

# 10. ベイジアンネットワーク分類 器

電気通信大学  
情報理工学研究科  
植野真臣

# スケジュール

4月8日 授業の概要とガイダンス

4月15日 ベイズの定理

4月22日 ベイズはどのように誕生したか？

5月13日 ベイズはコンピュータ、人工知能の父である！！

5月20日 ビリーフとベイズ

5月27日 6月3日 6月10日 尤度と最尤推定

6月17日 6月24日 7月1日 ベイズ推定と事前分布、階層ベイズ、因果推論

7月8日 自宅でオンデマンド授業

7月15日 自宅でオンデマンド授業

7月22日 7月29日 確率的グラフィカルモデル、ベイジアンネットワークと機械学習

8月5日 テストと総括

# 本日の目標

1. ベイジアンネットワーク分類器
2. Augmented Naive Bayes Classifiers (ANB)
3. 大規模ANBの学習
4. 分類影響パラメータ数最小化による  
ベイジアンネットワーク分類器学習

# 1. ベイジアンネットワーク分類器

## 1.1 ベイジアンネットワーク分類器 (Friedman et al., 1988)

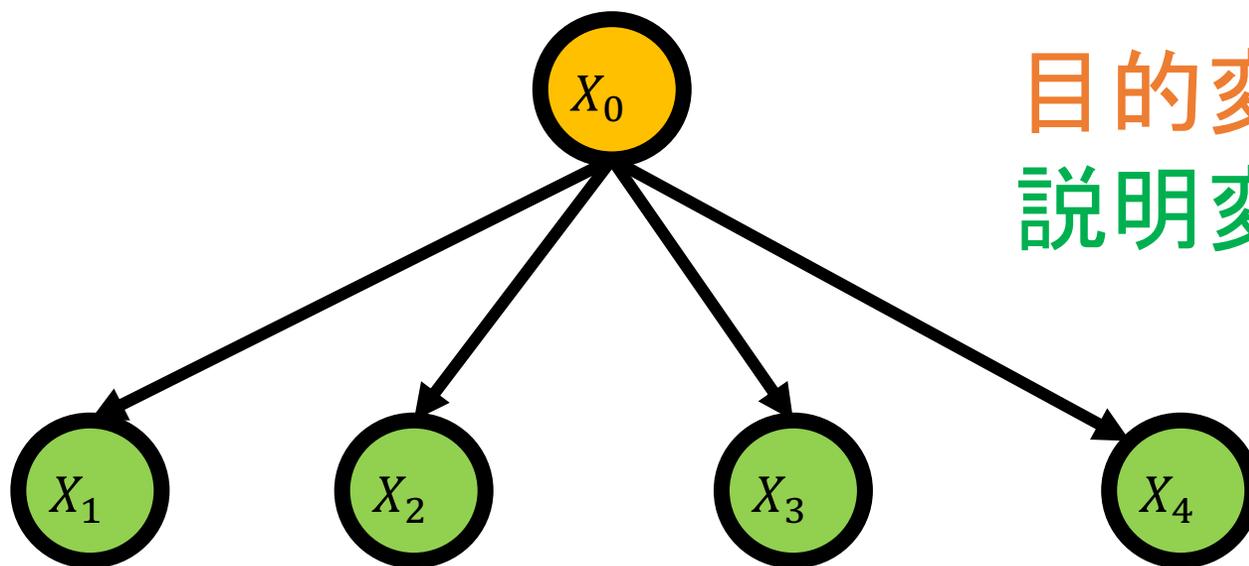
離散確率変数集合  $V = \{X_0, X_1, \dots, X_n\}$  をもつベイジアンネットワークについて,  $X_0$  を目的変数,  $\mathbf{F} = \{X_1, \dots, X_n\}$  を説明変数集合とする. いま  $\mathbf{F}$  の値  $\mathbf{x} = \{x_1, \dots, x_n\}$  を得たとき, 以下のように  $X_0$  の推定値  $\hat{c}$  を得る.

$$\begin{aligned}\hat{c} &= \operatorname{argmax}_{c=\{1, \dots, r_0\}} P(X_0 = c \mid \mathbf{F} = \mathbf{x}) \\ &= \operatorname{argmax}_{c=\{1, \dots, r_0\}} \frac{P(X_0 = c, \mathbf{F} = \mathbf{x})}{P(\mathbf{F} = \mathbf{x})} \\ &= \operatorname{argmax}_{c=\{1, \dots, r_0\}} P(X_0 = c, \mathbf{F} = \mathbf{x}).\end{aligned}$$

## 1.2 Friedman et al. (1997) による批判

周辺尤度で学習したベイジアンネットワークの分類精度が、単純な構造をとるNaive Bayes (Minsky, 1961) より劣ることが多々ある.

Naïve Bayesの例



目的変数 :  $X_0$

説明変数 :  $X_1, \dots, X_4$

N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian Network Classifiers,” Machine Learning, vol.29, no.2, pp.131–163, 1997.

Marvin Minsky. Steps toward Artificial Intelligence. In Proceedings of the IRE, volume 49, pp. 8–30, 1961.

## 1.2 Friedman et al. (1997) による批判

周辺尤度, MDL : 同時確率分布  $P(X_0, X_1, \dots, X_n | G)$  を表現する

生成モデルを学習

N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian Network Classifiers,”  
Machine Learning, vol.29, no.2, pp.131–163, 1997.

## 1.2 Friedman et al. (1997) による批判

~~周辺尤度, MDL: 同時確率分布  $P(X_0, X_1, \dots, X_n | G)$  を表現する~~

分類が**生成モデル**を学習する確率  $P(X_0 | X_1, \dots, X_n, G)$  を表現する  
**識別モデル**を学習するべき

N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian Network Classifiers,”  
Machine Learning, vol.29, no.2, pp.131–163, 1997.

## 1.2 Friedman et al. (1997) による批判

$$\begin{aligned} \text{対数尤度} &= \sum_{d=1}^N \log P(x_0^d, x_1^d, \dots, x_n^d \mid G, \Theta) \\ &= \underbrace{\sum_{d=1}^N \log P(x_0^d \mid x_1^d, \dots, x_n^d, G, \Theta)}_{\text{分類に関する CLL}(G, \Theta \mid D)} + \underbrace{\sum_{d=1}^N \log P(x_1^d, \dots, x_n^d \mid G, \Theta)}_{\text{分類に関与しない}} \end{aligned}$$

分類精度を高めるためには対数尤度  $LL(D \mid G, \Theta)$  ではなく、**条件付き対数尤度 (Conditional Log Likelihood: CLL)** のみを用いるべき。

N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian Network Classifiers,” Machine Learning, vol.29, no.2, pp.131–163, 1997.

## 定義1.3 Conditional MDL (Grossman and Domingos, 2004)

Conditional MDL (CMDL) スコアは以下で定義される.

$$CMDL(G, \Theta | D) = \frac{\log N}{2} \sum_{i=0}^n q_i (r_i - 1) - CLL(G, \Theta | D).$$

D. Grossman and P. Domingos, “Learning Bayesian Network classifiers by maximizing conditional likelihood,” Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004, pp.361–368, 2004.

## 1.4 CLLの問題点

CLLスコアは分解可能ではないため、構造探索に効率的なアルゴリズムが適用できず、厳密学習に膨大な時間がかかってしまう。

# 1.5 CLLの近似手法

- 構造探索に対して山登り法を適用した近似学習 (Grossman et al., 2004)
- CLLが分解可能となるように近似したapproximated CLL (aCLL) スコア (Carvalho et al., 2013)
- CLLスコアが等価な構造を重複して探索しないような貪欲学習アルゴリズム (Mihaljević et al., 2018)

分類精度：

CLLを用いた近似手法 > 周辺尤度を用いた近似手法

D. Grossman and P. Domingos. Learning Bayesian Network classifiers by maximizing conditional likelihood. In Proceedings of the International Conference on Machine Learning, pages 361–368, 2004.

A. M. Carvalho, P. Adão, and P. Mateus. Efficient Approximation of the Conditional Relative Entropy with Applications to Discriminative Learning of Bayesian Network Classifiers. Entropy, 15:2716–2735, 2013.

B. Mihaljević, C. Bielza, and P. Larrañaga. Learning Bayesian network classifiers with completed partially directed acyclic graphs. In Proceedings of the International Conference on Probabilistic Graphical Models, pages 272–283, 2018.

# 1.6 本当にCLLは良いのか？

- 周辺尤度最大化よりCLL最大化の方がなぜ良いのかという理由については未だ明らかになっていない。
- 周辺尤度を最大化する構造を厳密に学習できるにもかかわらず、先行研究では近似学習を行っている。このため、探索精度の悪さが結果に影響したのかもしれない。

# 1.7 周辺尤度厳密学習とCLL近似学習の分類精度比較

従来のCLLを用いた近似学習手法と、周辺尤度を用いた厳密学習手法の分類精度を比較する実験を行った。

# 1.8 周辺尤度厳密学習とCLL近似学習の分類精度比較

## 比較手法

- ・ GBN (厳密) : 周辺尤度を用いて厳密学習したBN
- ・ Naive Bayes (Minsky, 1961)
- ・ BN-CMDL (Grossman and Domingos, 2004) : CMDLを用いて近似学習したBN
- ・ BNC2P (Grossman and Domingos, 2004) : 各変数が最大 2 つまでしか親を持たない構造を候補として, CLLを用いて近似学習したBN
- ・ TAN-aCLL (Carvalho et al., 2013) : aCLLを用いて厳密学習したTAN
- ・ BN (山登) : 周辺尤度を用いて山登り法で近似学習したBN
- ・ MC-DAGGES : CLLスコアが等価な構造を重複して探索しないような貪欲学習アルゴリズム (Mihaljević et al., 2018)

Marvin Minsky. Steps toward Artificial Intelligence. In Proceedings of the IRE, volume 49, pages 8–30, 1961.

D. Grossman and P. Domingos, “Learning Bayesian Network classifiers by maximizing conditional likelihood,” Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004, pp.361–368, 2004.

A. M. Carvalho, P. Ado, and P. Mateus. Efficient Approximation of the Conditional Relative Entropy with Applications to Discriminative Learning of 15 Bayesian Network Classifiers. *Entropy*, 15(7): 2716–2735, 2013.

# 1.8 周辺尤度厳密学習とCLL近似学習の分類精度比較

No.	Dataset	Variables	Classes	Sample size	Naive-Bayes	GBN-CMDL	BNC2P	TAN-aCLL	GBN(山登)	MC-DAG GES	GBN(厳密)
1	Balance Scale	5	3	625	<b>0.9152</b>	0.3333	0.8560	0.8656	<b>0.9152</b>	0.7432	<b>0.9152</b>
2	banknote authentication	5	2	1372	0.8433	<b>0.8819</b>	0.8797	0.8761	<b>0.8819</b>	0.8768	0.8812
3	Hayes-Roth	5	3	132	0.8182	0.6136	0.6894	0.6742	0.7525	0.6970	0.6136
4	iris	5	3	150	0.7133	0.7800	0.8200	0.8200	0.8133	0.7800	<b>0.8267</b>
5	lenses	5	3	24	0.7500	0.8333	0.6667	0.7083	0.8333	0.8333	0.8333
6	Car Evaluation	7	4	1728	0.8571	<b>0.9497</b>	0.9416	0.9433	0.9416	0.9126	0.9416
7	liver	7	2	345	0.6319	0.6145	0.6290	<b>0.6609</b>	0.6029	0.6435	0.6087
8	MONK's Problems	7	2	432	0.7500	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b>	0.8449	<b>1.0000</b>	<b>1.0000</b>
9	mux6	7	2	64	0.5469	0.3750	0.5625	0.4688	0.4063	<b>0.7656</b>	0.4531
10	LED7	8	10	3200	0.7294	0.7366	<b>0.7375</b>	0.7350	0.7297	0.7331	0.7294
11	HTRU2	9	2	17898	0.7031	0.7096	0.7070	0.7018	0.7188	0.7214	<b>0.7305</b>
12	Nursery	9	5	12960	0.6782	<b>0.7126</b>	0.6092	0.5862	<b>0.7126</b>	0.6322	<b>0.7126</b>
13	pima	9	2	768	0.8966	0.9086	0.9118	0.9130	0.9092	0.9093	0.9112
14	post	9	3	87	0.9033	0.5823	<b>0.9442</b>	0.9177	0.9291	0.9046	0.9340
15	Breast Cancer	10	2	277	<b>0.9751</b>	0.8917	0.9473	0.9488	0.7058	0.6354	<b>0.9751</b>
16	Breast Cancer Wisconsin	10	2	683	0.7401	0.6209	0.6823	0.7184	0.7094	<b>0.9780</b>	0.7184
17	Contraceptive Method Choice	10	3	1473	0.4671	0.4501	<b>0.4745</b>	0.4705	0.4440	0.4576	0.4542
18	glass	10	6	214	0.5561	0.5654	0.5794	0.6308	0.4626	0.5888	0.5701
19	shuttle-small	10	6	5800	0.9384	0.9660	0.9703	0.9583	0.9683	0.9586	0.9693
20	threeOf9	10	2	512	0.8164	<b>0.9434</b>	0.8691	0.8828	0.8652	0.8750	0.8887
21	Tic-Tac-Toe	10	2	958	0.6921	<b>0.8841</b>	0.7338	0.7203	0.6754	0.7557	0.8340
22	MAGIC Gamma Telescope	11	2	19020	0.7482	0.7849	0.7806	0.7631	0.7844	0.7781	0.7873
23	Solar Flare	11	9	1389	0.7811	0.8265	0.8315	0.8229	<b>0.8431</b>	0.8013	<b>0.8431</b>
24	heart	14	2	270	0.8259	0.8185	0.8037	0.8148	0.8222	<b>0.8333</b>	0.8259
25	wine	14	3	178	0.9270	<b>0.9438</b>	0.9157	0.9326	0.9045	<b>0.9438</b>	0.9270
26	cleve	14	2	296	0.8412	0.8209	0.8007	<b>0.8378</b>	0.7973	0.8041	0.7973
27	Australian	15	2	690	0.8290	0.8312	0.8348	0.8464	0.8420	0.8406	<b>0.8536</b>
28	crx	15	2	653	0.8377	0.8346	0.8208	0.8560	<b>0.8622</b>	0.8576	0.8591
29	EEG	15	2	14980	0.5778	0.6787	0.6374	0.6125	0.6732	0.6182	0.6814
30	Congressional Voting Records	17	2	232	0.9095	0.9698	0.9612	0.9181	<b>0.9741</b>	0.9009	0.9655
31	zoo	17	5	101	0.9802	0.9109	0.9505	1.0000	0.9505	0.9802	0.9307
32	pendigits	17	10	10992	0.8032	0.9062	0.8719	0.8700	0.9253	0.8359	<b>0.9290</b>
33	letter	17	26	20000	0.4466	0.5796	0.5132	0.5093	0.5761	0.4664	0.5761
34	ClimateModel	19	2	540	0.9222	<b>0.9407</b>	0.9241	0.9333	0.9370	0.9296	0.9000
35	Image Segmentation	19	7	2310	0.7290	0.7918	0.7991	0.7407	0.8026	0.7476	0.8156
36	lymphography	19	4	148	<b>0.8446</b>	0.7939	0.7973	0.8311	0.7905	0.8041	0.7500
37	vehicle	19	4	846	0.4350	0.5910	0.5910	0.5816	0.5461	0.5414	0.5768
38	hepatitis	20	2	80	0.8500	0.7375	<b>0.8875</b>	0.8750	0.8500	<b>0.8875</b>	0.5875
39	German	21	2	1000	0.7430	0.6110	0.7340	<b>0.7470</b>	0.7140	0.7180	0.7210
40	bank	21	2	30488	0.8544	0.8618	0.8928	0.8618	0.8952	0.8708	<b>0.8956</b>
41	waveform-21	22	3	5000	0.7886	0.7862	0.7754	0.7896	0.7698	0.7926	0.7846
42	Mushroom	22	2	5644	0.9957	<b>1.0000</b>	<b>1.0000</b>	0.9995	<b>1.0000</b>	0.9986	0.9949
43	spect	23	2	263	0.7940	0.7940	0.7903	0.8090	0.7603	0.8052	0.7378
	average				0.7764	0.7721	0.7936	0.7943	0.7867	0.7944	0.7963

