鈴木讓「ベイジアンネットワーク入門」(培風館)
3. 統計的学習
3.2 データ圧縮 (1) 確率が既知の場合

鈴木讓

大阪大学

2009 年 11 月 19 日

あらまし

一意復号可能性

エントロピーと平均符号長

一意復号可能性

Z: 確率変数 ($|Z(\Omega)| < \infty$)
$\{0,1\}^*$: 有限の長さの 2 進列の集合

$\varphi(z) \in \{0,1\}^*$ が $z \in Z(\Omega)$ の符号語

$\varphi(z)$ が z の $\varphi : Z(\Omega) \longrightarrow \{0,1\}^*$ (符号化) についての像

$l_\varphi : Z(\Omega) \to \mathbb{N}$ が符号化 φ の長さ

$z \in Z(\Omega)$ に $\varphi(z) \in \{0,1\}^l$ なる $l \in \mathbb{N}$ (長さ) を対応させる写像

$\varphi^n : Z^n(\Omega) \to \{0,1\}^*$ が n 次の符号化

$z^n = (z_1, \cdots, z_n) \in Z^n(\Omega)$ に $(\varphi(z_1), \cdots, \varphi(z_n))$ を連結した長さ
$\sum_{i=1}^{n} l_\varphi(z_i)$ の 2 進列 $\varphi(z_1)\cdots\varphi(z_n) \in \{0,1\}^*$ を対応させる写像

一意復号可能性 (続)

φ が一意復号可能

各 $u^n, v^n \in Z^n(\Omega),\ n = 1, 2, \cdots$ で,

$$\varphi^n(u^n) = \varphi^n(v^n) \Longrightarrow u^n = v^n$$

$u, v \in Z(\Omega)$
$l_\varphi(u) \le l_\varphi(v)$

$\varphi(u)$ は $\varphi(v)$ の語頭, $\varphi(u) \prec \varphi(v)$

符号語 $\varphi(u)$ が符号語 $\varphi(v)$ の最初の長さ $l_\varphi(u)$ の 2 進列

φ が瞬時復号可能

$$\varphi(u) \prec \varphi(v) \Longrightarrow u = v$$

一意復号可能性 (続)

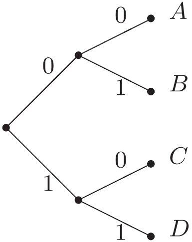
例: $Z(\Omega) = \{A, B, C, D\}$

- $\varphi_a, \varphi_b, \varphi_c$ は一意復号可能
- φ_a, φ_b は瞬時復号可能
- φ_c は瞬時復号可能ではない。
- $\varphi_d^6(A, C, A, D, B, A) = \varphi_d^6(D, B, A, C, A, A) = 011001100$

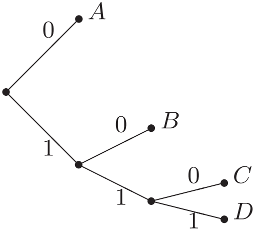
$z \in Z(\Omega)$	$\varphi_a(z)$	$\varphi_b(z)$	$\varphi_c(z)$	$\varphi_d(z)$
A	00	0	0	0
B	01	10	01	10
C	10	110	011	11
D	11	111	111	01
(A, C, A, D, B, A)	001000110100	01100111100	00110111010	011001100
(D, B, A, C, A, A)	110100100000	11110011000	11101001100	011001100

一意復号可能性 (続)

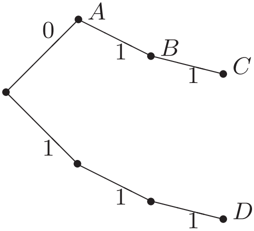
φ_a



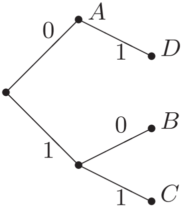
φ_b



φ_c



φ_d



瞬時復号可能 $\Longrightarrow$ 一意復号可能

命題 3.4

瞬時復号可能 $\Longrightarrow$ 一意復号可能

証明: φ が瞬時復号可能と仮定
符号化された後の系列 $\varphi(z_1)\cdots\varphi(z_n) \in \varphi^n(Z(\Omega))$ から

1. $\varphi(z) \prec \varphi(z_1)\cdots\varphi(z_n) \Longrightarrow z_1 = z$
   $\varphi(z_1)\varphi(z_2)\cdots\varphi(z_n)$ から $\varphi(z_1)$ を取り除く
2. $\varphi(z) \prec \varphi(z_2)\cdots\varphi(z_n) \Longrightarrow z_2 = z$
   $\varphi(z_2)\varphi(z_3)\cdots\varphi(z_n)$ から $\varphi(z_2)$ を取り除く
3. ......

符号化される前の系列 $z^n = (z_1\cdots z_n) \in Z^n(\Omega)$ が一意

一意復号可能であるための同値な条件

定理 3.3

以下の 3 つの条件は同値である。

1. l_φ が瞬時復号可能な符号化 φ の長さ
2. l_φ が一意復号可能な符号化 φ の長さ
3. l_φ が Kraft の不等式を満足：

$$\sum_{z \in Z(\Omega)} 2^{-l_\varphi(z)} \leq 1. \tag{1}$$

1. $\Longrightarrow$ 2. は，命題 3.4 から。

一意復号可能であるための同値な条件 (続)

