鈴木譲「ベイジアンネットワーク入門」(培風館)
3. 統計的学習
3.4 条件付確率の推定 (1) 状態分割

鈴木譲

大阪大学

2009 年 12 月 17 日

あらまし

状態分割の推定

過学習と未学習

状態分割の推定

X, Y: 確率変数 ($|Y(\Omega)| < \infty$)
例 $\{(x_i, y_i)\}_{i=1}^n \in X^n(\Omega) \times Y^n(\Omega)$ から、条件付確率 $\mu_{Y|X}$ を推定

同値関係 $x \sim x'$

$x, x' \in X(\Omega)$

$$\mu_{Y|X}(\{y\}|x) = \mu_{Y|X}(\{y\}|x'), \quad y \in Y(\Omega)$$

1. $x \sim x$
2. $x \sim x' \Longrightarrow x' \sim x$
3. $x \sim x', x' \sim x'' \Longrightarrow x \sim x''$

$s(x)$: $x \in X(\Omega)$ の代表元
$\mathcal{S} := \{s(x) \,|\, x \in X(\Omega)\}$ (状態分割) は、確率変数 $S := s(X)$ の像

統計的推測

本質的な仮定

1. $\sigma := |\mathcal{S}| < \infty$
2. $z^n := \{z_i\}_{i=1}^n = \{(x_i, y_i)\}_{i=1}^n$ の生起が独立

乱数発生 $\mu_{Y|X}$ から、$z^n \in Z^n(\Omega)$ を生成

統計的推測 $z^n \in Z^n(\Omega)$ から、$\mu_{Y|X}$ を推定

$\mathcal{S}$ が既知の場合: $\theta_{y,s}$ を推定するだけ

$c_{y,s}$: $y \in Y(\Omega)$, $s \in \mathcal{S}$ の同時頻度

$$n_s := \sum_{y \in Y(\Omega)} c_{y,s}$$

$$\hat{\theta}_{y,s} = \frac{c_{y,s} + a_{y,s}}{n_s + \sum_{y'} a_{y',s}}, \quad a_{y,s} > 0$$

$\theta_{y,s}$ の事前確率が未知

$$a_{y,s} = 1/2$$

$\mathcal{S}$ が未知の場合: $\theta_{y,s}$ のみならず $\mathcal{S}$ を推定

$$Q^{(\mathcal{S})}(z^n) := \prod_{s\in\mathcal{S}} \frac{\Gamma\left(\sum_{y\in Y(\Omega)} a_{y,s}\right) \prod_{j=0}^{m-1} \Gamma(c_{j,s}+a_{j,s})}{\left[\prod_{y\in Y(\Omega)} \Gamma(a_{y,s})\right] \Gamma\left(n_s + \sum_{y'\in Y(\Omega)} a_{y',s}\right)}$$

$a_{y,s} = 1/2$ のとき、

$$Q^{(\mathcal{S})}(z^n) := \prod_{s\in\mathcal{S}} \frac{\Gamma(m/2) \prod_{j=0}^{m-1} \Gamma(c_{j,s}+1/2)}{(1/2)^m \Gamma(n_s + m/2)} \tag{28}$$

(26) は，$\sigma := |\mathcal{S}|$ として，

$$L^{(\mathcal{S})}(z^n) := \sum_{s\in\mathcal{S}} \mathcal{H}_s(z^n) + \frac{\sigma(m-1)}{2} \log n \tag{29}$$

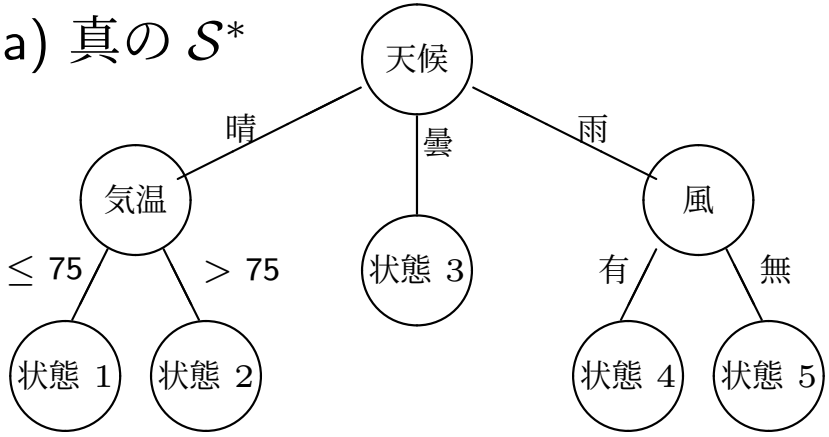
$$\mathcal{H}_s(z^n) := \sum_{y\in Y(\Omega)} c_{y,s} \log \frac{n_s}{c_{y,s}}$$

