

鈴木讓「ベイジアンネットワーク入門」(培風館)

### 3. 統計的学習

#### 3.5 有限型 Bayesian ネットワークの学習 (3)

コスト最小極大木アルゴリズムの適用

鈴木讓

大阪大学

2010年7月1日

あらまし

木への近似

## 木への近似

$$P_{X^{(j)}|X^{(0)}}(\cdot|x^{(0)}) = P_{X^{(j)}}(\cdot)$$

### Dendroid 分布

$$P_{X^{(1)}\dots X^{(N)}}(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}), \quad \mathbf{x}^{(1)} \in X^{(1)}(\Omega), \dots, \mathbf{x}^{(N)} \in X^{(N)}(\Omega)$$

$$\hat{P}_{X^{(1)}\dots X^{(N)}}(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}) := \prod_{j \in N} P_{X^{(a_j)}|X^{(a_{\lambda[j]}}}(\mathbf{x}^{(a_j)}|x^{(a_{\lambda[j]})}) \quad (43)$$

$$1 \leq \lambda[j] \leq j-1, \quad j \neq 1, \quad \lambda[1] = 0,$$

$\mathbf{U} := \{X^{(i)} \mid i = 1, \dots, N\}$ : 頂点集合

### 補題 3.3

無向グラフ  $G = (\mathbf{U}, \mathbf{E})$  が木であることと,

$\mathbf{E} = \{\{X^{(a_j)}, X^{(a_{\lambda[j]})}\} \mid j = 1, \dots, N\}$ ,  $1 \leq \lambda[j] \leq j-1$ ,  $j \neq 1$ ,

$\lambda[1] = 0$  なる  $(1, \dots, N)$  の順列  $(a_1, \dots, a_N)$  が存在することは同値

## Chow-Liu アルゴリズム

$$\begin{aligned} D(P_{X^{(1)} \dots X^{(N)}} \parallel \hat{P}_{X^{(1)} \dots X^{(N)}}) \\ &:= \sum_{\mathbf{x}^{(1)} \in X^{(1)}(\Omega), \dots, \mathbf{x}^{(N)} \in X^{(N)}(\Omega)} P_{X^{(1)} \dots X^{(N)}}(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}) \\ &\quad \cdot \log \frac{P_{X^{(1)} \dots X^{(N)}}(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)})}{\prod_{j \in N} P_{X^{(a_j)} | X^{(a_{\lambda[j]}}}(\mathbf{x}^{(a_j)} | \mathbf{x}^{(a_{\lambda[j]}})})} \end{aligned} \quad (44)$$

$$I(i, j) := \sum_{\mathbf{x}^{(i)} \in X^{(i)}(\Omega), \mathbf{x}^{(j)} \in X^{(j)}(\Omega)} P_{X^{(i)} X^{(j)}}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \log \frac{P_{X^{(i)} X^{(j)}}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})}{P_{X^{(i)}}(\mathbf{x}^{(i)}) P_{X^{(j)}}(\mathbf{x}^{(j)})}$$

### 命題 (Chow-Liu)

$D(P_{X^{(1)} \dots X^{(N)}} \parallel \hat{P}_{X^{(1)} \dots X^{(N)}})$  最小の  $(a_1, \dots, a_N), \lambda[k], 2 \leq k \leq N$  は, アルゴリズム 2.1 に  $\mathbf{U} := \{X^{(i)} \mid i = 1, \dots, N\}$ ,  $w(\{X^{(i)}, X^{(j)}\}) := I(i, j)$  を入力することで得られる。

## Chow-Liu アルゴリズム

証明:  $D(P_{X^{(1)} \dots X^{(N)}} \parallel \hat{P}_{X^{(1)} \dots X^{(N)}}) = \sum_{j=1}^N -I(a_j, a_{\lambda[j]}) + (\text{定数})$

## 例から出発して木を生成する

$n$  個の例

$$x^n = (x_1, \dots, x_n), \quad x_i \in X(\Omega) = \prod_{j=1}^N X^{(j)}(\Omega), \quad i = 1, \dots, n$$

$$I_n(i, j) := \sum_{x \in X^{(i)}, y \in X^{(j)}} c_{i,j}[x, y] \log \frac{c_{i,j}[x, y]}{c_i[x]c_j[y]} \geq 0$$

$c_i[x]$ :  $x_1, \dots, x_n$  における  $X^{(i)} = x \in X^{(i)}(\Omega)$  の頻度

$c_{i,j}[x, y]$ :  $x_1, \dots, x_n$  における

$X^{(i)} = x \in X^{(i)}(\Omega)$ ,  $X^{(j)} = y \in X^{(j)}(\Omega)$  の同時頻度

$$\sum_{x^{(1)}, \dots, x^{(N)}} c_{1, \dots, N}(x^{(1)}, \dots, x^{(N)}) \cdot \log \prod_{j=1}^N \frac{c_{a_{\lambda[j]}}(x^{(a_{\lambda[j]})})}{c_{a_j, a_{\lambda[j]}}(x^{(a_j)}, x^{(a_{\lambda[j]})})}$$

( $\mathcal{H}[\pi](x^n)$  とおく) の最小化をはかる

$c_{1, \dots, N}(x^{(1)}, \dots, x^{(N)})$ :  $X^{(1)} = x^{(1)} \in X^{(1)}(\Omega)$ ,  $\dots$ ,