2. $\Longrightarrow$ 3.: $l_{\max} := \max_{z \in Z(\Omega)} l(z)$

$\alpha^n(m)$: $\left\{ z^n \in Z^n(\Omega) \mid \sum_{i=1}^n l(z_i) = m \right\}$ の要素の数

$$\begin{aligned}\left(\sum_{z \in Z(\Omega)} 2^{-l(z)} \right)^n &= \sum_{z_1 \in Z(\Omega)} 2^{-l(z_1)} \cdots \sum_{z_n \in Z(\Omega)} 2^{-l(z_n)} \\ &= \sum_{z^n \in Z^n(\Omega)} 2^{-\sum_{i=1}^n l(z_i)} = \sum_{m=1}^{n l_{\max}} \alpha^n(m) 2^{-m}\end{aligned}$$

φ が一意復号可能 $\Longrightarrow$ 重複しない $\Longrightarrow$ $\alpha^n(m) \leq 2^m$ $\Longrightarrow$

$$\sum_{z \in Z(\Omega)} 2^{-l(z)} \leq (n l_{\max})^{1/n}$$

これが任意の n について成立

$$\sum_{z \in Z(\Omega)} 2^{-l(z)} \leq 1$$

一意復号可能であるための同値な条件 (続)

3. $\Longrightarrow$ 1.:

$$
\begin{aligned}
&\sum_{m=1}^{l_{\max}} \alpha^1(m) 2^{-m} \le 1 \\
\Longleftrightarrow \quad &\left\{
\begin{array}{lcl}
\alpha^1(1) & \le & 2, \\
\alpha^1(2) & \le & 2(2 - \alpha^1(1)), \\
\alpha^1(3) & \le & 2(2^2 - 2\alpha^1(1) - \alpha^1(2)), \\
\cdots & \cdots & \cdots \\
\alpha^1(l_{\max}) & \le & 2\{2^{l_{\max}-1} - 2^{l_{\max}-2}\alpha^1(1) \\
 & & - \cdots - \alpha^1(l_{\max} - 1)\}
\end{array}
\right.
\end{aligned}
$$

$\alpha^1(i) \le$ レベル i の葉に割り当て可能な葉の枚数

一意復号可能であるための同値な条件 (続)

φ が一意復号可能 $\not\Longrightarrow$ φ が瞬時復号可能

φ が一意復号可能のとき、φ が瞬時復号可能でなくとも,
$l_\varphi = l_{\varphi'}$ の瞬時復号可能な符号化 φ' が存在

例: φ_b, φ_c

エントロピーと平均符号長

Z: 確率変数 ($|Z(\Omega)| < \infty$)
$P_Z(z)$: 事象 $(Z = z)$ の確率

Z のエントロピー $H(Z)$ と φ についての平均符号長 $El_\varphi(Z)$

$$H(Z) := \sum_{z \in Z(\Omega)} -P_Z(z) \log_2 P_Z(z)$$

$$El_\varphi(Z) := \sum_{z \in Z(\Omega)} P_Z(z) l_\varphi(z)$$

命題 3.5

$$H(Z) \leq El_\varphi(Z) \leq H(Z) + 1$$

となる一意復号可能な符号化 φ が存在

エントロピーと平均符号長 (続)

証明: $Q(z) := 2^{-l_\varphi(z)}$

$$El_\varphi(z) = \sum_{z\in Z(\Omega)} -P_Z(z)\log_2 Q(z)$$

$\log x \le x-1,\ x>0$ と Kraft の不等式より，左の不等式が成立:

$$\begin{aligned}\sum_{z\in Z(\Omega)} P_Z(z)\log\frac{P_Z(z)}{Q(z)} &\ge \sum_{z\in Z(\Omega)} P_Z(z)\left\{1-\frac{Q(z)}{P_Z(z)}\right\}\\ &= 1-\sum_{z\in Z(\Omega)} Q(z) \ge 0\end{aligned}$$

$l_\varphi(z) := \lceil -\log P_Z(z)\rceil$ (切上げ) は Kraft の不等式を満足し，
$\lceil -\log P_Z(z)\rceil \le -\log P_Z(z)+1$ より，右の不等式が成立。

エントロピーと平均符号長 (続)

$Z^n := (Z_1, \cdots, Z_n)$: 確率変数の列
事象 $(Z_i = z_i)$, $(Z_j = z_j)$, $z_i, z_j \in Z(\Omega)$, $i \neq j$ が独立
$P_{Z^n}(z^n)$, $z^n \in Z^n(\Omega)$: 事象 $(Z^n = z^n)$ の確率

$$
\begin{aligned}
H(Z^n) &:= \sum_{z^n \in Z^n(\Omega)} -P_{Z^n}(z^n) \log P_{Z^n}(z^n) \\
&= \sum_{z^n \in Z^n(\Omega)} -P_{Z^n}(z^n) \sum_{i=1}^{n} \log P_Z(z_i) \\
&= \sum_{i=1}^{n} \sum_{z \in Z(\Omega)} -P_Z(z_i) \log P_Z(z_i) = nH(Z)
\end{aligned}
$$

エントロピーと平均符号長 (続)

$\varphi : Z^n(\Omega) \to \{0,1\}^*$ (n 次の符号化とは異なる)
$l_\varphi(z^n) := \lceil -\log P_{Z^n}(z^n) \rceil$
$Q(z^n) := 2^{-l_\varphi(z^n)},\ \ z^n \in Z^n(\Omega)$

$$nH(Z) \le El_\varphi(Z^n) \le nH(Z) + 1$$

平均圧縮率

平均の長さを n で割った値 $\dfrac{El_\varphi(Z^n)}{n}$

定理 3.4

$$\frac{El_\varphi(Z^n)}{n} \to H(Z)$$

$(n \to \infty)$ で一意復号可能な符号化 $\varphi : Z^n(\Omega) \to \{0,1\}^*$ が存在