# 1.9 高い分類精度を示す周辺尤度厳密学習

データセット	Variables	Classes	サンプルサイズ	Naive Bayes	GBN-CMDL	BNC2P	TAN-aCLL	GBN(貪欲)	MC-DAGGES	GBN(厳密)
HTRU2	9	2	17898	0.7031	0.7096	0.707	0.7018	0.7188	0.7214	0.7305
Nursery	9	5	12960	0.6782	0.7126	0.6092	0.5862	0.7126	0.6322	0.7126
MAGIC	11	2	19020	0.7482	0.7849	0.7806	0.7631	0.7844	0.7781	0.7873
EEG	15	2	14980	0.5778	0.6787	0.6374	0.6125	0.6732	0.6182	0.6814

サンプルサイズが大きいとき：

周辺尤度による厳密学習 > CLLによる近似学

習

# 1.10 周辺尤度厳密学習の著しい分類精度低下

サンプルサイズの小さいデータセットではGBN(厳密)の分類精度が他手法よりも著しく悪い。

No.	Dataset	Variables	Classes	Sample size	Naive-Bayes	GBN-CMDL	BNC2P	TAN-aCLL	GBN(山登)	MC-DAG GES	GBN(厳密)
1	Balance Scale	5	3	625	<b>0.9152</b>	0.3333	0.8560	0.8656	<b>0.9152</b>	0.7432	<b>0.9152</b>
2	banknote-authentication	5	2	1372	<b>0.8133</b>	<b>0.8819</b>	0.8797	<b>0.8761</b>	<b>0.8819</b>	0.8768	<b>0.8812</b>
3	Hayes-Roth	5	3	132	0.8182	0.6136	0.6894	0.6742	0.7525	0.6970	0.6136
4	iris	5	3	150	0.7133	0.7800	0.8200	0.8200	0.8133	0.7800	<b>0.8207</b>
5	lenses	5	3	24	0.7500	0.8333	0.6667	0.7083	0.8333	0.8333	0.8333
6	Car Evaluation	7	4	1728	0.8571	<b>0.9497</b>	0.9416	0.9433	0.9416	0.9126	0.9416
7	liver	7	2	345	0.6319	0.6145	0.6290	<b>0.6609</b>	0.6029	0.6435	0.6087
8	MONK's Problems	7	2	432	0.7500	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b>	0.8449	<b>1.0000</b>	<b>1.0000</b>
9	mux6	7	2	64	0.5469	0.3750	0.5625	0.4688	0.4063	<b>0.7656</b>	0.4531
10	LED7	8	10	3200	<b>0.7294</b>	<b>0.7366</b>	<b>0.7373</b>	<b>0.7350</b>	<b>0.7297</b>	<b>0.7331</b>	<b>0.7294</b>
11	HTRU2	9	2	17898	0.7031	0.7096	0.7070	0.7018	0.7188	0.7214	<b>0.7305</b>
12	Nursery	9	5	12960	0.6782	<b>0.7126</b>	0.6092	0.5862	<b>0.7126</b>	0.6322	<b>0.7126</b>
13	pima	9	2	768	0.8966	0.9086	0.9118	0.9130	0.9092	0.9093	0.9112
14	post	9	3	87	0.9033	0.5823	<b>0.9442</b>	0.9177	0.9291	0.9046	0.9340
15	Breast Cancer	10	2	277	<b>0.9751</b>	0.8917	0.9473	0.9488	0.7058	0.6354	<b>0.9751</b>
16	Breast Cancer Wisconsin	10	2	683	0.7401	0.6209	0.6823	0.7184	0.7094	<b>0.9780</b>	0.7184
17	Contraceptive Method Choice	10	3	1473	0.4671	0.4501	<b>0.4745</b>	0.4705	0.4440	0.4576	0.4542
18	glass	10	6	214	0.5561	0.5654	0.5794	0.6308	0.4626	0.5888	0.5701
19	shuttle-small	10	6	5800	0.9384	0.9660	0.9703	0.9583	0.9683	0.9586	0.9693
20	threeOf9	10	2	512	0.8164	<b>0.9434</b>	0.8691	0.8828	0.8652	0.8750	0.8887
21	Tic-Tac-Toe	10	2	958	0.6921	<b>0.8841</b>	0.7338	0.7203	0.6754	0.7557	0.8340
22	MAGIC Gamma Telescope	11	2	19020	0.7482	0.7849	0.7806	0.7631	0.7844	0.7781	0.7873
23	Solar Flare	11	9	1389	0.7811	0.8265	0.8315	0.8229	<b>0.8431</b>	0.8013	<b>0.8431</b>
24	heart	14	2	270	0.8259	0.8185	0.8037	0.8148	0.8222	<b>0.8333</b>	0.8259
25	wine	14	3	178	0.9270	<b>0.9438</b>	0.9157	0.9326	0.9045	<b>0.9438</b>	0.9270
26	cleve	14	2	296	0.8412	0.8209	0.8007	<b>0.8378</b>	0.7973	0.8041	0.7973
27	Australian	15	2	690	0.8290	0.8312	0.8348	0.8464	0.8420	0.8406	<b>0.8536</b>
28	crx	15	2	653	0.8377	0.8346	0.8208	0.8560	<b>0.8622</b>	0.8576	0.8591
29	EEG	15	2	14980	0.5778	0.6787	0.6374	0.6125	0.6732	0.6182	0.6814
30	Congressional Voting Records	17	2	232	0.9095	0.9698	0.9612	0.9181	<b>0.9741</b>	0.9009	0.9655
31	zoo	17	5	101	0.9802	0.9109	0.9505	1.0000	0.9505	0.9802	0.9307
32	pendigits	17	10	10992	0.8032	0.9062	0.8719	0.8700	0.9253	0.8359	<b>0.9290</b>
33	letter	17	26	20000	0.4466	0.5796	0.5132	0.5093	0.5761	0.4664	0.5761
34	ClimateModel	19	2	540	0.9222	<b>0.9407</b>	0.9241	0.9333	0.9370	0.9296	0.9000
35	Image Segmentation	19	7	2310	0.7290	0.7918	0.7991	0.7407	0.8026	0.7476	0.8156
36	lymphography	19	4	148	<b>0.8446</b>	0.7939	0.7973	0.8311	0.7905	0.8041	0.7500
37	vehicle	19	4	846	0.4350	0.5910	0.5910	0.5816	0.5461	0.5414	0.5768
38	hepatitis	20	2	80	0.8500	0.7375	<b>0.8875</b>	0.8750	0.8500	<b>0.8875</b>	0.5875
39	German	21	2	1000	0.7430	0.6110	0.7340	<b>0.7470</b>	0.7140	0.7180	0.7210
40	bank	21	2	30488	0.8544	0.8618	0.8928	0.8618	0.8952	0.8708	<b>0.8956</b>
41	waveform-21	22	3	5000	0.7886	0.7862	0.7754	0.7896	0.7698	0.7926	0.7846
42	Mushroom	22	2	5644	0.9957	<b>1.0000</b>	<b>1.0000</b>	0.9995	<b>1.0000</b>	0.9986	0.9949
43	spect	23	2	263	0.7940	0.7940	0.7903	0.8090	0.7603	0.8052	0.7378
	average					0.7764	0.7721	0.7936	0.7943	0.7867	0.7963

# 1.10 周辺尤度厳密学習の著しい分類精度低下

BN(厳密)の分類精度が他手法より著しく悪い時のデータセット

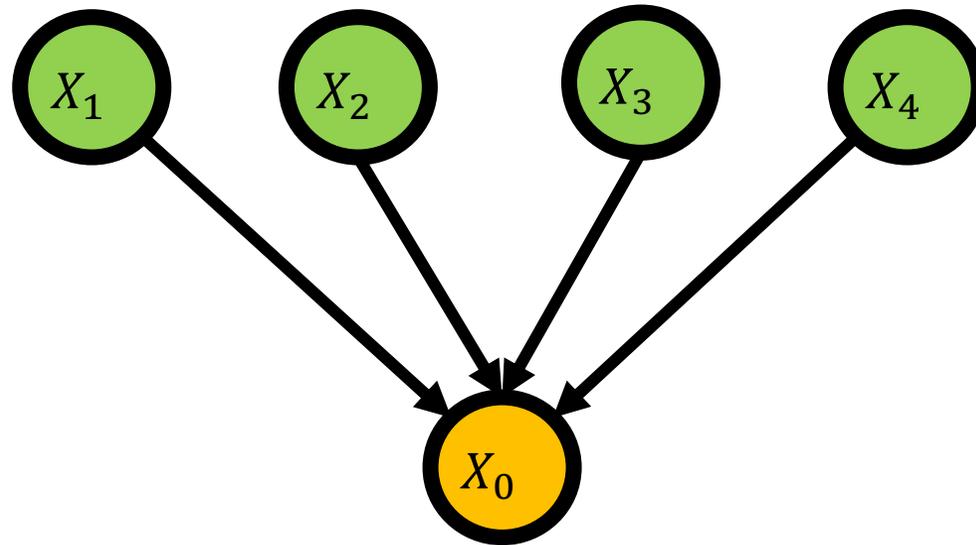
変数	サンプルサイズ	NB	BN-CMDL	BNC2P	TAN-aCLL	GBN(山登)	GBN(厳密)	GBN(厳密)の目的変数の親変数数
5	132	0.8182	0.8333	0.6364	0.6742	0.7879	0.6136	3.0
7	64	0.5469	0.3906	0.5625	0.4688	0.3750	0.4531	5.8
17	101	0.9802	0.8416	0.9505	1.0000	0.9406	0.9307	4.3

この時、周辺尤度による厳密学習では、**目的変数の親変数が多い構造を学習していることがわかった。**

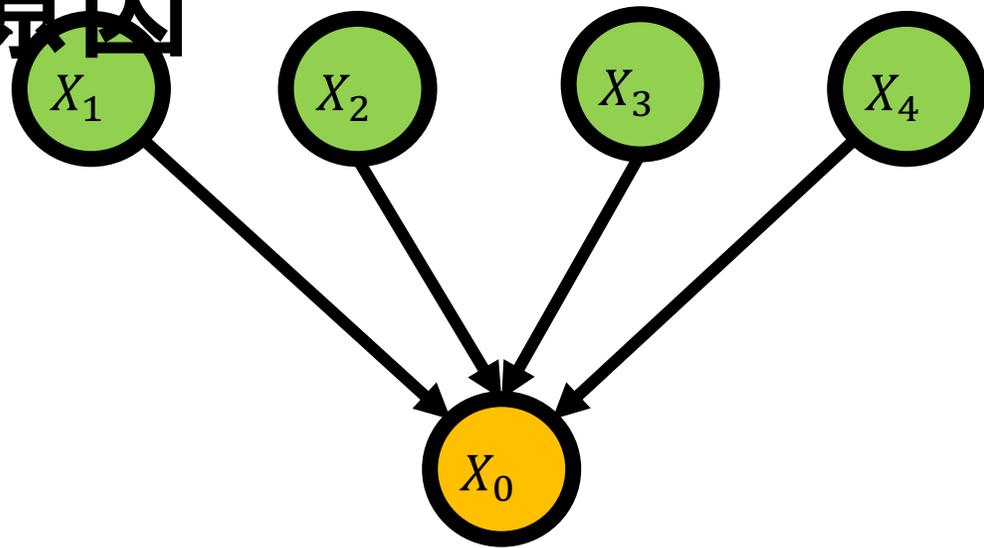
# 1.11 周辺尤度厳密学習の分類精度低下の原因

例. GBN (厳密) の分類精度が著しく悪い時の構造

親変数が多い



# 1.11 周辺尤度厳密学習の分類精度低下の原因



$X_0$ の条件付き確率パラメータ  
 $\theta_{0jk}$ のEAP推定量

$$\frac{N'_{0jk} + N_{0jk}}{N'_{0j} + N_{0j}}$$

目的変数の親変数が増加

- 目的変数の親変数集合の状態 $j$ のパターン数が指数的に増加
- データが欠測するパターン数が増加
- 目的変数のパラメータの推定精度が低下する.

