\cdots\psi_{N-1,N}(x_{N-1}, x_N) \quad \cdots(8.45)$$

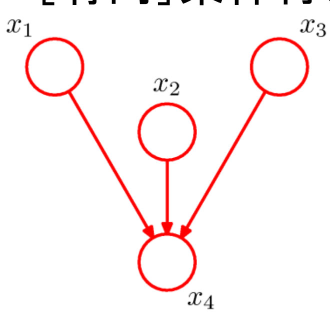
→対応付け

$$\begin{cases} \psi_{1,2}(x_1, x_2) = p(x_1)p(x_2 \mid x_1) \\ \psi_{2,3}(x_2, x_3) = p(x_3 \mid x_2) \\ \vdots \\ \psi_{N-1,N}(x_{N-1}, x_N) = p(x_N \mid x_{N-1}) \end{cases}$$

無向グラフ		有向グラフ
クリークポテンシャル関数	⇔	条件付き分布

変換の正確さ

[有向]条件付き分布の変数集合全て→ [無向]1つのクリーク集合に含まれる

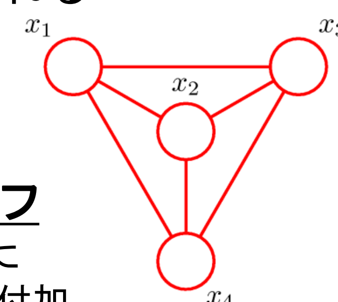


$$p(\mathbf{x}) = p(x_1)p(x_2)p(x_3)\underline{p(x_4 \mid x_1, x_2, x_3)}$$

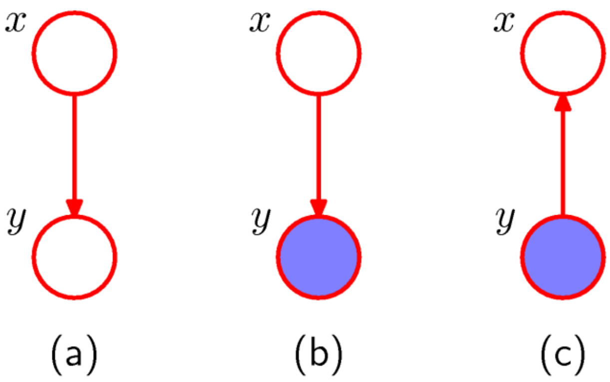
全ての変数が1つのクリーク
に属さなければならない

→ **モラルグラフ**

親同士の対に
無向リンクを付加



8.4 グラフィカルモデルにおける推論



変数yの値が観測される（図(b)）→p(y|x)

潜在変数xの周辺分布p(x)は事前分布

確率の加法定理，乗法定理より

$$p(y) = \sum_{x'} p(y \mid x') p(x') \quad \cdots (8.47)$$

ベイズの定理より

$$p(x \mid y) = \frac{p(y \mid x) p(x)}{p(y)} \quad \cdots (8.48)$$

xの事後分布p(x|y)が推論された．

8.4.1 連鎖における推論

グラフの同時分布は次のようになる

$$p(\mathbf{x}) = \frac{1}{Z}\psi_{1,2}(x_1, x_2)\psi_{2,3}(x_2, x_3)\cdots\psi_{N-1,N}(x_{N-1}, x_N) \quad \cdots(8.49)$$

ノードx_nの周辺分布は
$$p(x_n) = \sum_{x_1}\cdots\sum_{x_{n-1}}\sum_{x_{n+1}}\cdots\sum_{x_N} p(\mathbf{x}) \quad \cdots(8.50)$$

