第8章　グラフィカルモデル

本章では確率分布の図式的な表現である確率的グラフィカルモデルを扱う。これを扱うことのメリットは以下の通りである。

1. 確率モデルの構造を視覚化でき、新しいモデルの設計方針を決めるのに役立つ。
2. グラフの構造を調べることで、条件付き独立性などの性質がわかる。
3. 複雑なモデルにおける学習や推論をグラフ上の操作として簡潔に表現できる。

確率的グラフィカルモデルでは、確率変数をノードで表現し、変数間の確率的関係をリンクで表現する。そしてグラフ自体は「全確率変数上の同時分布が、一部の変数のみに依存する因子の積としてどのように分解可能か」という情報を表現する。（ここについては後で確認する。）

代表的なグラフィカルモデルの一つはベイジアンネットワーク（＝有効グラフィカルモデル）で、グラフのリンクが向きを持つ有向グラフで表現される。もう一つの代表的なモデルはマルコフ確率場（＝無向グラフィカルモデル）で、こちらは無向グラフで表現される。

また、推論問題を解く際にはこれらを因子グラフと呼ばれる別の表現に変換することもある。本章ではこれらグラフィカルモデルの特徴について学んでいく。

8.1 ベイジアンネットワーク

本節ではまず、変数間の確率的関係からベイジアンネットワークを構築する方法を学び、次にベイジアンネットワークから確率的関係を書き下す方法を学ぶ。また、これまで扱ってきたモデルがベイジアンネットワークでどのように表現できるかを確かめる。

ベイジアンネットワークは同時分布を条件付き確率に分解した次のような数式を表現するために使われる。

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

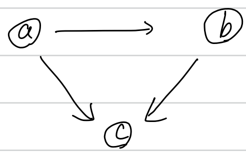
ノードは各変数に対応させ、リンクは条件つけられた変数から確率分布の変数に向かう方向に対応させる。

$$p(a,b,c) = p(c \mid a,b)\, p(b \mid a)\, p(a)$$

$a \to c$
$b \to c$

$a \to b$

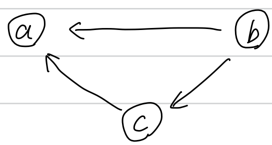
ベイジアンネットワーク：



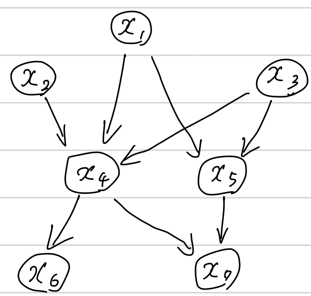
このような分解は任意の同時確率分布において可能で、さらに分解の仕方も複数存在しうる。例えば、

$$p(a,b,c) = p(a \mid b,c)\, p(c \mid b)\, p(b)$$

と分解するとベイジアンネットワークは次のように表現される。



これでは意味がないではないか、と思われるかもしれないが、実はグラフはリンクが**存在しないことをもって**分布の性質を表現する。これを確かめるために、今度は右のベイジアンネットワークから同時確率分布の分解の表式を得る。



$$p(x_1)p(x_2)p(x_3)p(x_4|x_1,x_2,x_3)p(x_5|x_1,x_3)p(x_6|x_4)p(x_7|x_4,x_5)$$

この式からわかるように、同時確率分布 $p(x)=p(x_1,\cdots,x_K)$ は一般的に $p(x)=\prod_{k=1}^{K}p(x_k|p(x_k))$ のように分解でき、同時分布の分解特性を表現している。

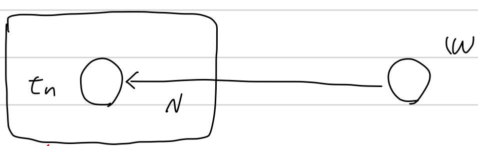
8.1.1 例：多項式曲線フィッティング

本項では、1.2.6 項で紹介したベイズ多項式回帰モデルをベイジアンネットワークでどのように表現されるかを確認する。

入力データを x 、ラベルを t 、係数ベクトルを w とし、ノイズの分散 σ^2 と係数ベクトルのハイパーパラメータ α を導入する。t と w の同時分布は次のように事前分布と尤度の積で現されていた。

$$p(t,w)=p(w)\prod_{n=1}^{N}p(t_n|w)$$

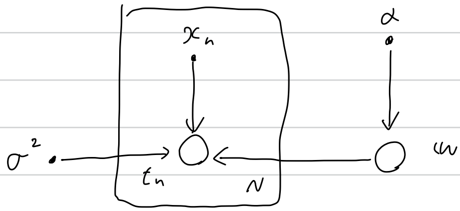
これは次のようなグラフィカルモデルで表現できる。



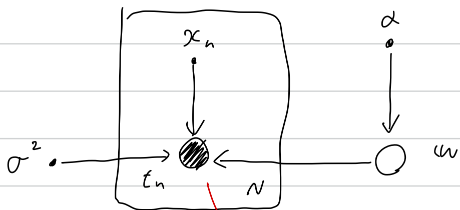
プレートと呼ばれ、Nコのノードを1つで表している。

また、決定的な変数も含めた下記の式展開に対応したしたグラフィカルモデルも存在する。

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



確率変数のうち、実際に値が観測される変数を観測変数と呼び、観測されない変数を潜在変数と呼ぶ。これらの違いはグラフィカルモデルではノードに影をつけることで判別する。



今回 t_n は観測されるが、w は観測されない。

また、新たに与えられたデータに対する予測値を表現するためには、新たなラベルを潜在変数として次のようなモデルで表現できる。