

「ベイジアンネットワーク概説」

- 3.7 シミュレーションによる各種法の比較
- 3.8 MDL(BIC)とベイジアンネットワーク
の予測分布との関係
- 3.9 実データによる解析例

新納浩幸

モデル選択手法の比較

AIC、MDL、クーパの手法を比較


モデル選択の手法



シミュレーションでデータを発生

K2 アルゴリズムでモデルを推定

この中でモデル選択の手法が必要
AIC, MDL, クーパの手法を各々使って比較



シミュレーションの詳細

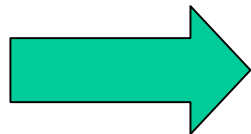
(1) 図3.3(p.61) の構造を適当に選ぶ

(2) 上記の構造からサンプルを k 個発生

K2 アルゴリズムで正しく構造を推定できれば○

(3) (1)(2) の実験を 1000 回行い、○の数をカウント

(4) k の値を増やして、(2)、(3)を繰り返す



結果、次ページの表

シミュレーション結果

サンプル数	AIC	MDL	クーパー
50	205	75	18
100	423	347	222
200	486	641	487
300	494	786	634
400	497	908	752
500	498	951	782
600	487	972	835
700	480	983	870
800	458	988	872
900	462	988	881
1000	424	989	897

MDL が最も良い

サンプル数が多くなっても一貫性を
持たない AIC は正解率が上がらない

MDL (BIC) と予測分布

定理3.2

$$\log p(X | B_S) \leq \log p(X, \Theta | B_S) - \frac{1}{2} k_m \log N$$

予測分布

最大にする B_S
を求める

(クーパー)

事後分布

最大にする Θ を求める (MAP推定)

$$BIC_m = -2 \cdot \left[l_m(\theta_m | X) - \frac{1}{2} k_m \log N \right]$$

MDL は右辺を最小にしようとする B_S

予測分布の強一致性

$$\log p(X, \Theta | B_S) - \frac{1}{2} k_m \log \frac{N'+N}{2\pi} + \text{定数} + O\left(\frac{1}{n}\right)$$

p.65,66 の
計算

→ $\leq \log p(X | B_S)$ ← (不等号式 1)

定理3.2

→ $\leq \log p(X, \Theta | B_S) - \frac{1}{2} k_m \log N$

定理3.2 から予測分布は強一致性をもつ

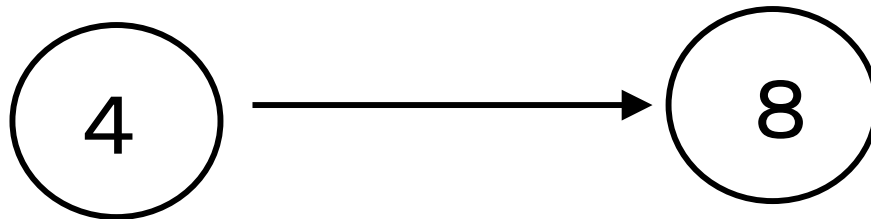
(不等号式 1)より収束は MDL よりも緩慢

解析例(1/2)

中学1年生256名、1次方程式に関するテスト15問

各問題を BN のノード、その問題を解答できるかどうか

例)



問題4を解けるならば問題8を解ける、、、関連有り

解析例(2/2)

