第一回
ベイジアンネットワーク

1.4.3　データによる学習
1.4.4　モデルの評価
1.4.5　モデル全体の評価
補足　マルコフ性の定義

発表者：鈴木朋央　発表日：06/10/20

データによる学習

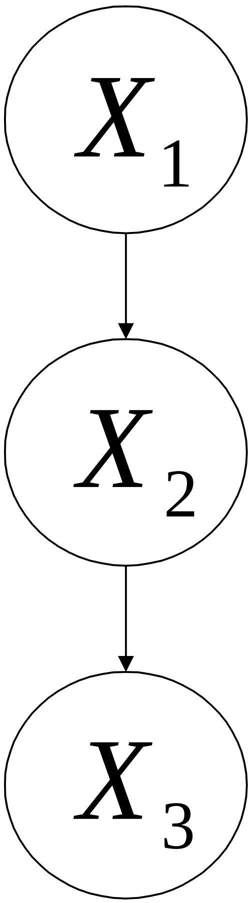
ノードは不確実な事象を示す確率変数である

不確実な事象が観測されたとき
→　確率変数が単なる数値に特定されるとき

↓

他の不確実な事象に関する評価も変化する

ケース1



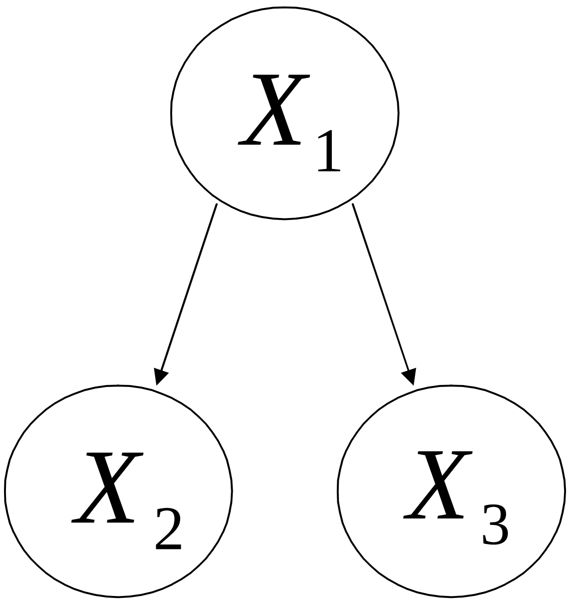
X2が観測された場合のX1,X3の確率

P(X3 | X2)の場合、X2を観測された特定の値x2に固定した分布を取り出せばよい

X1の確率は逆方向の推論であるため、ベイズの定理（逆確率の定理）を用いて評価する

$$P(X_1 \mid X_2) = \frac{P(X_2 \mid X_1)P(X_1)}{P(X_2)}$$

ケース2～手法1

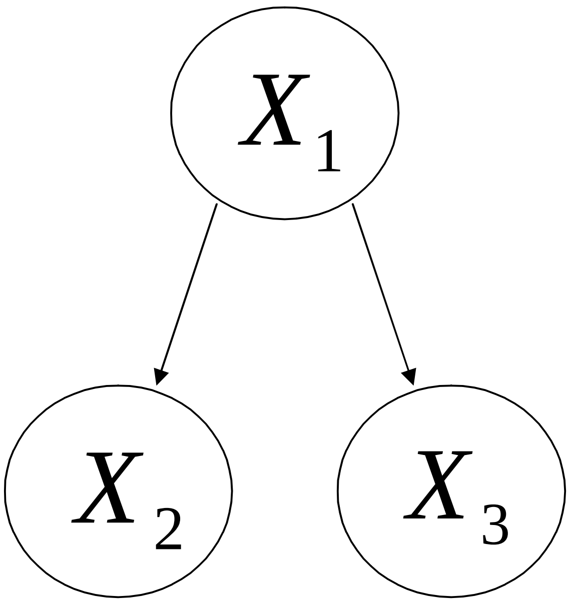


X3が観測された場合のP(X2 | X3)

1. $P(X_1, X_2, X_3) = P(X_2 \mid X_1) P(X_3 \mid X_1) P(X_1)$
2. $\sum_{X_1} P(X_1, X_2, X_3)$
3. $P(X_3) = \sum_{X_2} P(X_2, X_3)$
4. $P(X_2 \mid X_3) = \dfrac{P(X_2, X_3)}{P(X_3)}$

$\sum_{X_j}$ は確率変数Xjのとる全ての値について加算することを示す

ケース2～手法2



X3が観測された場合のP(X2 | X3)

1. $$P(X_1 \mid X_3) = \frac{P(X_3 \mid X_1)P(X_1)}{P(X_3)}$$

2. $$P(X_2 \mid X_3) = \sum_{X_1} P(X_2 \mid X_1)P(X_1 \mid X_3)$$

x3のみに依存するため、計算は後者のほうが速い

ケース3

X3が特定の値x3であると観測された場合、

→　その他の確率変数の事後分布は

$$P(X_1, X_2, X_4, X_5, X_6 \mid x_3) = \frac{P(X_1, X_2, x_3, X_4, X_5, X_6)}{P(X_3 = x_3)}$$

で与えられる

例：X3のデータを得たときのX2の確率P(X2 | x3)

$$P(x_2 \mid x_3) = \sum_{x_1}\sum_{x_4}\sum_{x_5}\sum_{x_6} P(X_1, X_2, X_4, X_5, X_6 \mid X_3 = x_3)$$

で表される

モデルの評価

ベイジアンネットワークモデルは、m個のノードの確率変数X1,X2,,,,,Xmの同時確率をモデル化する

このモデルが得られたデータ x=(x1,x2,,,,xm)をどの程度予測するか

↓

予測とデータとの距離を定義する方法として・・・

→　スコア関数 $S = -\log P(x)$

これが小さいほど、優れたモデルであるとされる

モデルの評価

XについてN個のデータ $x^{(1)}, x^{(2)}, \ldots, x^{(N)}$ が継時的に得られたとする。

＊ネットワークが学習機能を持たない場合

→　データによるモデルの変化は無く、N個のデータ全体の評価は

$$S = \sum_i \left\{ -\log P(x^{(i)}) \right\}$$

＊学習機能を持ち、1回ごとにモデル分布が変化する場合

→　第i番目のデータを取る次点で更新された確率分布をPi(x)とするとき

$$S = \sum_i \{ -\log P_i(x^{(i)} \mid x^{(1)}, x^{(2)}, \ldots, x^{(i-1)}) \}$$

モデルの評価

スコア関数Sはモデルの良し悪しの指標ではあるが、数値自体の絶対的意味は無い。

しかし、事前の各ノードへの確立の評価が、他の基準と比べてどの程度よいかを比較することが出来る

基準となる確率評価による分布を P_{ref} とするとき

$$
\begin{aligned}
d &= S - S_{ref} \\
&= S + \sum_i \log P_{ref}(x^{(i)} \mid x^{(1)}, x^{(2)}, \ldots, x^{(i-1)})
\end{aligned}
$$

モデル全体の評価

l番目のモデルに含まれるパラメータや潜在ノードの確率変数を積分や加算によって消去した観測変数のみを予測する分布（予測分布）をPl(x)とする。

各モデルl（l=1,..,2）を真とする事前確率が等しいとき、モデルlとl'は、次に定義される比（λ）によって比較することができる

$$\lambda = \frac{P_l(x)}{P_{l'}(x)}$$

ベイズ比：モデル選択基準

（値が大きいほど支持度が強い）

1より大きければ l' よりも l を支持している