グラフィカルモデル（短縮版）

櫻井彰人

一言で

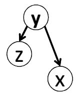
- グラフィカルモデル：
  - 条件付き独立性の有無を図示したもの

確率変数の独立性

- 確率変数 x と y が独立とは、次式が成立すること
$$p(x, y) = p(x)p(y)$$
- 条件付き確率の定義から
$$p(x, y) = p(x|y)p(y)$$
- 従って、x と y が独立 if and only if
$$p(x) = p(x|y)$$

条件付き独立性

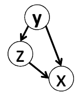
- 確率変数 x と z が y を条件として条件付き独立であるとは、



$$p(x, z|y) = p(x|y)p(z|y)$$

しばしば、 $x \perp\!\!\!\perp z|y$ と書く

- 条件付き確率の定義から



$$p(x, z|y) = p(x|z, y)p(z|y)$$
- 従って、 $x \perp\!\!\!\perp z|y$ if and only if
$$p(x|y) = p(x|z, y)$$

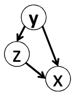
グラフィカルモデル（後程）

因子への分解（積で表示）

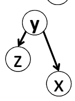
- 条件付き確率の定義から

グラフィカルモデル（後程）

$$p(x, y, z) = p(x|y, z)p(z|y)p(y)$$



- もし、 $x \perp\!\!\!\perp z|y$ であれば



$$p(x, y, z) = p(x|y)p(z|y)p(y)$$

グラフィカルモデル

- 構成要素
  - 有向無閉路グラフ（DAG or directed acyclic graph）
  - 節: 確率変数
  - 枝: 条件付き独立ではないことを表す
  - 枝がない: 条件付き独立であることを表す
  - 方向: 節の親子関係を表す
    - 親節: （しばしば）説明変数を表す
    - 子節: （しばしば）被説明変数（応答変数）を表す

例

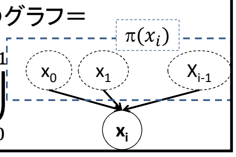
$p(x,y) = p(x|y)p(y)$　　$p(x,y) = p(x)p(y)$

- 節x の親ノード集合を $\pi(x)$で表す
- 結合確率とグラフは、次のように分解できる

$$p(x_0, x_1, \dots, x_{n-1}) = \prod_{i=0}^{n-1} p(x_i|\pi(x_i))$$

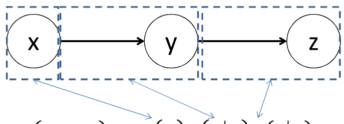
$\pi(x_i)$ は空集合も可

全結合のグラフ＝ $\bigcup_{i=0}^{n-1}$



例

- X = 曇り(Y/N)
- Y = 降雨あり(Y/N)
- Z = 地面が濡れている(Y/N)

