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
3.3 MDL 基準

鈴木譲

大阪大学

2010 年 6 月 24 日

あらまし

記述長

一般的なエントロピー

Markov 過程

MDL 基準

記述長

$0 \log 0 := 0$ $(\lim_{p\to 0} p \log p = 0)$

$$\mathcal{H}(z^n) := \sum_{j=0}^{m-1} c_j \log \frac{n}{c_j}$$

Stirling の公式

$$\Gamma(x) = (2\pi)^{1/2} x^{x-1/2} \exp[-x + \gamma(x)], \quad 0 < \gamma(x) < \frac{1}{12x}$$

$$\begin{aligned} l_\varphi(z^n) &= -\log Q_W(z^n) = \log\{\frac{\Gamma(1/2)}{\Gamma(m/2)} \cdot \frac{\Gamma(n+m/2)}{\prod_{j=0}^{m-1} \Gamma(c_j + 1/2)}\} \\ &= \mathcal{H}(z^n) + \frac{m-1}{2} \log n + O(1) \end{aligned} \tag{21}$$

(証明: 106-107 ページ)

適応的符号化

$c_{z_i}(z^{i-1})$: z^{i-1} における z_i の頻度

$$Q_W(z^n) = \prod_{i=1}^{n} \frac{Q_W(z^i)}{Q_W(z^{i-1})} = \prod_{i=1}^{n} \frac{c_{z_i}(z^{i-1}) + 1/2}{(i-1) + m/2}$$

適応的符号化: 送信側と受信側で各時点までで同じ推定結果を共有し、逐次符号化の仕方を変える

切上げを除くと、各時点で $-\log \dfrac{c_{z_i}(z^{i-1}) + 1/2}{(i-1) + m/2}$ の長さで符号化

$m = 2$, $z^{10} = 0100011010$

$$\frac{0+1/2}{0+1} \cdot \frac{0+1/2}{1+1} \cdot \frac{1+1/2}{2+1} \cdot \frac{2+1/2}{3+1} \cdot \frac{3+1/2}{4+1} \cdot \frac{1+1/2}{5+1} \cdot \frac{2+1/2}{6+1} \cdot \frac{4+1/2}{7+1} \cdot \frac{3+1/2}{8+1} \cdot \frac{5+1/2}{9+1}$$

θ の推定

- 前半: θ_{z_i} の推定精度が良くない
- θ_{z_i} 既知の場合と比較しての平均符号長の増分: $\dfrac{m-1}{2}\log n$

$z^n \in Z^n(\Omega)$ から $j \in Z(\Omega)$ の確率を推定

$$\frac{c_j(z^n) + a_j}{n + \sum_{k=0}^{m-1} a_k}, \quad a_0 = \cdots = a_{m-1} = \frac{1}{2} \tag{22}$$

一般的なエントロピー

独立性を仮定しない:

$$P_{Z^n}(z^n) = \prod_{i=1}^{n} P_{Z_i|Z_1^{i-1}}(z_i|z_1^{i-1}), \quad z^n \in Z^n(\Omega)$$

定常性を仮定:

$$P_{Z_{i+1}|Z_{i-k+1}^{i}}(z_i|z_{i-k}^{i-1}) = P_{Z_i|Z_{i-k}^{i-1}}(z_i|z_{i-k}^{i-1}), \quad k \geq 0, \quad z_{i-k}^{i+1} \in Z_{i-k}^{i+1}(\Omega) \tag{23}$$

$\Longrightarrow$ エントロピー $H := \lim_{n\to\infty} \frac{H(Z^n)}{n}$ が存在

一般的なエントロピー (続)

証明:

$$H(Z_{i+1}|Z_2^i) = H(Z_i|Z_1^{i-1}).$$

$Z_2^i = z_2^i$ のもとでの Z_1, Z_{i+1} の相互情報量

$$H(Z_{i+1}|Z_2^i = z_2^i) - H(Z_{i+1}|Z_1, Z_2^i = z_2^i) \geq 0$$

の $Z_2^i = z_2^i \in Z_2^i(\Omega)$ についての平均値

$$H(Z_{i+1}|Z_2^i) - H(Z_{i+1}|Z_1^i) \geq 0$$

$$H(Z_i|Z_1^{i-1}) \geq H(Z_{i+1}|Z_1^i) \geq 0$$

$$a_n \to a \Longrightarrow \frac{1}{n}\sum_{i=1}^{n} a_i \to a$$

$$\frac{H(Z^n)}{n} = \frac{H(Z_1) + H(Z_2|Z_1) + \cdots + H(Z_n|Z_1^{n-1})}{n} \to H$$

Markov 過程

$\{Z_i\}_{i=1}^{\infty}$ が h 次 Markov 過程

$i \geq h+1$ で

$$P_{Z_i|Z_1^{i-1}}(z_i|z_1^{i-1}) = P_{Z_i|Z_{i-h}^{i-1}}(z_i|z_{i-h}^{i-1}) \tag{24}$$

有限の $h \geq 0$ が存在

$h = 0$ であれば，$\{Z_i\}_{i=1}^{\infty}$ は独立

$P_{Z_i|Z_{i-h}^{i-1}}(z_i|z_{i-h}^{i-1})$ の $z_{i-h}^{i-1} \in Z^h(\Omega)$ の分割

$s^{(1)}, \cdots, s^{(k)} \subseteq Z^h(\Omega)$:

$$s^{(i)} \cap s^{(j)} = \{\}, \quad i \neq j, \quad \bigcup_{i=1}^{k} s^{(i)} = Z^h(\Omega)$$