K状態変数ノードがN個分の計算→xのとりうる状態はK^N個

→周辺分布を求めるには指数オーダーの計算量

<u>**グラフィカルモデルの効率利用**</u>

(8.49)を(8.50)に代入　$\sum_{x_N}$ に関係するのは　$\psi_{N-1,N}(x_{N-1}, x_N)$のみ

x_Nについては $\sum_{x_N}\psi_{N-1,N}(x_{N-1}, x_N)$ を計算．x_{N-1}に関係するのは…

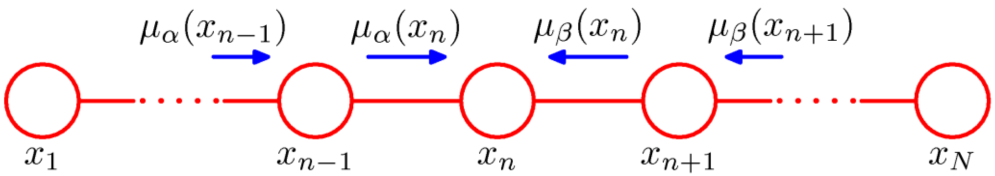
ノードx_nの周辺分布は

$$p(x_n) = \frac{1}{Z}\underbrace{\left[\sum_{x_{n-1}}\psi_{n-1,n}(x_{n-1}, x_n)\cdots\left[\sum_{x_2}\psi_{2,3}(x_2, x_3)\left[\sum_{x_1}\psi_{1,2}(x_1, x_2)\right]\right]\cdots\right]}_{\mu_\alpha(x_n)}\underbrace{\left[\sum_{n+1}\psi_{n,n+1}(x_n, x_{n+1})\cdots\left[\sum_{x_N}\psi_{N-1,N}(x_{N-1}, x_N)\right]\right]}_{\mu_\beta(x_n)} \quad (8.52)$$

8.4.1 連鎖における推論

局所的なメッセージの伝搬

(8.52)より $p(x_n)=\frac{1}{Z}\mu_\alpha(x_n)\mu_\beta(x_n)\ \cdots(8.54)$



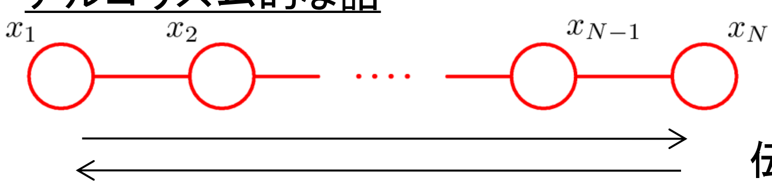
μ_α :ノード番号大へ前向きに伝わるメッセージ
μ_β :ノード番号小へ後向きに伝わるメッセージ

メッセージ μ_α は $\mu_\alpha(x_n)=\sum_{x_{n-1}}\psi_{n-1,n}(x_{n-1},x_n)\left[\sum_{x_{n-2}}\cdots\right]=\sum_{x_{n-1}}\psi_{n-1,n}(x_{n-1},x_n)\mu_\alpha(x_{n-1})\ \cdots(8.55)$

なので， $\mu_\alpha(x_2)=\sum_{x_1}\psi_{1,2}(x_1,x_2)\ \cdots(8.56)$ から再帰的に所望のノードに到達するまで繰り返す

後向きも同様．マルコフ連鎖と呼ばれる．

アルゴリズム的な話



伝播中の全ての（中間的な）メッセージを保存

途中にいくつかのノードが観測されている場合は，観測値固定

まとめ

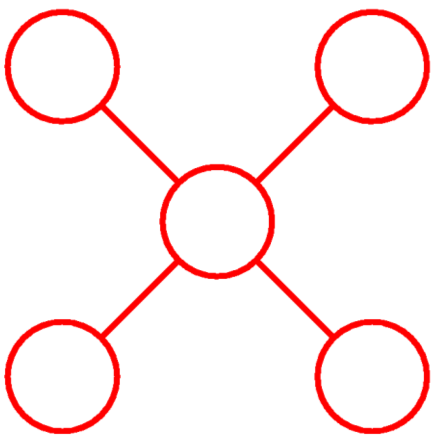
- 有向グラフィカルモデルと無向グラフィカルモデル
- 条件付き確率分布の関係性を明示的に扱う
- グラフの関係性に従い，事前分布・観測変数から他の変数を求めることができる
- 有向グラフィカルモデルでは，有向分離が必要

- 因子グラフを用いての，推論．．

8.4.2 木

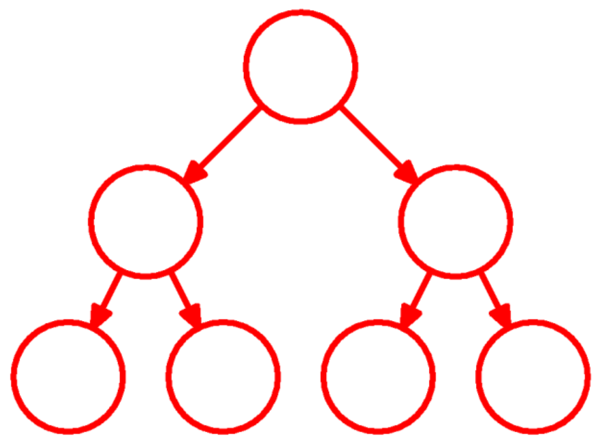
連鎖⇔メッセージパッシング　←厳密推論
木構造グラフ⇔積和アルゴリズム　←一般化，厳密推論