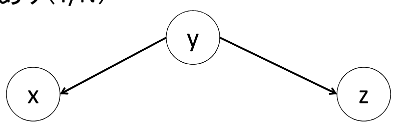


$p(x,y,z) = p(x)p(y|x)p(z|y)$

注: x と z は y を条件として条件付き独立である

例: 説明変数1個, 応答変数2個

- X = 痛みあり（Y/N）
- Y = インフルエンザに感染（Y/N）
- Z = 熱あり（Y/N）



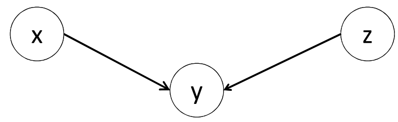
$p(x,y,z) = p(y)p(x|y)p(z|y)$

$p(x,y,z) = p(y)p(x,z|y)$

注: x と z は y を条件として条件付き独立である

例: 説明変数2個, 応答変数1個

- X = 地震発生（Y/N）
- Y = （自宅の）Alarmがなる（Y/N）
- Z = （自宅に）空き巣が入る（Y/N）

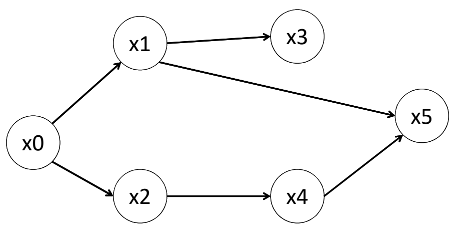


$p(x,y,z) = p(x)p(z)p(y|x,z)$ $= p(x)p(z)\frac{p(x,y,z)}{p(x,z)}$

から xとzとは独立

注: 一般的には、x と z は y を条件として条件付き独立ではない。何故か？

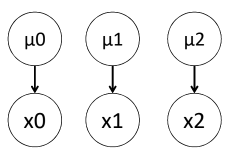
例: 積で記述

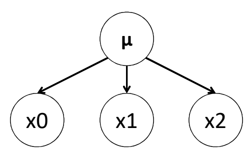


$$p(x_0, x_1, x_2, x_3, x_4, x_5) = p(x_0)p(x_1|x_0)\, p(x_2|x_0)p(x_3|x_1)\, p(x_4|x_2)p(x_5|x_1, x_4)$$

パラメータも書こう

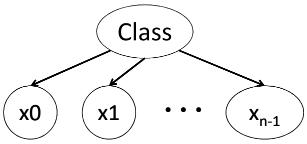
- モデルのパラメータも（ハイパーパラメータも）確率変数と同様に書いてしまおう
- 例えば、多変数ベルヌーイや多項分布





naïve Bayes は

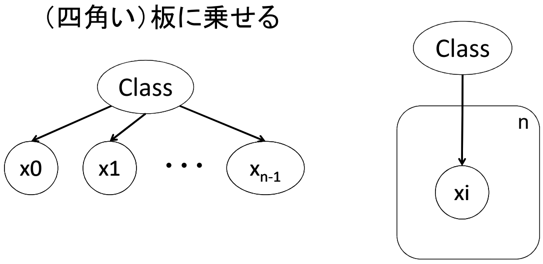
- 説明変数は、クラス変数を条件として、条件付き独立と仮定する



13

plate 記法

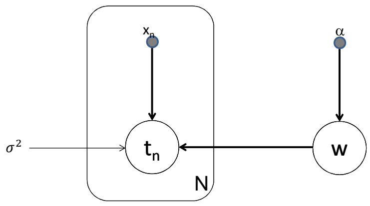
- 同じ依存関係にある変数が複数個あるとき、(四角い)板に乗せる



14

例

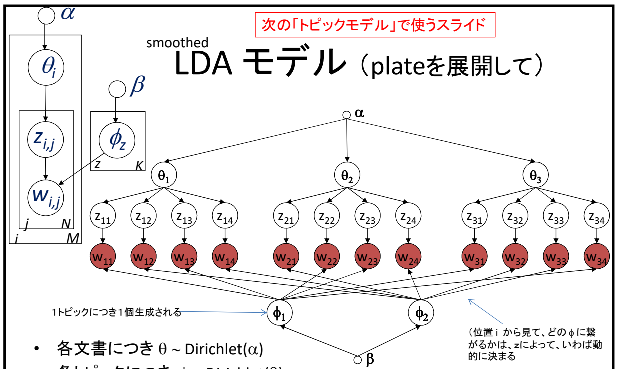
- パラメータは「点」にしてみた



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

次の「トピックモデル」で使うスライド

smoothed LDA モデル（plateを展開して）



- 各文書につき θ ~ Dirichlet(α)
- 各トピックにつき ϕ ~ Dirichlet(β)
- 全N単語のそれぞれ（のindex n where n=1,…,N）につき:
  - トピック z_n ~ Multinomial(θ)
  - 単語 w_n ~ $p(w_n|z_n,\phi)$, ただし、p は多項分布, i.e., $w_n \sim Multinomial(\phi_{z_n})$.

次の「トピックモデル」で使うスライド

Unigram と mixture of unigrams