# 1.11 周辺尤度厳密学習の分類精度低下の原因

実際に、目的変数の親変数が過剰な構造では、  
データが欠測するパターン数が増加していた。

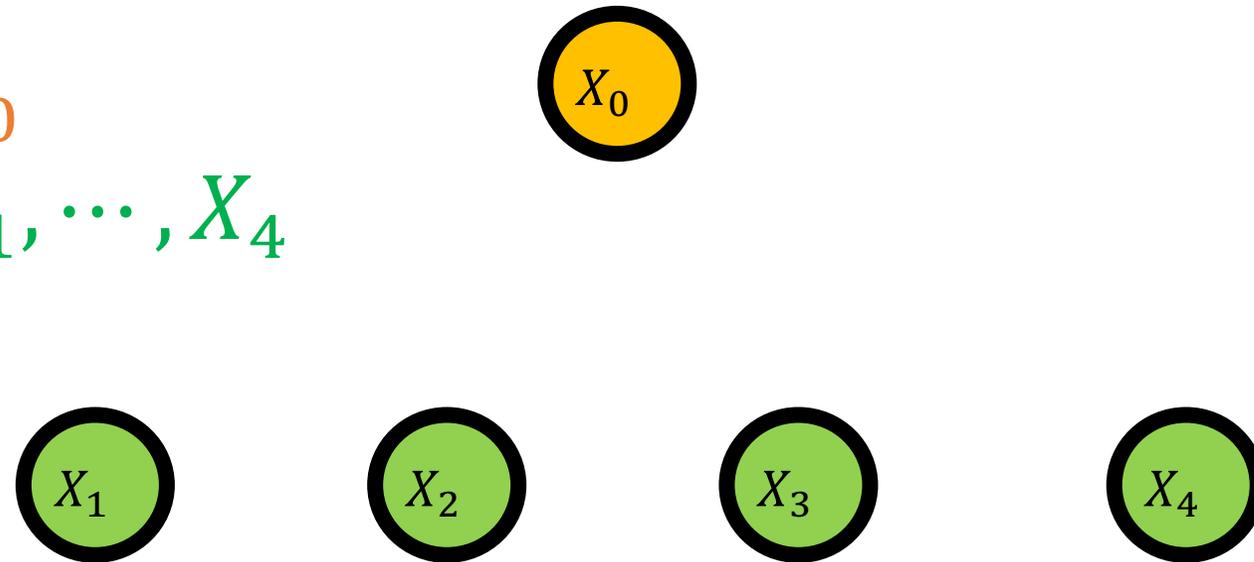
データセット	変数数	サンプルサイズ	目的変数の親変数数	データが欠測するパターン数
Hayes-Roth	5	132	3.0	17.2
mux6	7	64	5.8	5.2
zoo	17	101	4.3	20.3

# 2. Augmented Naive Bayes Classifiers (ANB)

## 2.1 Augmented Naïve Bayes Classifiers (ANB)

目的変数 :  $X_0$

説明変数 :  $X_1, \dots, X_4$



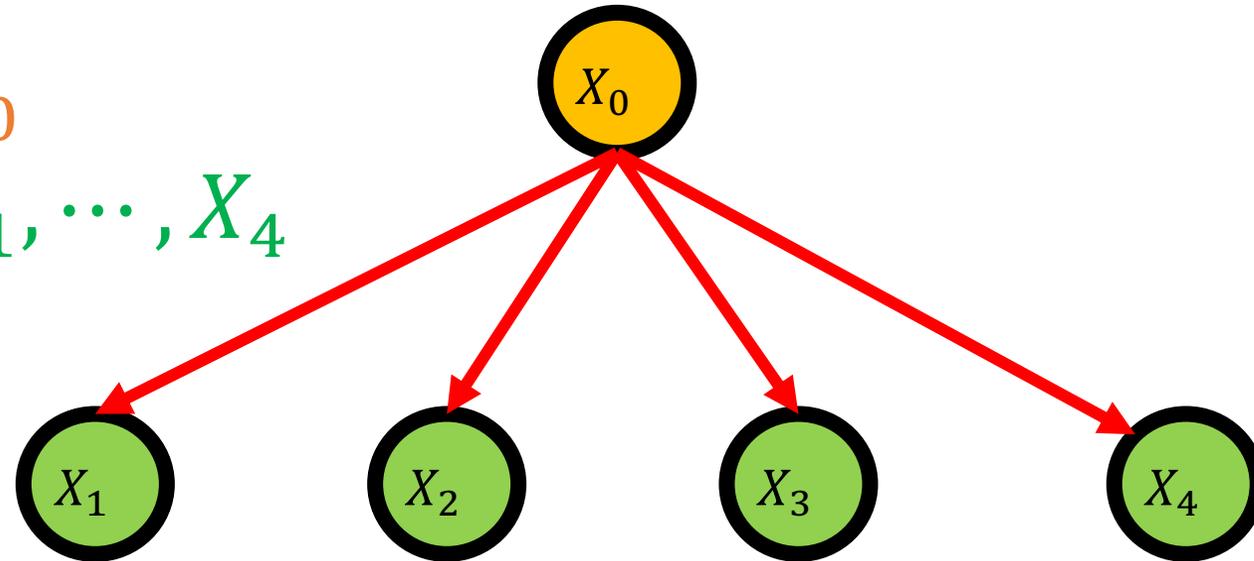
## 2.1 Augmented Naïve Bayes Classifiers

(ANB)

まず強制的に目的変数から全説明変数へエッジを引く。

目的変数 :  $X_0$

説明変数 :  $X_1, \dots, X_4$

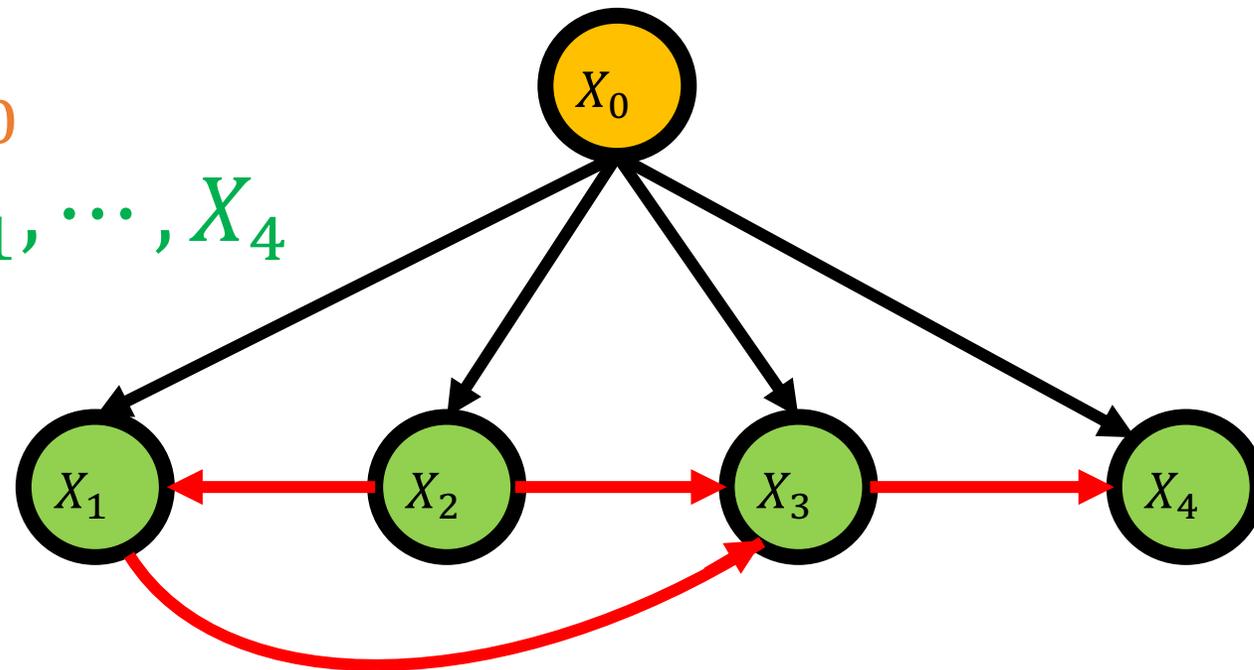


## 2.1 Augmented Naïve Bayes Classifiers

(ANB) まず強制的に目的変数から全説明変数へエッジを引く。

目的変数 :  $X_0$

説明変数 :  $X_1, \dots, X_4$



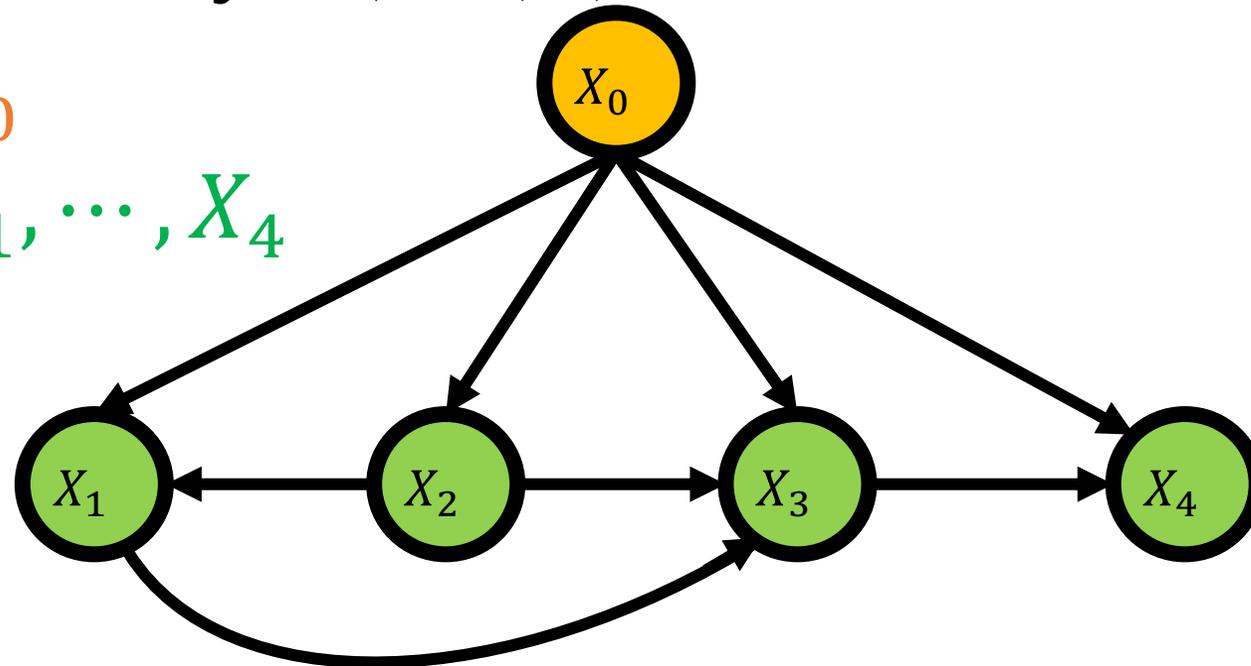
目的変数からのエッジを所与として  
説明変数間の構造を厳密学習。

## 2.1 Augmented Naïve Bayes Classifiers

(ANB) 目的変数から全説明変数にエッジが引かれている構造を, Augmented Naïve Bayes (ANB) (Friedman et al., 1997) と呼ぶ.

目的変数 :  $X_0$

説明変数 :  $X_1, \dots, X_4$



目的変数の親変数が0で, 全ての説明変数を子変数とするため, 先述した分類精度低下の問題が生じない.

N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian Network Classifiers,"  
Machine Learning, vol.29, no.2, pp.131–163, 1997.

## 2.2 ANBの厳密学習アルゴリズム

全てのANB構造の集合から，周辺尤度を最大にする構造を探索．

ANBでは目的変数は親を持たないため，以下のスコアを最大にすればよい．

$$BDeu_{ANB}(G, D) = BDeu(G, D) - LocalBDeu_0(\emptyset, D).$$

Silander and Myllymäki (2006)が提案した動的計画法による通常のベイジアンネットワークのための厳密学習アルゴリズムを，ANB学習に修正．

T. Silander and P. Myllymäki. A Simple Approach for finding the Globally Optimal Bayesian Network Structure. In *Proceedings of Uncertainty in Artificial Intelligence*, pages 445–452, 2006.

## 2.3 Notation

**定義3.3 シンク (Pearl, 1988)**

子変数を持たない変数をシンクと呼ぶ。

**定義3.4 最適親変数集合 (Silander and Myllymäki, 2006)**

変数集合 $\mathbf{Z}$ , ( $X_0 \in \mathbf{Z}$ )のべき集合の中で $X_0$ を含むものの集合を $\Pi(\mathbf{Z})$ とすると,  $X_i$ の $\mathbf{Z}$ に関する最適親変数集合は以下で定義される。

$$g_i^*(\mathbf{Z}) = \operatorname{argmax}_{\mathbf{W} \in \Pi(\mathbf{Z})} LocalBDeu_i(\mathbf{W}, D).$$

- ・ 変数集合 $\mathbf{Z}$ , ( $X_0 \in \mathbf{Z}$ )で構成されるANB構造の中で周辺尤度最大の構造を $G^*(\mathbf{Z})$ で表す。
- ・  $G^*(\mathbf{Z})$ におけるあるシンクを $X_s^*(\mathbf{Z})$ で表す。

T. Silander and P. Myllymäki. A Simple Approach for finding the Globally Optimal Bayesian Network Structure. In *Proceedings of Uncertainty in Artificial Intelligence*, pages 445–452, 2006.

## 2.4 ANBの厳密学習のステップ

1. 説明変数 $X_i \in F$ と変数集合 $Z \subseteq V \setminus \{X_i\}$ , ( $X_0 \in Z$ )の考えられる全ての組み合わせについて, ローカルスコア $LocalBDeu_i(Z, D)$ を計算する.
2. 説明変数 $X_i \in F$ と変数集合 $Z \subseteq V \setminus \{X_i\}$ , ( $X_0 \in Z$ )の考えられる全ての組み合わせについて, 最適親変数集合 $g^*(Z)$ を計算する.
3. すべての変数集合 $Z \subseteq V$ , ( $X_0 \in Z$ )について, シンク $X_S^*(Z)$ を計算する.
4. ステップ2と3を用いて $G^*(V)$ を計算する.

## 2.5 ステップ3の計算方法

$G^*(\mathbf{Z})$ はシンク $X_S^*(\mathbf{Z})$ をもつ.

$G^*(\mathbf{Z})$ において $X_S^*(\mathbf{Z})$ は $g^*(\mathbf{Z} \setminus \{X_S^*(\mathbf{Z})\})$ を親変数集合としてもっている.

$G^*(\mathbf{Z})$ における $X_S^*(\mathbf{Z})$ 以外の変数は $G^*(\mathbf{Z} \setminus \{X_S^*(\mathbf{Z})\})$ を構成する.

したがって,

$$X_S^*(\mathbf{Z}) = \operatorname{argmax}_{X_i \in \mathbf{Z} \setminus \{X_0\}} \{LocalBDeu_i(g_i^*(\mathbf{Z} \setminus \{X_i\}), D) + BDeu_{ANB}(G^*(\mathbf{Z} \setminus \{X_i\}), D)\}.$$

## 2.6 ステップ4の計算方法

$X_S^*(\mathbf{Z}) = \operatorname{argmax}_{X_i \in \mathbf{Z} \setminus \{X_0\}} \{LocalBDeu_i(g_i^*(\mathbf{Z} \setminus \{X_i\}), D) + BDeu_{ANB}(G^*(\mathbf{Z} \setminus \{X_i\}), D)\}$  の式より,  $G^*(\mathbf{Z})$  は  $G^*(\mathbf{Z} \setminus \{X_S^*(\mathbf{Z})\})$  と,  $g^*(\mathbf{Z} \setminus \{X_S^*(\mathbf{Z})\})$  から  $X_S^*(\mathbf{Z})$  に向かうエッジ集合と  $X_S^*(\mathbf{Z})$  に分解できる.