Markov 過程 (続)

$\mathcal{S} := \{s^{(1)}, \cdots, s^{(k)}\}$
$s : z^h \in Z^h(\Omega) \to s(z^h) \in \mathcal{S}$

$$P_{Z_i|Z^{i-1}}(z_i|z^{i-1}) = P_{Z_i|s(Z_{i-h}^{i-1})}(z_i|s(z_{i-h}^{i-1})), \quad z_{i-h}^{i} \in Z_{i-h}^{i}(\Omega) \quad (25)$$

- $s(z_{i-h}^{i-1})$ が同じであれば，$Z_i = z_i \in Z_i(\Omega)$ の確率が同じ
- 定常性より、
  $P_{S_i}(s_i)$, $s_i := s(z_{i-h}^{i-1}) \in \mathcal{S}$, $S_i := s(Z_{i-h}^{i-1})$: i によらない

エントロピー

$$H(Z_i|Z_{i-h}^{i-1}) = H(Z_i|S_i) = \sum_{s\in\mathcal{S}} P_{S_i}(s)H(Z_i|S_i = s): \ i \text{ によらない}$$

Markov 過程で θ が既知の場合

$z_{-h+1}^{0} = z_{-h+1} \cdots z_0 \in Z^h(\Omega)$: 未知 ($h$ は有限)

全体 z^n として、$O(1/n)$ の平均圧縮率増加

$c_{z,s}$: $s \in \mathcal{S}$ における $z \in Z(\Omega)$ の頻度
$\theta_{z,s}$: $s \in \mathcal{S}$ における $z \in Z(\Omega)$ の確率:

$$\theta_{z,s} \in \left\{ (\theta_{z,s})_{z \in Z(\Omega), s \in \mathcal{S}} \mid \theta_{z,s} \geq 0,\ z \in Z(\Omega),\ \sum_{z \in Z(\Omega)} \theta_{z,s} = 1,\ s \in \mathcal{S} \right\}$$

$$P_\theta(z^n) := \prod_{s \in \mathcal{S}} \prod_{z \in Z(\Omega)} \theta_{z,s}^{c_{z,s}}$$

$\theta_{z,s}$ が既知

平均圧縮率とエントロピーの差を $O(1/n)$ にする符号化が存在

Markov 過程で θ が未知の場合

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

$$n_s := \sum_{z\in Z(\Omega)} c_{z,s}$$

$$\begin{aligned} l^{(\mathcal{S})}(z^n) &= \sum_{s\in\mathcal{S}}\left[\mathcal{H}_s(z^n)+\frac{m-1}{2}\log n_s + O(1)\right] \\ &\le \sum_{s\in\mathcal{S}}\mathcal{H}_s(z^n)+\frac{|\mathcal{S}|(m-1)}{2}\log n + O(1) \end{aligned} \tag{26}$$

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

(22) に相当する確率の推定式:

$$\frac{c_{z,s}+1/2}{n_s+m/2} \tag{27}$$

$\mathcal{S}$ が未知の場合: 2 段階符号化

$\mathcal{S}$ の候補は有限個なので，各 $\mathcal{S}$ の事前確率 $\pi(\mathcal{S})$ を決められる

2 段階符号化

$\lceil -\log \pi(\mathcal{S}) \rceil$ の長さで $\mathcal{S}$ を符号化し，その $\mathcal{S}$ を送受信側が共有した後，長さ $l^{(\mathcal{S})}(z^n)$ で $z^n \in Z(\Omega)^n$ を送信

$$l(z^n) := \min_{\mathcal{S}} \{ \lceil -\log \pi(\mathcal{S}) \rceil + \lceil l^{(\mathcal{S})}(z^n) \rceil \}$$

S に重み付けして、適応的符号化で実現

$$Q(z^n) := \sum_{\mathcal{S}} \pi(\mathcal{S}) Q^{(\mathcal{S})}(z^n)$$

$$l(z^n) := \lceil -\log Q(z^n) \rceil$$

MDL 基準

後者の符号長が前者の符号長を上回ることはない:

$$
\begin{aligned}
-\log Q(z^n) &= -\log \sum_{\mathcal{S}} \pi(\mathcal{S}) Q^{(\mathcal{S})}(z^n) \\
&\leq -\log \max_{\mathcal{S}} \pi(\mathcal{S}) Q^{(\mathcal{S})}(z^n) \\
&\leq \min_{\mathcal{S}} [-\log \pi(\mathcal{S}) - \log Q^{(\mathcal{S})}(z^n)] \\
&\leq \min_{\mathcal{S}} \{\lceil -\log \pi(\mathcal{S}) \rceil + l^{(\mathcal{S})}(z^n)\}.
\end{aligned}
$$

メリットは無いが、$z^n \in Z^n(\Omega)$ から $\mathcal{S}$ を推定する手段として利用

$$
\frac{\pi(\mathcal{S}) Q^{(\mathcal{S})}(z^n)}{\sum_{\mathcal{S}'} \pi(\mathcal{S}') Q^{(\mathcal{S}')}(z^n)}
$$

(分子) を最大にする $\mathcal{S}$ を求めること (誤り確率最小) と等価