無向木



ループを持たない

有向木

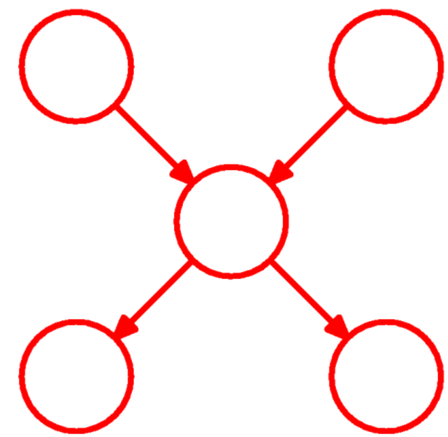


ループを持たない

親なしノード１つ
他ノードは親１つ

無向変換時　モラル化必要なし

有向多重木



ループを持たない

親はいくつでもいい

モラル化必要あり

8.4.3 因子グラフ

●有向グラフ・無向グラフ

多くの変数に依存する大域的な関数が，局所的な変数の部分集合のみに依存する

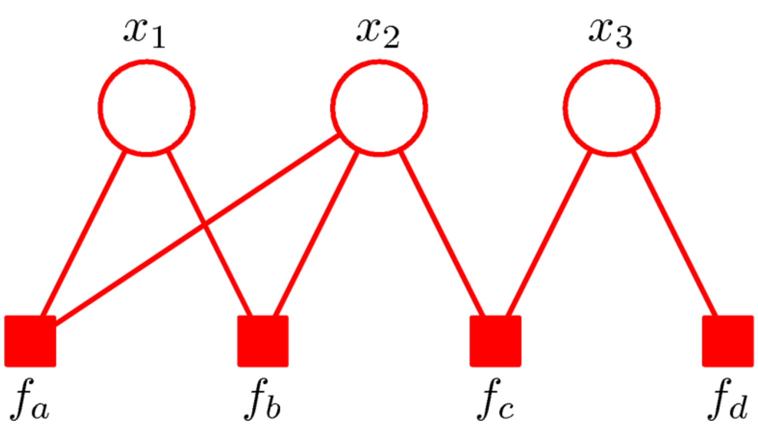
●因子グラフ

変数を表現するノードに因子（関数）そのものに対応するノードを付け加える

ある変数上の同時分布を因子の積の形で表す

$$p(\mathbf{x}) = \prod_s f_s(\mathbf{x}_s) \quad \cdots(8.59)$$

有向グラフ $p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k \mid pa_k) \quad \cdots(8.5)$ (8.59)の特別な場合