$G^*(V)$  から再帰的に分解を行うことで, 最終的にシンクとその最適親変数集合のペア  $n$  組に分解できる.

要素数の小さい  $\mathbf{Z}$  から順に  $G^*(\mathbf{Z})$  と  $X_S^*(\mathbf{Z})$  を計算していくことで, 最後に  $G^*(V)$  が得られる.

## 2. / ANB厳密学習アルゴリズムの効率

性

このアルゴリズムで計算されるローカルスコア, 最適親変数集合, シンクの数はいずれもそれぞれ  $(n-1)2^{n-2}$ ,  $(n-1)2^{n-2}$ ,  $2^{n-1}$  である.

ANB制約を課さないアルゴリズム (Silander and Myllymäki, 2006) におけるそれぞれの計算個数は  $n2^{n-1}$ ,  $n2^{n-1}$ ,  $2^n$  である.

ANB厳密学習アルゴリズムはANB制約を課さないアルゴリズムより約2倍速いと考えられる.

T. Silander and P. Myllymäki. A Simple Approach for finding the Globally Optimal Bayesian Network Structure. In *Proceedings of Uncertainty in Artificial Intelligence*, pages 445–452, 2006.

## 2.8 ANB厳密学習の漸近的性質①

### 定理2.8

$N \rightarrow \infty$ のとき, 厳密学習したANBはパラメータ数最小のI-map ANBに概収束する.

## 2.9 ANB厳密学習の漸近的性質②

### 定理2.9

以下の仮定1～3のもとで，厳密学習したANBは，真の構造 $G^*$ について，以下の関係を満たすような構造 $\hat{G}$ に概収束する．

任意の有限データセット $D'$ について，

$$P(X_0 \mid \mathbf{F}, \hat{G}, D') = P(X_0 \mid \mathbf{F}, G^*, D').$$

この関係を $\hat{G}$ と $G^*$ が分類等価であると呼ぶ．

仮定1 パーフェクトマップが存在する．

仮定2 すべての説明変数が，分類に影響を及ぼす変数集合（真の構造における目的変数のマルコフブランケット）に含まれる．

仮定3 真の構造において目的変数のマルコフブランケットに含まれる変数は目的変数と隣接する．



## 2.11 分類等価による利点

ANBの制約は一般にパラメータ数を増加させてしまうデメリットがあるが、仮定1~3のもとで、 $N \rightarrow \infty$ のとき、厳密学習されたANBは真の構造と全く同じ分類性能を持つ構造に概収束する。

## 2.12 定理の実証実験

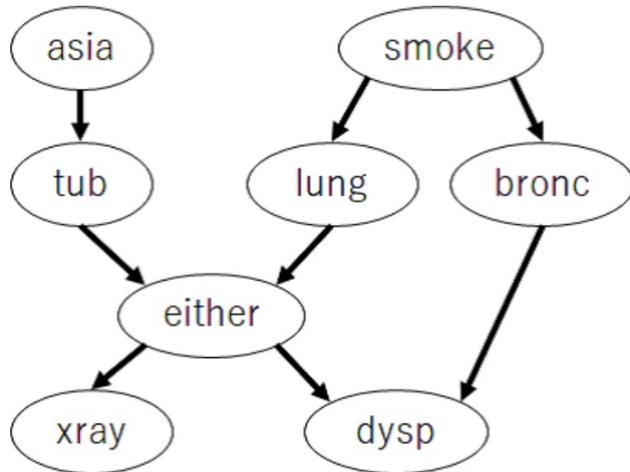
仮定2と3を共に満たさないネットワークASIAと、共に満たすネットワークCancerを用いる。

真の構造とANB厳密学習それぞれの分類確率を計算し、Kullback-Leibler divergence (KLD)を測定した。

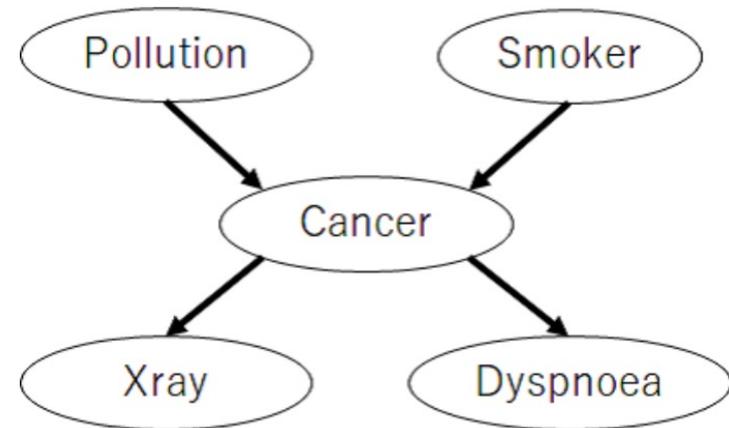
パラメータ数最小のANBと、ANB厳密学習の推定構造のstructural Hamming distance (SHD)を測定した。

SHDは構造間の距離のようなものを表す。

各ネットワークから100, 500, 1,000, 10,000, 50,000, 100,000サンプル発生させ、それぞれについてKLDとSHDを測定した。



仮定2と3を共に満たさないネットワークASIA



仮定2と3を共に満たすネットワークCancer

## 2.13 パラメータ数最小のI-map ANBを学習できることの実証

Network	Variables	Sample size	SHD-(Proposal, I-map ANB)	KLD-(Proposal, True structure)
仮定2と3を満たさないネットワーク ASIA	8	100	3	$2.31 \times 10^{-2}$
		500	2	$1.24 \times 10^{-1}$
		1000	2	$7.63 \times 10^{-2}$
		5000	1	$3.67 \times 10^{-3}$
		10000	0	$9.26 \times 10^{-4}$
		50000	0	$6.28 \times 10^{-4}$
		100000	0	$3.59 \times 10^{-5}$
仮定2と3を満たすネットワーク CANCER	5	100	1	$8.79 \times 10^{-2}$
		500	1	$2.43 \times 10^{-3}$
		1000	0	0.00
		5000	0	0.00
		10000	0	0.00
		50000	0	0.00
		100000	0	0.00

## 2.14 真の構造と分類等価な構造を学習できることの実証

Network	Variables	Sample size	SHD-(Proposal, I-map ANB)	KLD-(Proposal, True structure)
仮定 2 と 3 を満たさないネットワーク ASIA	8	100	3	$2.31 \times 10^{-2}$
		500	2	$1.24 \times 10^{-1}$
		1000	2	$7.63 \times 10^{-2}$
		5000	1	$3.67 \times 10^{-3}$
		10000	0	$9.26 \times 10^{-4}$
		50000	0	$6.28 \times 10^{-4}$
		100000	0	$3.59 \times 10^{-5}$
仮定 2 と 3 を満たすネットワーク CANCER	5	100	1	$8.79 \times 10^{-2}$
		500	1	$2.43 \times 10^{-3}$
		1000	0	0.00
		5000	0	0.00
		10000	0	0.00
		50000	0	0.00
		100000	0	0.00

## 2.14 真の構造と分類等価な構造を学習できることの実証

Network	Variables	Sample size	SHD-(Proposal, I-map ANB)	KLD-(Proposal True structure)
仮定 2 と 3 を満たさないネットワーク ASIA	8	100	3	$2.31 \times 10^{-2}$
		500	2	$1.24 \times 10^{-1}$
		1000	2	$7.63 \times 10^{-2}$
		5000	1	$3.67 \times 10^{-3}$
		10000	0	$9.26 \times 10^{-4}$
		50000	0	$6.28 \times 10^{-4}$
		100000	0	$3.59 \times 10^{-5}$
仮定 2 と 3 を満たすネットワーク CANCER	5	100	1	$8.79 \times 10^{-2}$
		500	1	$2.43 \times 10^{-3}$
		1000	0	0.00
		5000	0	0.00
		10000	0	0.00
		50000	0	0.00
		100000	0	0.00

## 2.15 ANB厳密学習が分類精度を改善

	Naive-Bayes	GBN-CMDL	BNC2P	TAN-aCLL	BN(山登)	MC-DAG GES	BN(厳密)	ANB(厳密)
average	0.7764	0.7721	0.7936	0.7943	0.7867	0.7944	0.7963	0.8061
<i>p</i> -value (ANB-BDeu vs. the other methods)	0.00308	0.04136	0.00672	0.05614	0.06876	0.06010	0.22628	-

BN(厳密)で目的変数が親変数を多くもっていた時のデータセットの分類精度

変数数	サンプルサイズ	BN(厳密)	ANB(厳密)
5	132	0.6136	0.8333
7	64	0.4531	0.5469
17	101	0.9307	0.9505

## 2.16 変数選択

分類等価の定理では，すべての説明変数が真の構造における目的変数のマルコフブランケットに含まれることを仮定しているが，それは一般には成り立たない。

この問題を解決するには，事前に目的変数のマルコフブランケットのみを変数選択する必要がある。

仮定3が成り立つとき，目的変数のマルコフブランケットは目的変数の親変数と子変数 (Parents and Children : PC) の集合に一致するため，これを探索する。

PC集合の厳密学習法として，SSL (Niinimäki and Parviainen, 2012) や  $S^2$ TMB (Gao and Ji, 2017) があるが，変数数が増加すると計算時間が指数的に増加するため，30変数程度が学習の限界。

Teppo Niinimäki and Pekka Parviainen. Local Structure Discovery in Bayesian Networks. In Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence, UAI'12, pages 634–643. AUAI Press, 2012. ISBN 9780974903989.

Tian Gao and Qiang Ji. Efficient score-based Markov Blanket discovery. International Journal of Approximate Reasoning, 80:277–293, 2017. ISSN 0888-613X. doi: <https://doi.org/10.1016/j.ijar.2016.09.009>.

## 2.17 PC探索手法

PCの厳密学習手法より効率的な手法として条件付き独立性検定(CIテスト)を用いた以下のような変数選択手法が知られている.

- MMPC (Tsamardinos et al. 2006)
- HITON-PC (Aliferis et al. 2003)
- PCMB (Pena 2007)

これらの手法では, 目的変数と説明変数の間でCIテストを行い, 独立性が検出された説明変数はPC集合から取り除く.

しかし, 上記の手法はすべてCIテストとして条件付き相互情報量や統計的検定を用いており, 漸近的に真の独立性を検出する保証がない.

I. Tsamardinos, L.E. Brown, and C.F. Aliferis. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*, 65(1):31–78, 2006.

C.F. Aliferis, I. Tsamardinos, and A. Statnikov. HITON: A Novel Markov Blanket Algorithm for Optimal Variable Selection. *AMIA Annual Symposium proceedings*, pages 21–25, 2003.

J.M. Pena, R. Nilsson, J. Björkegren, J. Tegnér. Towards scalable and data efficient learning of Markov boundaries *International Journal of Approximate Reasoning*, 45 (2) (2007), pp. 211-232

## 2.18 Bayes factor

漸近的に真の独立性を検出するCIテストとして, Steck and Jaakkola (2002)が周辺尤度を用いたBayes factorを提案している.

**定義3.12** 周辺尤度によるBayes factor (Steck and Jaakkola (2002))

二変数 $X, Y$ と変数集合 $\mathbf{Z}$ について, 周辺尤度を用いた対数Bayes factor  $\log BF_D(X, Y | \mathbf{Z})$ は以下で定義される.

$$\log BF_D(X, Y | \mathbf{Z}) = \text{LocalBDeu}_X(\mathbf{Z}, D) - \text{LocalBDeu}_Y(\mathbf{Z} \cup \{Y\}, D).$$

$BF_D(X, Y | \mathbf{Z}) \geq 1$ のとき $I(X, Y | \mathbf{Z})$ と判定し,

$BF_D(X, Y | \mathbf{Z}) < 1$ のとき $\neg I(X, Y | \mathbf{Z})$ と判定する.

Harald Steck and Tommi S. Jaakkola. On the Dirichlet Prior and Bayesian Regularization. In Proceedings of the 15th International Conference on Neural Information Processing Systems, NIPS'02, pages 713–720. MIT Press, 2002b.

## 2.19 Bayes factorを用いた変数選択を適用した ANB厳密学習は最高の分類精度を示す

	Naive-Bayes	GBN-CMDL	BNC2P	TAN-aCLL	GBN(山登)	MC-DAG GES	GBN(厳密)	ANB(厳密)	ANB(厳密, 変数選択適用)
average	0.7764	0.7721	0.7936	0.7943	0.7867	0.7944	0.7963	0.8061	<b>0.8184</b>
<i>p</i> -value ( <i>fsANB-BDeu</i> vs. the other methods)	0.00001	0.00014	0.00013	0.00280	0.00015	0.00212	0.00064	0.01101	-

分類精度：

ANB(厳密, 変数選択適用) > 比較手法 (有意水準 0.05)

# 3. 大規模ANBの学習

## 3.1 スコアベースアプローチの問題点

スコアベースアプローチは、構造の探索数がノード数に対し指数的に増加してしまう。

## 3.2 厳密学習手法（学習可能な変数数）

- 動的計画法：29変数
  - T. Silander and P. Myllymaki, “A simple approach for finding the globally optimal Bayesian network structure,” in Uncertainty in Artificial Intelligence (UAI), 445–452, AUAI Press, 2006
- A\* 探索：29変数
  - Yuan, C., and Malone, B., Learning optimal Bayesian networks: A shortest path perspective. *Journal of Artificial Intelligence Research*, 48, 23–65, 2013.
- 幅優先分枝限定法：33変数
  - Malone, B., Yuan, C., Hansen, E., and Bridges, Improving the scalability of optimal Bayesian network learning with external-memory frontier breadth-first branch and bound search. in *Uncertainty in Artificial Intelligence*, 479–488, 2011
- 整数計画法：60変数
  - J. Cussens, “Bayesian network learning with cutting planes,” in *Uncertainty in Artificial Intelligence (UAI)*, 153–160, AUAI Press, 2011.

## 3.2 厳密学習手法（学習可能な変数数）

- 動的計画法：29変数

- T. Silander and P. Myllymaki, “A simple approach for finding the globally optimal Bayesian network structure,” in Uncertainty in Artificial Intelligence (UAI), 445–452, AUAI Press, 2006.