$X^{(N)} = x^{(N)} \in X^{(N)}(\Omega)$  が同時に成立する頻度

## 例から出発して木を生成する

### アルゴリズム 3.1: Chow-Liu アルゴリズム

入力:  $n$  個の例  $x^n$

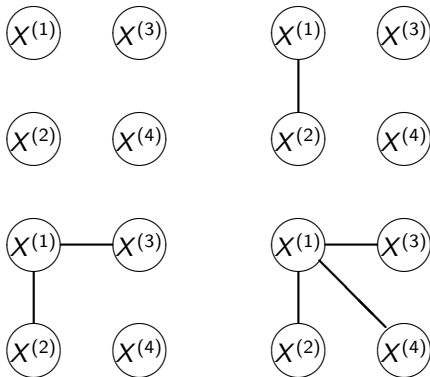
出力:  $\mathcal{H}[\pi](x^n)$  を最小とする木  $T$

1.  $\mathbf{E} \leftarrow \{\}$
2.  $\mathcal{E} \neq \{\}$  である限り, 以下を繰り返す。
  - (1)  $I_n(i, j)$  が最大の  $\{X^{(i)}, X^{(j)}\} \in \mathcal{E}$  について,  
 $\mathcal{E} \leftarrow \mathcal{E} - \{\{X^{(i)}, X^{(j)}\}\}$
  - (2)  $(\mathbf{U}, \mathbf{E} \cup \{\{X^{(i)}, X^{(j)}\}\})$  が巡回経路をもたないとき,  
 $\mathbf{E} \leftarrow \mathbf{E} \cup \{\{X^{(i)}, X^{(j)}\}\}$
3.  $T \leftarrow (\mathbf{U}, \mathbf{E})$

$\mathcal{H}[\pi](x^n)$ (3.45) の値は, 例  $x^n$  の木への適合性とみなせる

# 実行例 (1)

$i$	$j$	$l_n(i,j)$
1	2	12
1	3	10
2	3	8
1	4	6
2	4	4
3	4	2





## 例から出発して森を生成する

$\mathcal{H}[\pi](x^n)$  の最小化ではなく, パラメータ数  $K[\pi]$  を加えた

$$L[\pi](x^n) := \mathcal{H}[\pi](x^n) + \frac{K[\pi]}{2} \log n$$

の最小化

$$J_n(i, j) := I_n(i, j) - \frac{(\alpha^{(i)} - 1)(\alpha^{(j)} - 1)}{2} \log n$$

### 補題

無向グラフ  $G = (\mathbf{U}, \mathbf{E})$  が森であることと,

$$\mathbf{E} = \{ \{X^{(a_j)}, X^{(a_{\lambda[j]})}\} \mid j = 1, \dots, N \},$$

$0 \leq \lambda[j] \leq j - 1, j = 1, \dots, N, \lambda[1] = 0$  なる  $(1, \dots, N)$  の順列  $(a_1, \dots, a_N)$  が存在することは同値。

木では  $1 \leq \lambda[j] \leq j - 1$ , 森では  $0 \leq \lambda[j] \leq j - 1$

## 例から出発して森を生成する (続)

### アルゴリズム 3.2: Suzuki, 1993

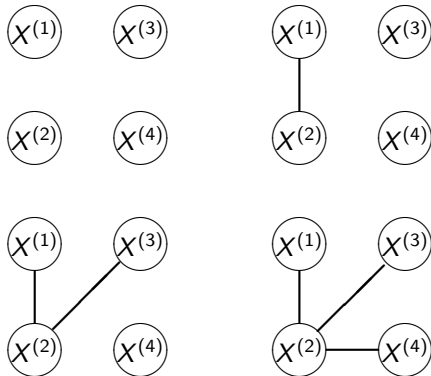
入力:  $n$  個の例  $x^n$

出力:  $L(x^n)[\pi]$  を最小とする森  $F$

1.  $\mathbf{E} \leftarrow \{\}$
2.  $\mathcal{E} \neq \{\}$  である限り, 以下を繰り返す.
  - (1)  $J_n(i, j)$  が最大の  $\{X^{(i)}, X^{(j)}\} \in \mathcal{E}$  について,  
 $\mathcal{E} \leftarrow \mathcal{E} - \{\{X^{(i)}, X^{(j)}\}\}$
  - (2)  $(\mathbf{U}, \mathbf{E} \cup \{\{X^{(i)}, X^{(j)}\}\})$  が巡回経路をもたず, かつ  
 $J_n(i, j) \geq 0$  のとき,  $\mathbf{E} \leftarrow \mathbf{E} \cup \{\{X^{(i)}, X^{(j)}\}\}$
3.  $F \leftarrow (\mathbf{U}, \mathbf{E})$

## 実行例 (2)

$i$	$j$	$l_n(i, j)$	$\alpha^{(i)}$	$\alpha^{(j)}$	$J_n(i, j)$
1	2	12	5	2	8
1	3	10	5	3	2
2	3	8	2	3	6
1	4	6	5	4	-6
2	4	4	2	4	1
3	4	2	3	4	-4



## 例から出発して森を生成する (続)

### 定理 3.10, Suzuki, 1993

アルゴリズム 10 は,  $n$  個の例  $x^n$  から  $L[\pi](x^n)$  を最小とする森  $T$  を正しく出力する。

証明 :

$$K[\pi] = \sum_{j=1}^N (\alpha^{(a_j)} - 1) \alpha^{(a_{\lambda[j]})}$$

$L[\pi](x^n)$  の最小化は, 以下の最小化と同値

$$\begin{aligned} & \sum_{j=1}^N -I_n(a_j, a_{\lambda[j]}) + \frac{1}{2} \sum_{j=1}^N (\alpha^{(a_j)} - 1)(\alpha^{(a_{\lambda[j]})} - 1) \log n \\ &= - \sum_{j=1}^N J_n(a_j, a_{\lambda[j]}) \end{aligned}$$