Quinlan のゴルフの例

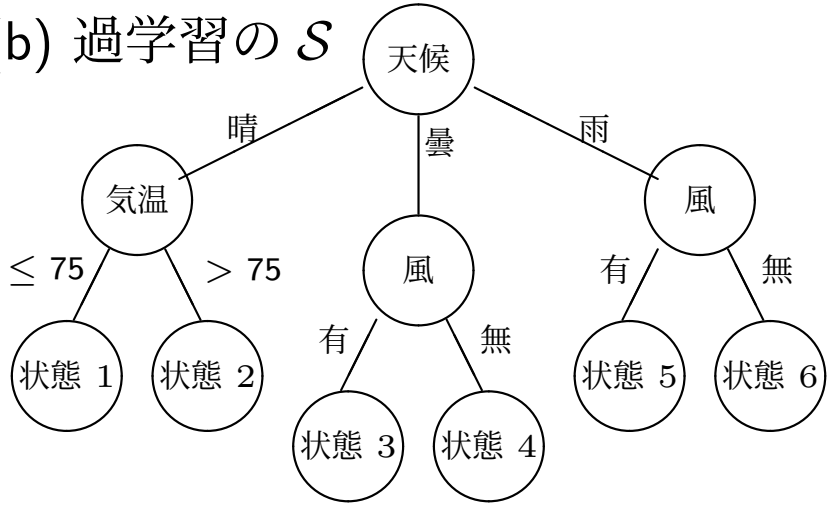
天候	気温 (華氏)	湿度	風	プレイ
晴	75	70	有	した
晴	80	90	有	しなかった
晴	85	85	無	しなかった
晴	72	95	無	しなかった
晴	69	70	無	した
曇	72	90	有	した
曇	83	78	無	した
曇	64	65	有	した
曇	81	75	無	した
雨	71	80	有	しなかった
雨	65	70	有	しなかった
雨	75	80	無	した
雨	68	80	無	した
雨	70	96	無	した

Quinlan のゴルフの例 (続)

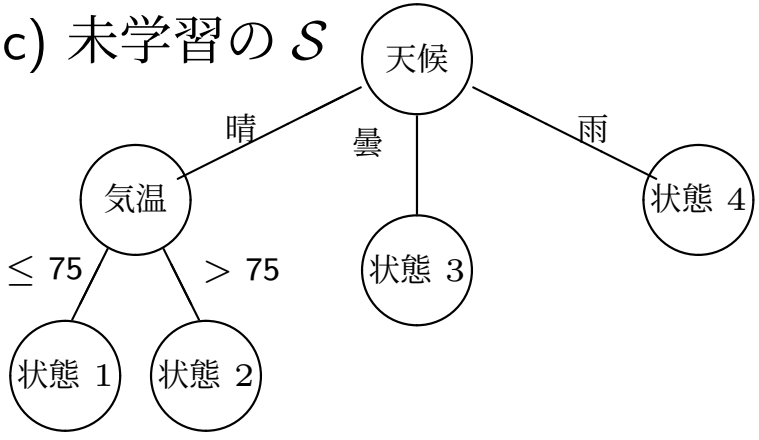
(a) 真の $\mathcal{S}^*$



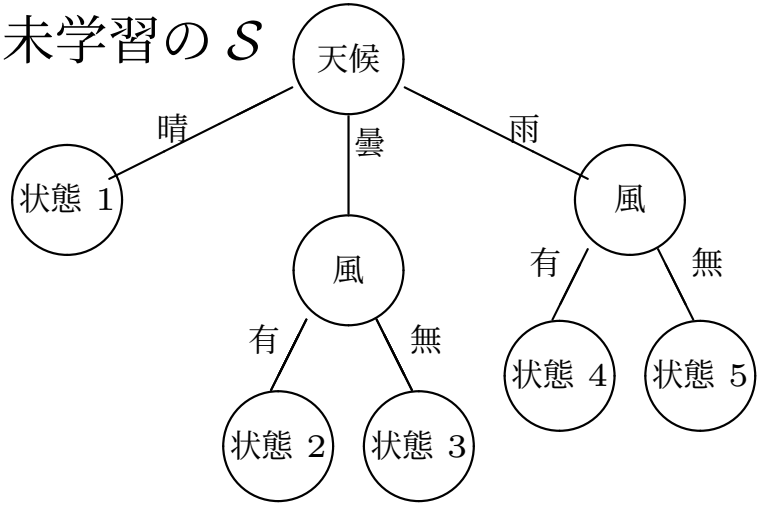
(b) 過学習の $\mathcal{S}$



(c) 未学習の $\mathcal{S}$



(d) 未学習の $\mathcal{S}$



Quinlan のゴルフの例 (続)

(a):

1. 天候 = 晴, 気温 ≤ 75
2. 天候 = 晴, 気温 > 75
3. 天候 = 曇
4. 天候 = 雨, 風 = 有
5. 天候 = 雨, 風 = 無

$\mathcal{S} = \{1,2,3,4,5\}$, $Y(\Omega) = \{0,1\}$ として, (28), (29) を計算:

1. $c_{0,1} = 1,\ c_{1,1} = 2,\ n_1 = 3,\ Q_1(z^n) = \frac{1/2}{1} \cdot \frac{1+1/2}{2} \cdot \frac{0+1/2}{3} = \frac{1}{2^4}$,
   $\mathcal{H}_1(z^n) = -\log(1/3) - 2\log(2/3)$
2. $c_{0,2} = 2,\ c_{1,2} = 0,\ n_2 = 2,\ \ Q_2(z^n) = \frac{1/2}{1} \cdot \frac{1+1/2}{2} = \frac{3}{2^3}$,
   $\mathcal{H}_2(z^n) = -2\log(2/2) - 0\log(0/2) = 0$
3. $c_{0,3} = 4,\ c_{1,3} = 0,\ n_3 = 4$,
   $Q_3(z^n) = \frac{1/2}{1} \cdot \frac{1+1/2}{2} \cdot \frac{2+1/2}{3} \cdot \frac{3+1/2}{4} = \frac{5\times 7}{2^7}$,
   $\mathcal{H}_3(z^n) = -4\log(4/4) - 0\log(0/4) = 0$

Quinlan のゴルフの例 (続)

4. $c_{0,4} = 2,\ c_{1,4} = 0,\ n_4 = 2,\ \ Q_4(z^n) = \frac{1/2}{1} \cdot \frac{1+1/2}{2} = \frac{3}{2^3},$
   $\mathcal{H}_4(z^n) = -2\log(2/2) - 0\log(0/2) = 0$
5. $c_{0,5} = 0,\ c_{1,5} = 3,\ n_5 = 3,$
   $Q_5(z^n) = \frac{1/2}{1} \cdot \frac{1+1/2}{2} \cdot \frac{2+1/2}{2} = \frac{3\times 5}{2^5},$
   $\mathcal{H}_5(z^n) = -0\log(0/3) - 3\log(3/3) = 0$

$$Q^{(\mathcal{S})}(z^n) = \prod_{s\in\mathcal{S}} Q_s(z^n) = \frac{1}{2^4} \cdot \frac{3}{2^3} \cdot \frac{5\times 7}{2^7} \cdot \frac{3}{2^3} \cdot \frac{3\times 5}{2^5} = \frac{3^3\times 5^2\times 7}{2^7},$$

$$\mathcal{H}^{(\mathcal{S})}(z^n) = \sum_{s\in\mathcal{S}} \mathcal{H}_s(z^n) = 3\log 3 - 2\log 2,$$

$$L^{(\mathcal{S})}(z^n) = 3\log 3 - 2\log 2 + \frac{5}{2}\log 14 = 3\log 3 + \frac{1}{2}\log 2 + \frac{5}{2}\log 7$$

Quinlan のゴルフの例 (続)

(b):

1. 天候 = 晴, 気温 ≤ 75
2. 天候 = 晴, 気温 > 75
3. 天候 = 曇, 風 = 有
4. 天候 = 曇, 風 = 無
5. 天候 = 雨, 風 = 有
6. 天候 = 雨, 風 = 無

(c):

1. 天候 = 晴, 気温 ≤ 75
2. 天候 = 晴, 気温 > 75
3. 天候 = 曇
4. 天候 = 雨

(d)

1. 天候 = 晴
2. 天候 = 曇, 風 = 有
3. 天候 = 曇, 風 = 無
4. 天候 = 雨, 風 = 有
5. 天候 = 雨, 風 = 無

過学習と未学習

$\mathcal{S}$ が真	$\mathcal{S}^*$
$\mathcal{S}$ が過学習	真ではなく、$\mathcal{S}$ の各要素が $\mathcal{S}^*$ の要素の部分集合
$\mathcal{S}$ が未学習	真でも過学習でも、過学習でもない

過学習の状態分割の性質

$\{\theta^*_{y,s}\}_{y\in Y(\Omega),s\in\mathcal{S}^*}$: 真の確率パラメータ
$p[s] := \mu(X \in s)$, $s \in \mathcal{S}^*$
$Over(\mathcal{S}^*)$: 過学習の状態分割の集合
すべての状態 $s \in \mathcal{S} \neq \mathcal{S}^*$ について $p[s] > 0$ のとき

$$\theta_{y,s} := \sum_{t\in\mathcal{S}^*} \frac{p[s\cap t]}{p[s]}\theta^*_{y,t}$$

$$D(\mathcal{S}^*||\mathcal{S}) := \sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{S}^*} p[s\cap t] \sum_{y\in Y(\Omega)} \theta^*_{y,t}\log\frac{\theta^*_{y,t}}{\theta_{y,s}}$$

$$\mathcal{S} \in \{\mathcal{S}^*\}\cup Over(\mathcal{S}^*) \iff \theta^*_{y,t} = \theta_{y,s},\ \mathcal{S}^* \ni t \supseteq s \in \mathcal{S},\ y \in Y(\Omega)$$
$$\implies D(\mathcal{S}^*||\mathcal{S}) = 0$$