厳密学習手法の問題点

最先端手法を用いても**60変数程度が限界**

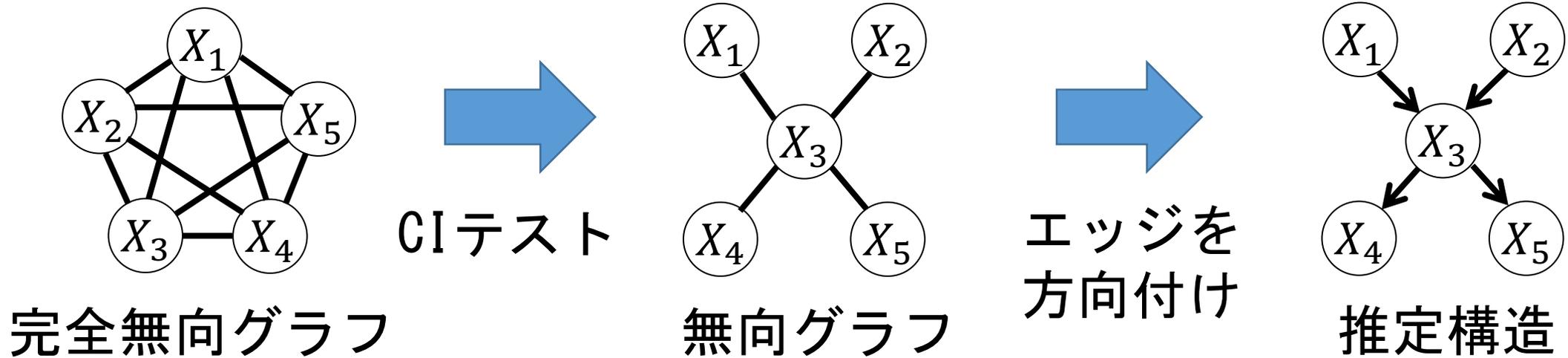
first branch and bound search. in Uncertainty in Artificial Intelligence, 479–488, 2011

- 整数計画法：60変数

- J. Cussens, “Bayesian network learning with cutting planes,” in Uncertainty in Artificial Intelligence (UAI), 153–160, AUAI Press, 2011.

# 3.3 制約ベースアプローチ

- 漸近的に真の構造を学習する保証は持たないが効率的な学習法
- 条件付き独立性検定 (CI テスト) によるエッジの削除とエッジの方向付けによる構造学習手法



## 3.4 制約ベースアプローチの問題点

### 制約ベースアプローチの従来手法

- ・ PCアルゴリズム (Spirtesら, 2000)
- ・ MMHCアルゴリズム (Tsamardinos, 2006)
- ・ RAIアルゴリズム (Yehezkel and Lerner, 2009)

従来手法が用いるCIテストは漸近的に真の独立性を検出する保証がない。

P. Spirtes, C. Glymour, and R. Scheines, Causation, Prediction, and Search, MIT press, 2000.

I. Tsamardinos, L.E. Brown, and C.F. Aliferis, “The max-min hill-climbing Bayesian network structure learning algorithm,” Machine Learning, vol.65, no.1, pp.31–78, 2006.

R. Yehezkel and B. Lerner, “Bayesian network structure learning by recursive autonomy identification,” Journal of Machine Learning Research, vol.10, pp.1527–1570, 2009.

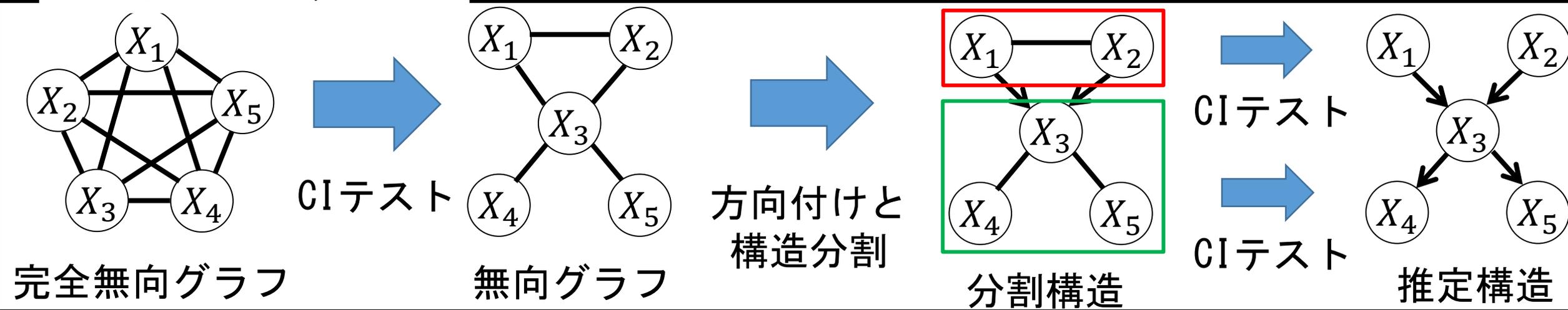
### 3.5 Bayes factorを用いたRAIアルゴリズム (Natoriら, 2017)

漸近一致性を有しつつ最も大規模な構造を学習できる手法

Bayes factor

$$CI(X, Y | \mathbf{Z}) = \begin{cases} \text{従属} & (\log BF_D(X, Y | \mathbf{Z}) \geq 0) \\ \text{独立} & (\log BF_D(X, Y | \mathbf{Z}) < 0) \end{cases}$$

#### RAIアルゴリズム



# 3.6 Bayes factorを用いたRAIアルゴリズム

## 学習手順

入力：データ，完全無向グラフ，CIテストの次数 $n_z = 0$

出力：推定されたグラフ

終了条件：各ノードが $n_z + 1$ 個未満の親ノードを持つ

1. 各エッジに対してBayes factorを用いたCIテストとエッジの削除
2. 各エッジを方向付け
3. グラフを部分グラフに分割
4.  $n_z \leftarrow n_z + 1$ として終了条件を満たすまで再帰的に動作

# 3.7 RAIANBアルゴリズムによる学習

完全無向グラフに制約を加え, CIテストの実行範囲を制限することでANB構造を学習する.

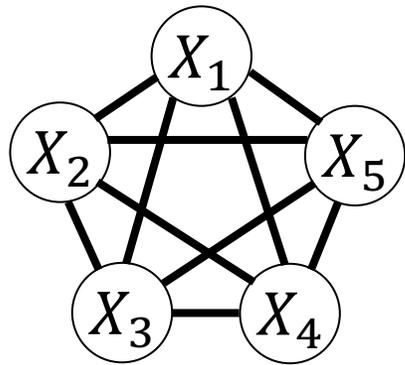
## 学習手順

入力: データ, 説明変数からなる完全無向グラフ,  
CIテストの次数 $n_z = 1$  (条件部が $X_0$ を必ず含む), 出力: 推定されたグラフ

1. 説明変数間の各エッジに対してBayes factorを用いたCIテストとエッジの削除
2. 各エッジを方向付け
3. 全体グラフを部分グラフに分割
4.  $n_z + 1$ 個未満の親を持つノードが存在する場合,  $n_z \leftarrow n_z + 1$ として再帰的に動作
5. 目的変数から全ての説明変数へ向けてエッジを引く.

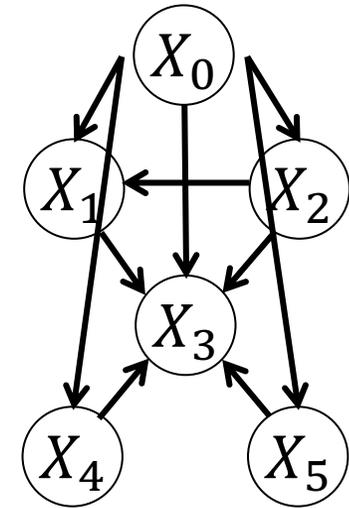
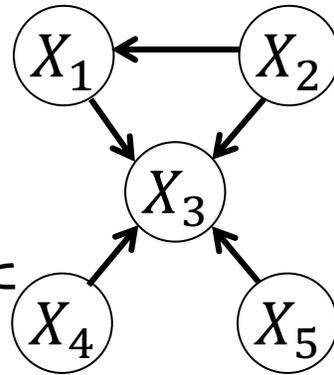
# 3.8 RAIANBの動作例

- 目的変数を含まない2変数に対してCIテストを行いエッジを削除
- 目的変数から全ての説明変数へ向けてエッジを方向付ける



説明変数からなる完全グラフ

→  
 $X_0$ を所与とした  
CIテスト



推定されたANB

## 3.9 RAIANBアルゴリズムの効率性

RAIANBアルゴリズムは以下の理由により通常のRAIアルゴリズムよりも高速に学習できる。

- ・ RAIANBアルゴリズムでは目的変数と説明変数間のCIテストを行う必要がない。
- ・ 真の構造がANB構造である場合，最初から目的変数を所与としたCIテストを行うRAIANBの方が通常のRAIより早く独立性を検出するため，早期にエッジを除去することができる。これにより，アルゴリズム中の構造の分解が加速する。

# 3.10 RAIANBアルゴリズムの漸近的性質

## 定理4.10

$N \rightarrow \infty$ のとき, RAIANBアルゴリズムで学習した構造は, パラメータ数最小のI-map ANBに概収束する.

# 3.11 RAIANBと従来手法の分類精度比較

- Naïve Bayes
  - 全ての説明変数が目的変数のみを親に持つ
- Tree Augmented Naïve Bayes (TAN)
  - 全ての説明変数が目的変数を親に持ち、説明変数間で木構造をとる
- 厳密学習手法 (数十変数が限界)
  - GBN (厳密)
    - 動的計画法で厳密学習したGBN
  - ANB (厳密)
    - 動的計画法で厳密学習したANB
- 制約ベース手法
  - RAI-GBN
    - CIテストにBayes factorを用いて構造学習したGBN
  - RAI-ANB
    - RAIANBアルゴリズムで学習したANB

### 3.12 小規模データにおける精度

	dataset	variable	number of data	classes	Naive Bayes	TAN	GBN-CMDL	BNC2P	TAN-aCLL	GBN (厳密)	ANB (厳密)	RAI-GBN	RAI-ANB
1	magic	11	19020	2	0.7447	0.7767	0.7849	0.7806	0.7631	<b>0.7865</b>	0.7863	0.7793	0.7790
2	Flare	11	1389	9	0.7804	0.7976	0.8265	0.8315	0.8229	<b>0.8430</b>	0.8265	0.8423	0.8178
3	heart	14	270	2	0.8296	0.8407	0.8185	0.8037	0.8148	<b>0.8444</b>	0.8148	0.7666	0.8333
4	wine	14	178	3	0.9205	0.9212	0.9438	0.9157	0.9326	0.9424	<b>0.9490</b>	0.9212	0.9150
5	Cleve	14	296	2	0.8309	0.8175	0.8209	0.8007	<b>0.8378</b>	0.8144	0.8309	0.7771	0.8212
6	Australian	15	690	2	0.8362	0.8304	0.8312	0.8348	0.8464	<b>0.8492</b>	0.8449	0.8405	0.8463
7	crx	15	653	2	0.8391	0.8483	0.8346	0.8208	<b>0.8560</b>	0.8481	0.8482	0.8544	0.8436
8	EEG	15	14980	2	0.5774	0.6298	0.6787	0.6374	0.6125	0.6843	<b>0.6937</b>	0.6421	0.6709
9	Congressional	17	232	2	0.9137	0.9398	0.9698	0.9612	0.9181	<b>0.9699</b>	<b>0.9699</b>	0.9655	0.9438
10	zoo	17	101	5	0.9709	0.9427	0.9109	0.9505	<b>1.0000</b>	0.9900	0.9700	0.8809	0.9418
11	pendigits	17	10992	10	0.7998	0.8477	0.9062	0.8719	0.8700	<b>0.9329</b>	0.9326	0.8757	0.9254
12	letter	17	20000	26	0.4456	0.4866	0.5796	0.5132	0.5093	0.5777	0.5950	0.5560	<b>0.6145</b>
13	ClimateModel	19	540	2	0.9203	0.9314	<b>0.9407</b>	0.9241	0.9333	0.9259	0.9055	0.9074	0.9203
14	ImageSegmentation	19	2310	7	0.7324	0.7510	0.7918	0.7991	0.7407	0.8233	<b>0.8290</b>	0.7839	0.8121
15	lymphography	19	148	4	0.8523	0.8109	0.7939	0.7973	0.8311	<b>0.8647</b>	0.7909	0.6842	0.8514
16	vehicle	19	846	4	0.4266	0.5472	0.5910	0.5910	0.5816	0.5910	<b>0.6417</b>	0.4893	0.6028
17	hepatitis	20	80	2	0.8750	0.8750	0.7375	0.8875	0.8750	<b>0.9250</b>	0.9000	0.8125	0.8875
18	German	21	5000	2	0.7440	0.7340	0.6110	0.7340	0.7470	0.7320	0.7420	0.7000	<b>0.7540</b>
19	bank	21	30488	2	0.8542	0.8774	0.8618	0.8928	0.8618	0.8954	0.8956	<b>0.8959</b>	0.8926
20	waveform-21	22	5000	3	0.7894	0.7914	0.7862	0.7754	0.7896	0.7938	<b>0.8048</b>	0.7328	0.7870
21	Mushroom	22	5644	2	0.9962	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b>	0.9995	0.9946	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b>
22	spect	23	263	2	0.7868	0.8101	0.7940	0.7903	0.8090	0.7759	<b>0.8207</b>	0.7937	0.8096
	Classification accuracy	Arithmetic average			0.7939	0.8094	0.8097	0.8143	0.8160	<b>0.8366</b>	0.8360	0.7955	0.8304
		p-value			0.0024	0.0117	0.0324	0.0099	0.0574	> 0.1	> 0.1	0.0013	-
	Runtime (s)	Arithmetic average			0.00	2.58	30.53	21.11	10.05	1790.93	500.76	26.06	3.14
		Geometric average			0.00	0.798	9.50	6.87	3.27	201.76	110.69	7.90	1.26

### 3.13 大規模データにおける精度

	dataset	variables	num of data	classes	Naive Bayes	TAN	RAI-GBN	RAI-ANB
1	kr-vs-kp	37	3196	2	0.8773	0.9239	0.9405	<b>0.9518</b>
2	Connect-4	43	67557	3	0.7212	0.7643	0.7467	<b>0.7973</b>
3	Flowmeters D	44	180	4	<b>0.8388</b>	<b>0.8388</b>	0.8055	0.8277
4	movement libras	91	360	15	0.5027	0.5388	0.1611	<b>0.5666</b>
5	dota2	117	102944	2	<b>0.5980</b>	0.5810	0.5435	0.5957
6	Musk1	167	478	2	0.6538	0.7565	0.6658	<b>0.8219</b>
7	Musk2	167	6598	2	0.7443	0.8408	0.8808	<b>0.9639</b>
8	Epileptic Seizure	179	11500	5	0.2344	0.3650	0.1886	<b>0.3820</b>
9	mfeat-fac	219	2000	10	0.3520	0.4590	0.3030	<b>0.4730</b>
10	semeion	257	1600	10	0.8556	0.8719	0.4106	<b>0.8794</b>
11	madelon	501	2000	2	0.5905	0.5270	<b>0.6280</b>	0.5830
12	pd speech features	755	756	2	0.7182	0.7897	0.7657	<b>0.8228</b>
13	pure-spectra-matrix	<b>1301</b>	571	20	0.9088	0.8984	0.4833	<b>0.9159</b>
	Classification accuracy	Arithmetic average			0.6612	0.7042	0.5787	<b>0.7370</b>
		p-value			0.0044	0.0012	0.0015	-
	Runtime (s)	Arithmetic average			0.0	545.7	2002.1	1665.9
		Geometric average			0.0	52.6	375.3	227.4