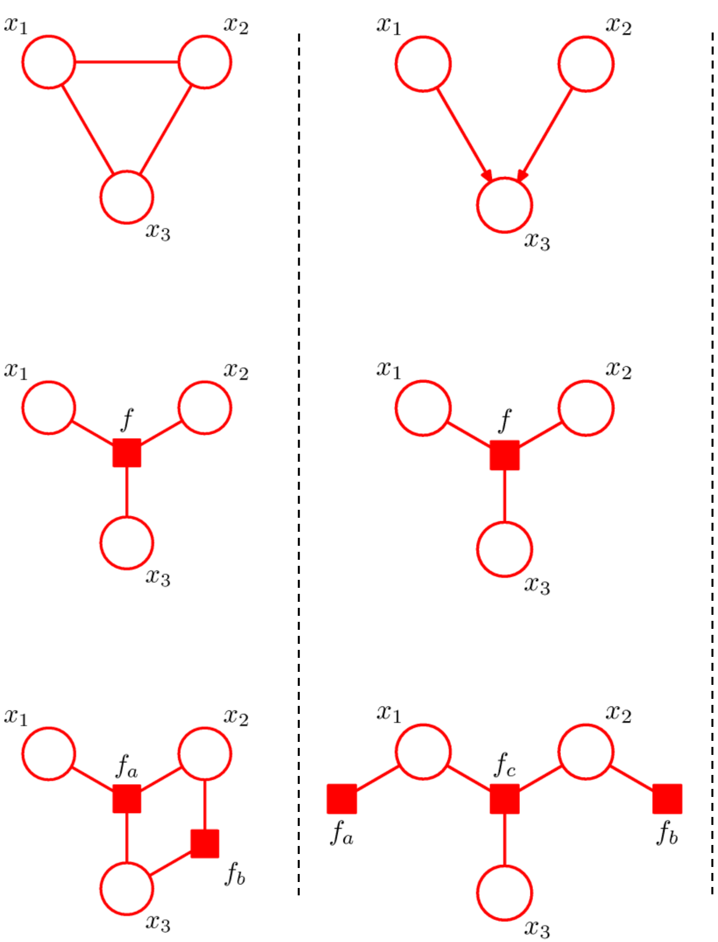


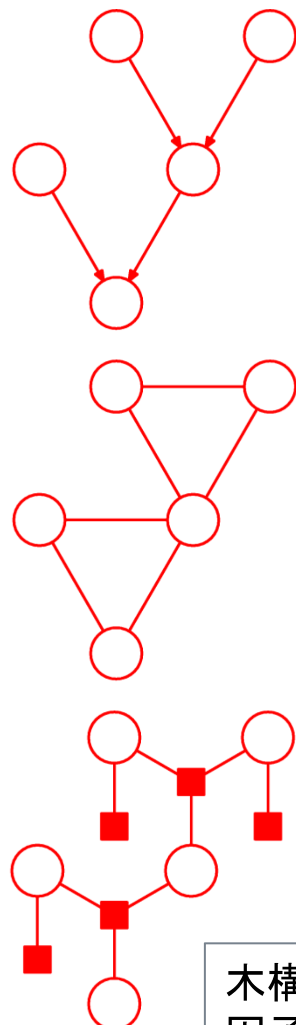
$$p(\mathbf{x}) = f_a(x_1, x_2) f_b(x_1, x_2) f_c(x_2, x_3) f_d(x_3)$$

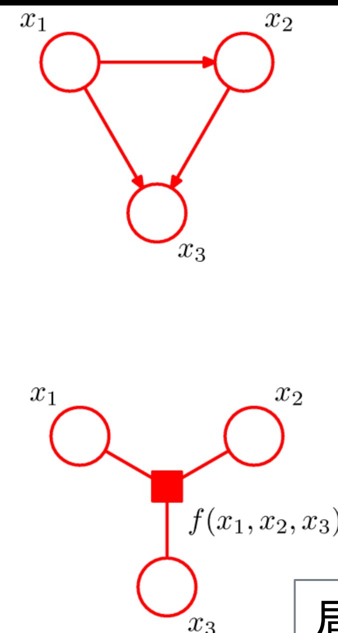
まとめていない

Pattern Recognition and Machine Learning

8.4.3 因子グラフ







局所的な
ループの回避

同じグラフでの違った因子グラフ表現
→より正確に因数分解を表現

木構造を
因子グラフでは保持

8.4.4 積和アルゴリズム

ノード(ノード部分集合)上の局所的な**周辺分布**の計算アルゴリズム

木構造の因子グラフに適応(どのグラフからも変換可能)

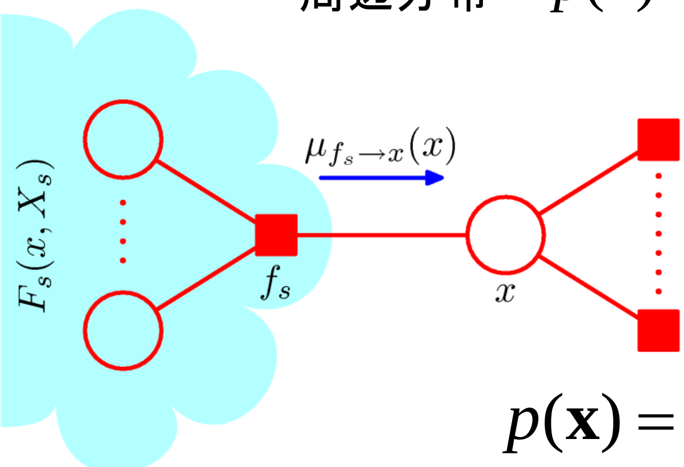
(i) 周辺分布を求めるための
効率の良い厳密推論アルゴリズムを得る

(ii) 複数の周辺分布を計算した場合に,
計算の重複をなくして効率化する

8.4.4 積和アルゴリズム

ある特定の変数x上の周辺分布p(x)を求める問題を考える（変数は離散的）

周辺分布 $$p(x)=\sum_{\mathbf{x}\backslash x} p(\mathbf{x}) \quad \cdots(8.61)$$ x以外の変数の同時分布の和



グラフは木構造なので，
同時分布を
因子の変数ノードxに隣接する各因子ノードごとに
グループ分けできる．

$$p(\mathbf{x})=\prod_{s\in ne(x)} F_s(x, X_s) \quad \cdots(8.62)$$

部分木内の同時分布の積が
全体の同時分布

ne(x)：xに隣接する因子のノード集合
X_s：f_sを通して変数ノードxに接続される部分木の変数集合
$F_s(x,X_s)$：f_sに関連するグループすべての因子の積

$$p(x)=\prod_{s\in ne(x)}\left[\sum_{X_s} F_s(x, X_s)\right]=\prod_{s\in ne(x)} \mu_{f_s\to x}(x) \quad \cdots(8.63)$$

8.4.4 積和アルゴリズム