### 3.13 大規模データにおける精度

	dataset	variables	num of data	classes	Naive Bayes	TAN	RAI-GBN	RAI-ANB
1	kr-vs-kp	37	3196	2	0.8773	0.9239	0.9405	<b>0.9518</b>
2	Connect-4	43	67557	3	0.7212	0.7643	0.7467	<b>0.7973</b>
3	Flowmeters D	44	180	4	<b>0.8388</b>	<b>0.8388</b>	0.8055	0.8277
4	movement libras	91	360	15	0.5027	0.5388	0.1611	<b>0.5666</b>
5	dota2	117	102944	2	<b>0.5980</b>	0.5810	0.5435	0.5957
6	Musk1	167	478	2	0.6538	0.7565	0.6658	<b>0.8219</b>
7	Musk2	167	6598	2	0.7443	0.8408	0.8808	<b>0.9639</b>
8	Epileptic Seizure	179	11500	5	0.2344	0.3650	0.1886	<b>0.3820</b>
9	mfeat-fac	219	2000	10	0.3520	0.4590	0.3030	<b>0.4730</b>
10	semeion	257	1600	10	0.8556	0.8719	0.4106	<b>0.8794</b>
11	madelon	501	2000	2	0.5905	0.5270	<b>0.6280</b>	0.5830
12	pd speech features	755	756	2	0.7182	0.7897	0.7657	<b>0.8228</b>
13	pure-spectra-matrix	1301	571	20	0.9088	0.8984	0.4833	<b>0.9159</b>
	Classification accuracy	Arithmetic average			0.6612	0.7042	0.5787	<b>0.7370</b>
		p-value			0.0044	0.0012	0.0015	-
	Runtime (s)	Arithmetic average			0.0	545.7	2002.1	1665.9
		Geometric average			0.0	52.6	375.3	<b>227.4</b>

# 4. 分類影響パラメータ数最小化による ベイジアンネットワーク分類器学習

Shouta Sugahara, Koya Kato and Maomi Ueno: Learning Bayesian Network Classifiers to Minimize the Class Variable Parameters. Proceedings of the AAAI Conference on Artificial Intelligence, 38(18), 20540-20549. (2024) <https://doi.org/10.1609/aaai.v38i18.30039>.

## 4.1 NCP最小のI-map

BN分類器では、分類確率の推定に全ての変数のパラメータを用いるわけではなく、分類確率の推定に必要な変数のパラメータのみを用いる。

従来手法はパラメータ数最小のI-map ANBが得られることは保証するが、分類確率の推定に全ての変数のパラメータ数 (Number of Class variable Parameters: NCP) を最小化した方が高精度な分類器が構成できると考えられる。

$$NCP(G) = \sum_{i=0}^n NCP_i(\mathbf{Pa}_{X_i}^G), \quad NCP_i(\mathbf{Pa}_{X_i}^G) = \begin{cases} (r_i - 1)q_i & (i = 0 \vee X_0 \in \mathbf{Pa}_{X_i}^G) \\ 0 & (\text{その他の場合}) \end{cases}$$

## 4.2 変数順序

変数順序： 構造 $G$ の各変数を要素とするベクトル $\pi$ に対し、 $\pi$ の $i$ 番目の要素を $X_{\pi_i}$ で表すと、 $\forall i, \text{Pa}_{X_{\pi_i}}^G \subseteq \cup_{j=1}^{i-1} \{X_{\pi_j}\}$  が成り立つ時、 $\pi$ を変数順序という。

ここで、 $\text{Pa}_{X_i}^G$ は変数 $X_i$ の親変数集合。

例) 変数順序 $(X_0, X_1, X_2, X_3)$ に従う構造

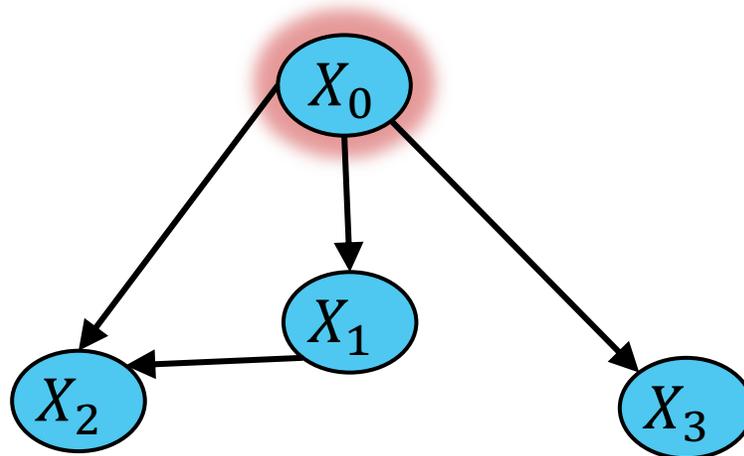


図 変数順序 $(X_0, X_1, X_2, X_3)$ に従う構造の例

## 定理 4.3

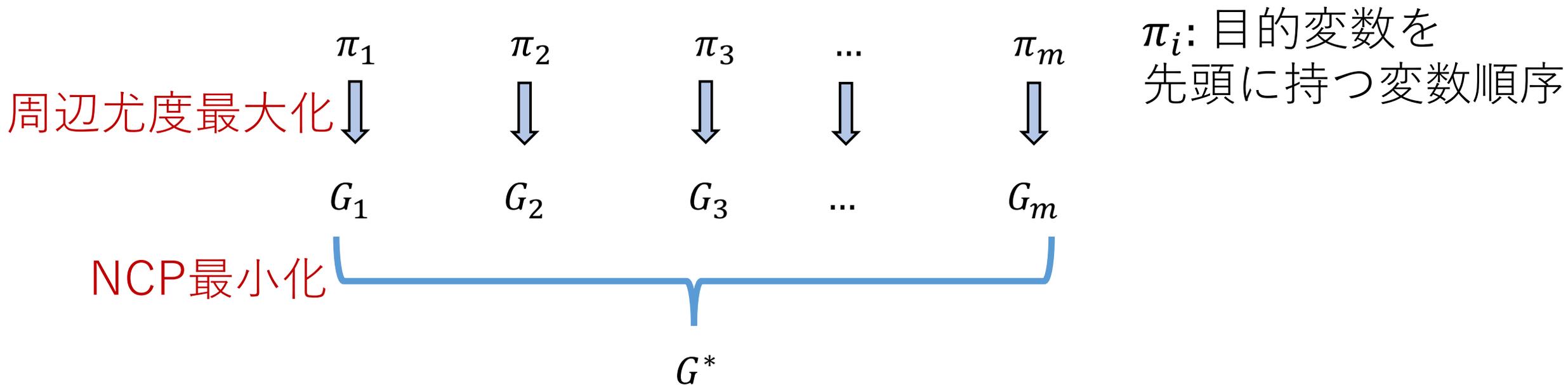
変数順序を所与として周辺尤度を最大にする構造は、その変数順序に従う真の分類確率に漸近収束する構造の中で目的変数パラメータ数が最小の構造に一致する。

## 4.4 学習手順

以下の2つのステップから構成される

1. 目的変数から始まる全ての変数順序について，周辺尤度を最大化する構造をそれぞれ求める
2. 第1ステップで得られた構造のうち目的変数パラメータ数  
最小の構造を探索する

## 4.5 学習手順の図



$G^*$  : 目的変数パラメータ数を最小にして  
真の分類確率に漸近収束する構造

## 4.6 最短パス探索問題への定式化

目的変数パラメータ数をコストとした重み付きグラフ  
(NPCリバーソオーダーグラフ)の最短パス探索問題として  
定式化.

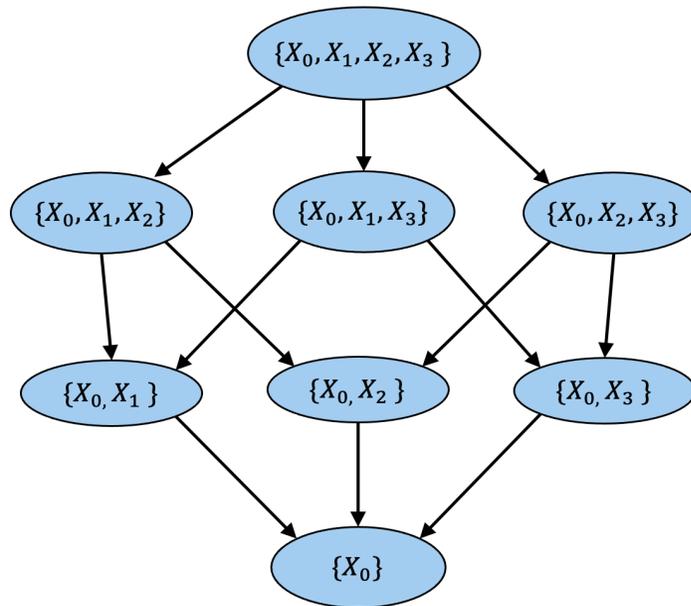
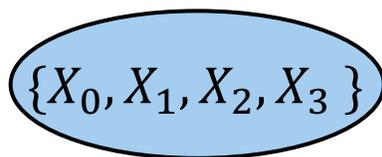


図 四変数に対するNPCリバーソオーダーグラフ

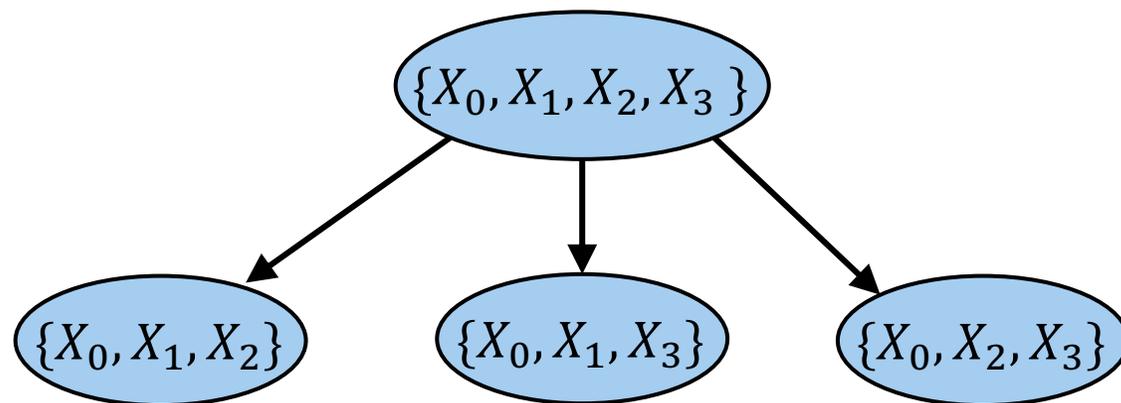
## 4.7 幅優先探索

NPCリバーソードグラフを幅優先探索する



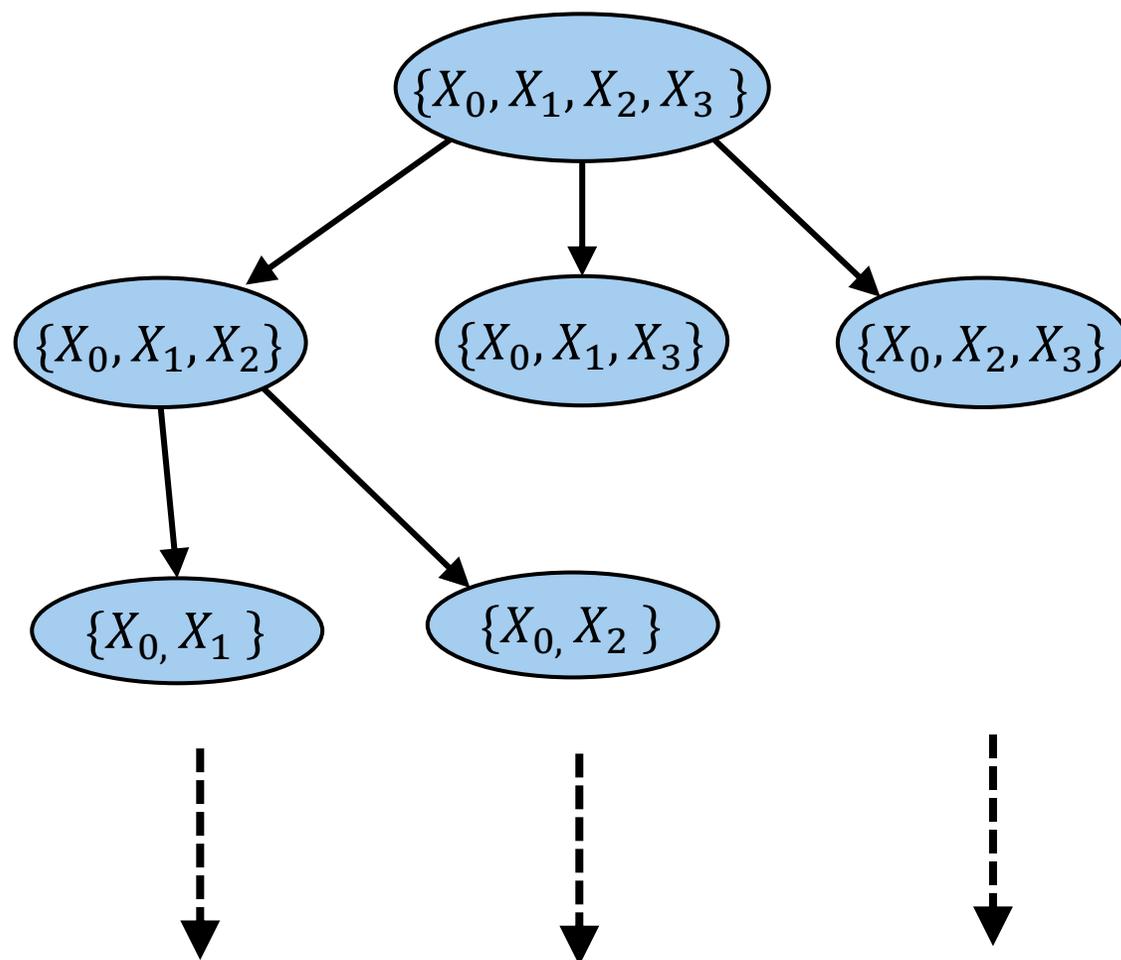
## 4.7 幅優先探索

NPCリバースオーダグラフを幅優先探索する



## 4.7 幅優先探索

NPCリバースオーダグラフを幅優先探索する



## 4.8 幅優先探索の問題

- 変数数の増加に伴い指数的に計算時間が増加する  
20変数程度の構造学習が限界
- 探索の終了まで構造を得ることができない

## 4.9 深さ優先分枝限定法

幅優先探索の問題を解決する効率的な学習法：**深さ優先分枝限定法**

### 定理5.7

本手法はNCPを最小にして真の分類確率に漸近収束する構造を学習する

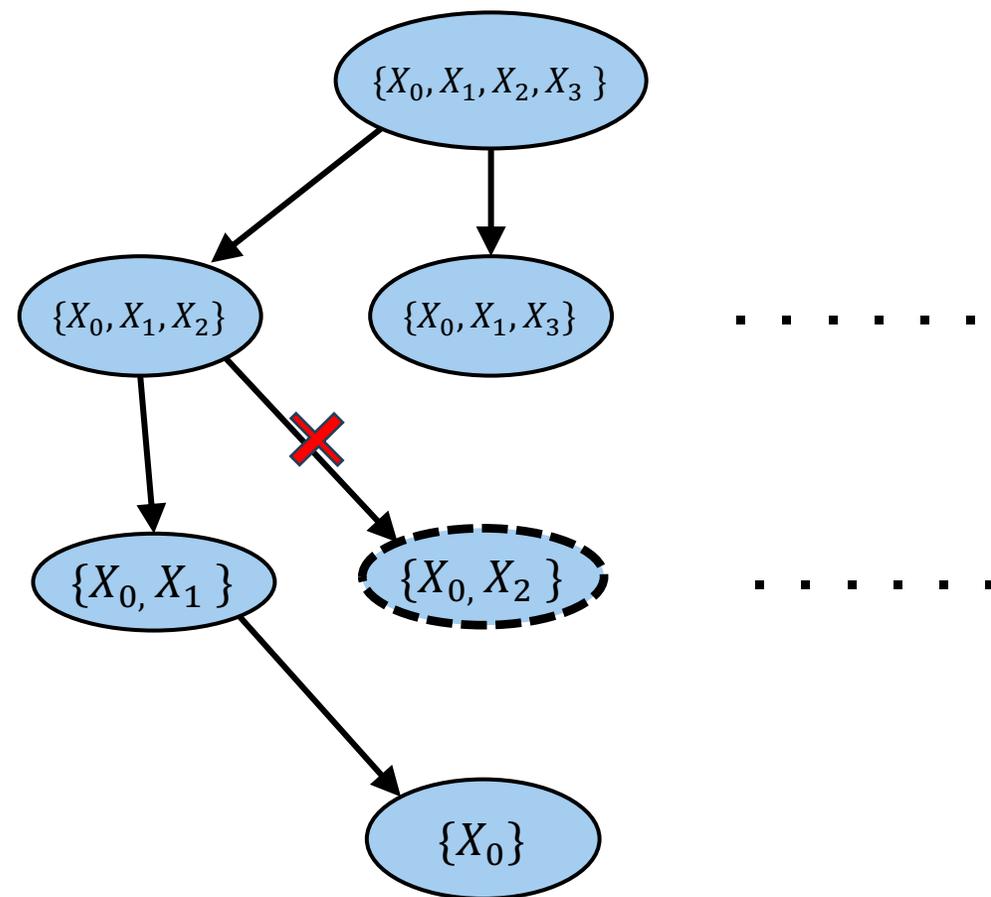


図 四変数に対するNPCリバーソオダグラフ

## 4.10 NPCリバーソオーダーグラフと枝刈り

深さ優先分枝限定法の計算時間は**枝刈り**によって向上する

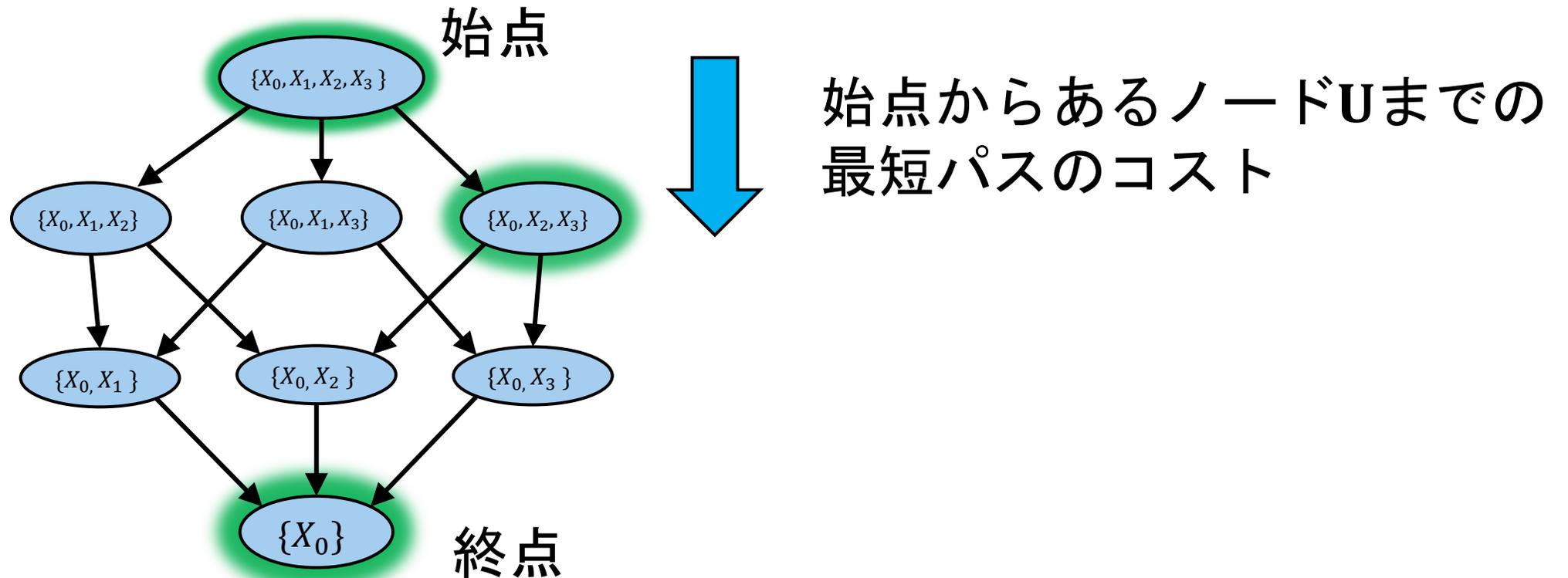


図 NPCリバーソオーダーグラフ

## 4.10 NPCリバーソオーダーグラフと枝刈り

深さ優先分枝限定法の計算時間は**枝刈り**によって向上する

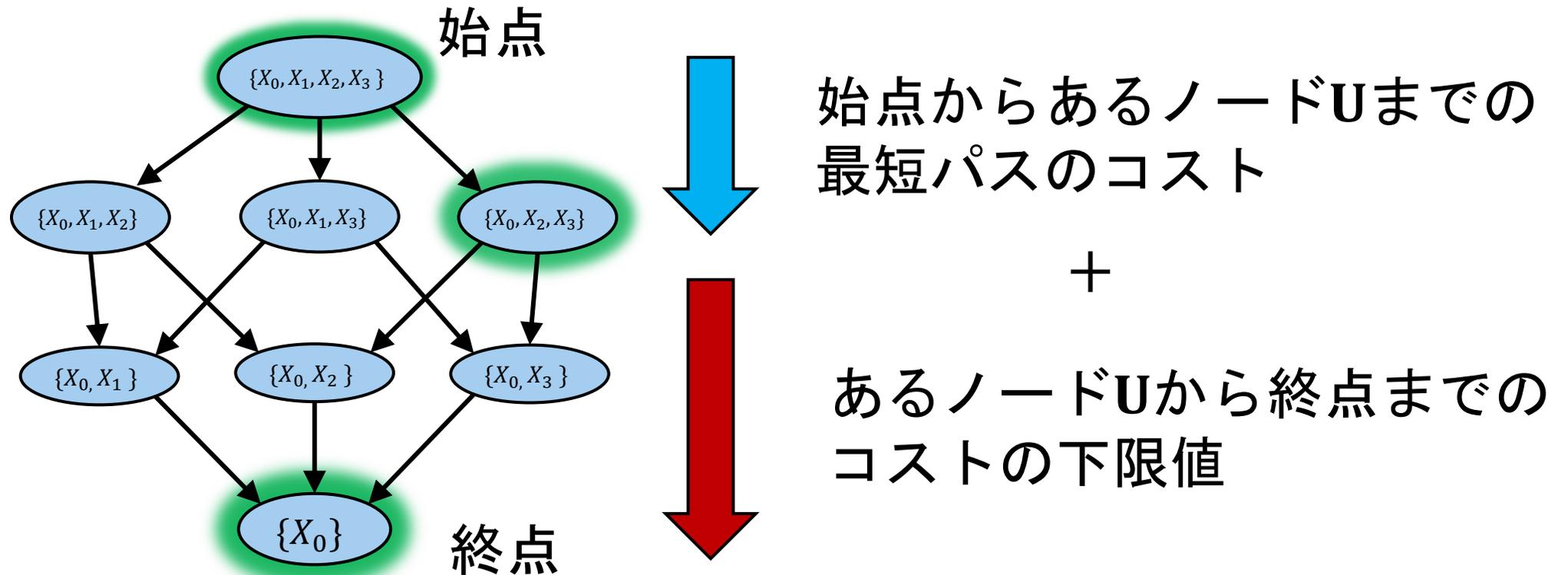
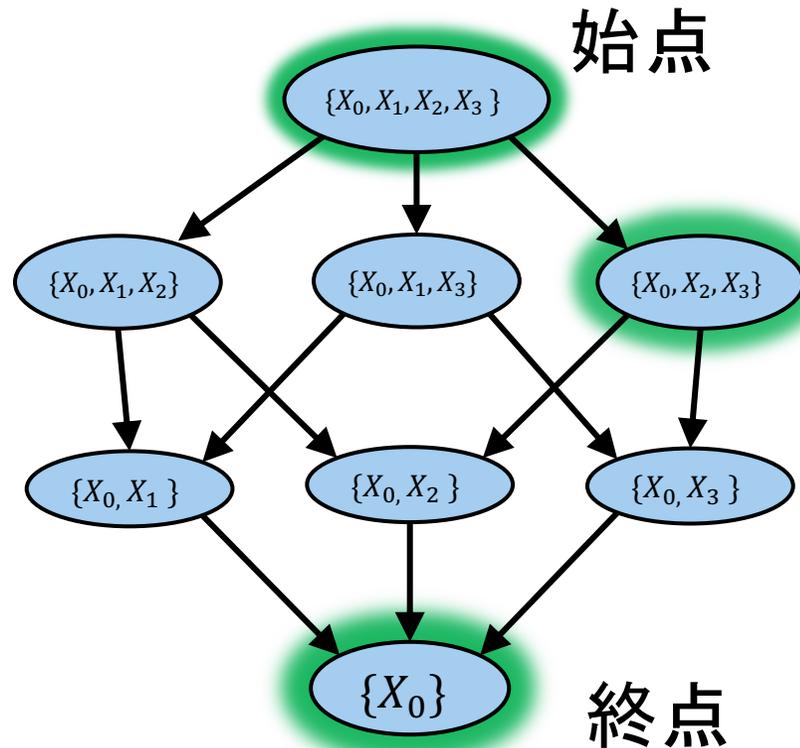


図 NPCリバーソオーダーグラフ

## 4.10 NPCリバーソオーダーグラフと枝刈り

深さ優先分枝限定法の計算時間は**枝刈り**によって向上する



始点からあるノードUまでの  
最短パスのコスト

+

あるノードUから終点までの  
コストの下限值

$\geq$  これまでに得られている  
最短パスのコスト

図 NPCリバーソオーダーグラフ

## 4.10 NPCリバーソオーダーグラフと枝刈り

深さ優先分枝限定法の計算時間は**枝刈り**によって向上する

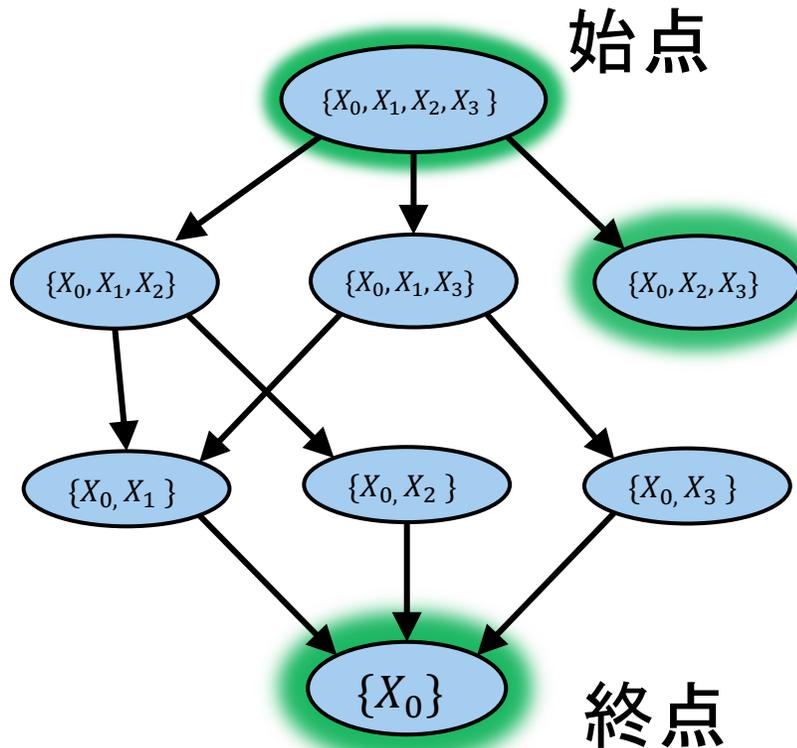


図 NPCリバーソオーダーグラフ

始点からあるノードUまでの  
最短パスのコスト

+

あるノードUから終点までの  
コストの下限值

≥ これまでに得られている  
最短パスのコスト

## 4.10 NPCリバーソオーダーグラフと枝刈り

深さ優先分枝限定法の計算時間は**枝刈り**によって向上する

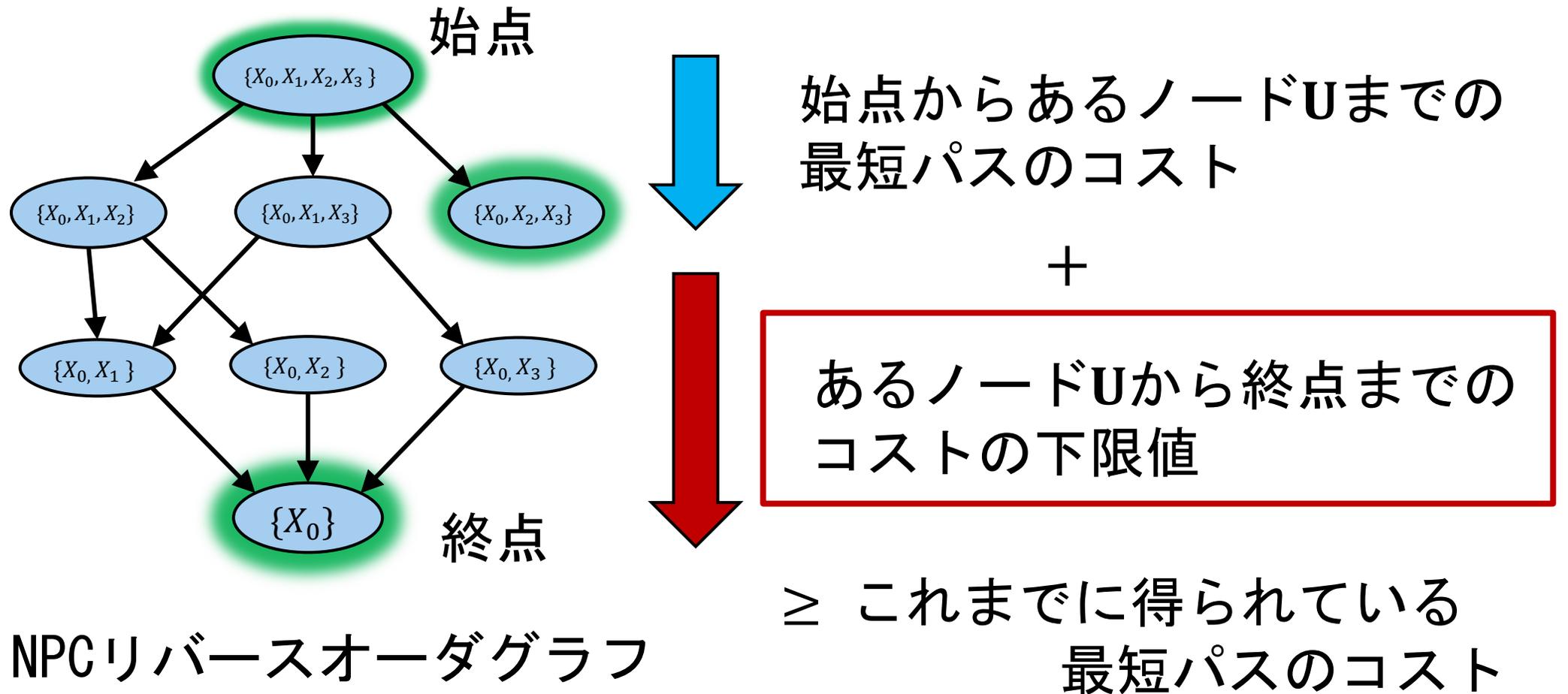


図 NPCリバーソオーダーグラフ

## 4.11 コストの下限值

定理4.11

Naive Bayesの目的変数パラメータ数

$\leq$  目的変数パラメータ数を最小にする構造の目的変数パラメータ数

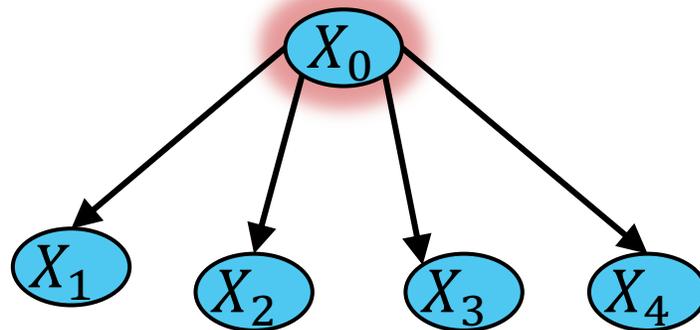


図 Naive Bayes

## 4.12 深さ優先分枝限定法の利点

1. 枝刈りにより幅優先探索に比べて**計算時間を削減**できる
2. 実行途中にメモリ等のリソースが不足しても**それまでの最適構造を得ることが可能**になる

## 4.13 評価実験

- 比較手法：
  1. Naive Bayes
  2. ANB厳密学習（変数選択適用）
  3. 幅優先探索によりNCP最小の構造を学習
  4. **深さ優先分枝限定法**：Naive Bayesによる下限値を用いた深さ優先分枝限定法  
パラメータはすべて期待事後推定量 (expected a posterior: EAP) で推定した。
- 実データ：
  - UCIレポジトリデータベースに登録されているベンチマークデータセット
- 実験手順：
  - 各手法，各データセットに対して，10分割交差検証によるテストデータの平均一致率を求め，分類精度とし，計算時間を測定した

# 4.14 分類精度

データセット	変数数	データ数	Naive Bayes	ANB	深さ優先 幅優先探索		分枝限定法								
					幅優先探索	深さ優先	BreastCancer	Heart	HTRU2	CVR	SolarFlare	Glass	CMC	Hayes-Roth	BalanceScale
Image Seg	19	2310	0.9286	0.9468	0.9550	0.9558	10	277	0.7401	0.7040	0.7401	0.7401			
Pendigits	17	10992	0.8805	0.9636	0.9601	0.9609	14	270	0.8444	0.8407	0.8074	0.8074			
Letter	17	20000	0.7384	0.8454	0.8608	0.8616	9	17898	0.9689	0.9779	0.9783	0.9784			
Lympho	19	148	0.8446	0.7770	0.8041	0.7905	17	232	0.9095	0.9483	0.9655	0.9698			
EEG	15	14980	0.6874	0.7644	0.7304	0.7285	11	1389	0.7811	0.8229	0.8431	0.8431			
WCW	10	683	0.9751	0.9751	0.9751	0.9751	10	214	0.5561	0.6449	0.6262	0.6075			
Zoo	17	101	0.9406	0.9505	0.9505	0.9307	10	1473	0.4671	0.4481	0.4623	0.4467			
Hepatitis	20	80	0.8500	0.5750	0.7875	0.8000	5	132	0.8182	0.7879	0.8333	0.8333			
Wine	14	178	0.9831	0.9663	0.9775	0.9775	5	625	0.9152	0.9152	0.9152	0.9152			
Australian	15	690	0.8464	0.8420	0.8551	0.8507	5	24	0.7500	0.7500	0.8750	0.8750			
Vehicle	19	846	0.4350	0.6253	0.6019	0.5827	5	150	0.9400	0.9400	0.9467	0.9467			
							8	3200	0.7294	0.7294	0.7316	0.7325			
							5	1372	0.9249	0.9410	0.9410	0.9410			
									0.8106	0.8201	0.8385	0.8354			

# 4.15 計算時間と枝刈り回数

No.	Dataset	Variables	深さ優先		枝刈り回数							
			幅優先探索	分枝限定法								
1	Lenses	5	0.0131	<b>0.0100</b>	3.9	12	Heart	14	9.9224	<b>3.8845</b>	6275.3	
2	Hayes-Roth	5	0.0170	<b>0.0149</b>	3.3	13	Australian	15	18.3376	<b>8.1574</b>	22030.4	
3	Iris	5	0.0181	<b>0.0134</b>	3.2	14	EEG	15	166.2700	<b>24.3754</b>	42485.8	
4	Balance Scale	5	<b>0.0134</b>	0.0170	4.0	15	Zoo	17	459.3530	<b>21.2139</b>	29600.5	
5	Banknote	5	<b>0.0188</b>	0.0268	4.0	16	Congressional	17	427.0858	<b>34.8079</b>	42485.8	
6	LED7	8	<b>0.1214</b>	0.1371	27.6	17	Pendigits	17	744.8170	<b>145.3891</b>	29600.5	
7	HTRU2	9	<b>0.2785</b>	0.3636	58.7	18	Letter	17	530.7353	<b>99.6420</b>	22001.9	
8	BC	10	0.3236	<b>0.1719</b>	176.1	19	Lymphography	19	555.6909	<b>97.3051</b>	189638.4	
9	BCW	10	0.3175	<b>0.1221</b>	98.0	20	Image Segmentation	19	5588.0012	<b>261.5876</b>	154339.2	
10	Solar Flare	11	0.8851	<b>0.4564</b>	326.2	21	Hepatitis	20	10044.8238	<b>250.7541</b>	386621.3	
11	Wine	14	16.0481	<b>4.6310</b>	5945.4							
									average	883.9567	<b>45.3848</b>	44368.1

計算時間が短い方を赤字

## 4.16 大規模データセットを用いた評価実験

- 比較手法  
小規模での実験と同様の手法
- 実データ  
小規模での実験より大規模な31~116変数のベンチマークデータセット
- 実験手順
  - 各手法, 各データセットに対して, 10分割交差検証によるテストデータの平均一致率を求め, 分類精度とし, 計算時間を測定した
  - 構造学習は, 6時間の制限時間を設け, 超過する場合は打ち切った

## 4.17 分類精度

No.	Dataset	Variables	Naïve Bayes	ANB	幅優先探索	深さ優先 分枝限定法
1	wdbc	31	0.9139	TO	TO	0.9350
2	turkiye	33	0.3442	TO	TO	0.4897
3	ionosphere	35	0.7550	TO	TO	0.8832
4	kr-vs-kp	37	0.6640	TO	TO	0.9252
5	Flowmeters_D	44	0.8333	TO	TO	0.8833
6	Parkinson	48	0.7625	TO	TO	0.7708
7	PAMAP2	53	0.6864	TO	TO	0.8634
8	spam	58	0.8794	TO	TO	0.9331
9	molecular	61	0.9433	TO	TO	0.9464
10	Nuclear	75	0.9303	TO	TO	0.9914
11	MI	116	0.9154	TO	TO	0.9375
average			0.7843	-	-	0.8690

TO: 6時間以内に構造を得ることができなかったことを表す

## 4.18 さらに精度の向上

- 深さ優先分枝限定法で学習した構造についてCLLを最大化するようパラメータを推定することで、さらに精度の向上が期待できる。
- しかし、CLLを最大化するパラメータは解析的に解けないため、勾配法で数値的に求める必要があるが、ほとんどの場合局所解に陥ってしまう問題がある。
- 構造がコーダルグラフである場合、CLLが単峰性を有するため、パラメータの初期値によらず勾配法で大域的最適解が得られることが知られている (Roos et. al, 2005)。
- 学習構造に対しコーダルグラフとなるようエッジを追加し、CLL最大推定量でパラメータ推定することで、局所解の問題を解決できる。

## 4.19 CLL最大化推定のアルゴリズム

1. 分類に影響する目的変数パラメータ数 (NCP) を最小にして真の分類確率に漸近収束する構造を学習する.
2. 得られた構造をモラル化し, その後コーダル化する.
3. 2で得られたコーダルグラフ  $G$  をマルコフネットワークの構造として扱い, マルコフネットワークのCLLを最大化するようパラメータを推定する. 訓練データを  $D = \cup_{d=1}^N (x_0^d, x_1^d, \dots, x_n^d)$  とすると, パラメータ  $\phi_C(\mathbf{x})$  ( $X_0 \in C$ ) についてのCLLの偏導関数は以下である.

$$\frac{\partial}{\partial \phi_C(\mathbf{x})} CLL(\Theta) = N_{\mathbf{x}} - \sum_{d=1}^N P(\mathbf{x}^{X_0} | x_1^d, \dots, x_n^d),$$

ただし,  $N_{\mathbf{x}}$  は訓練データ中の  $\mathbf{x}$  の頻度であり,  $\mathbf{x}^{X_0}$  は  $\mathbf{x}$  における  $X_0$  の値である. この等式を用いた勾配法によってパラメータを推定する.

マルコフネットワークの構造がコーダルグラフの場合, CLLは単峰性を有する. ステップ2を行うことで, 勾配法によってCLL最大化推定量の大域解が得られる.

## 4.20 分類精度評価実験

- 比較手法:

1. 深さ優先分枝限定法 (EAP): 深さ優先分枝限定法で学習した構造のパラメータを EAP で推定した分類器
2. RF: Random forest
3. DL: Deep learning
4. 深さ優先分枝限定法 (CLL): 深さ優先分枝限定法で学習した構造についてコードグラフとなるようエッジを追加した後, CLL を最大化するようパラメータを推定した分類器

- 実データ:

- UCI レポジトリデータベースに登録されているベンチマークデータセット

- 実験手順:

- 各手法, 各データセットに対して, 10分割交差検証によるテストデータの平均一致率を求め, 分類精度とし, 計算時間を測定した

## 4.21 分類精度

データセット	変数数	サンプルサイズ	深さ優先 分枝限定法 (EAP)	RF	Deep	深さ優先 分枝限定法 (CLL)
Hayes-Roth	5	132	0.8333	0.8088	0.7725	0.8032
Balance Scale	5	625	0.9152	0.8287	0.9840	0.9856
Banknote authentication	5	1372	0.9410	0.9432	0.9403	0.9428
Hepatitis	20	80	0.8000	0.8625	0.8750	0.8375
Zoo	17	101	0.9307	0.9500	0.9300	0.9604
Pendigits	17	10992	0.9609	0.9914	0.9899	0.9743
平均			0.8968	0.8974	0.9153	<b>0.9173</b>

# まとめ

- 周辺尤度による厳密学習手法は, CLLによる近似学習手法と比べて, 必ずしも劣るわけではない.
- サンプルサイズが小さいときにBNCでは目的変数の親変数数が増え, 子変数が減ると分類精度が劣化するので, 目的変数が全説明変数を子変数として持つ制約をおくANB構造の周辺尤度スコアを用いた厳密学習法がCLLを含む従来手法の分類精度を有意に改善することができる.
- RAIANBアルゴリズムは1000変数を超えるデータを学習でき, 従来のRAIアルゴリズムよりも分類精度が高い.
- 深さ優先分枝限定法は真のモデルがBNに従っていない場合でも, NCP最小のI-mapを学習できる.
- 深さ優先分枝限定法の学習構造のパラメータをCLL最大化推定量で推定した場合, ランダムフォレストやディープラーニングと同等以上の分類精度を示す.