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
3.2 データ圧縮 (2) 確率が未知の場合

鈴木譲

大阪大学

2009 年 12 月 10 日

あらまし

問題の定式化

$\min_{W} \max_{\theta} D^n(P_\theta || Q_W)$ の上界

$\min_{W} \max_{\theta} D^n(P_\theta || Q_W)$ の下界

Jeffreys の事前確率

問題の定式化: $\min_W \max_\theta D^n(P_\theta||Q_W)$ とそれに到達する W

c_j: $z^n \in Z^n(\Omega)$ における $j \in Z(\Omega) := \{0, 1, \cdots, m-1\}$ の頻度

$$\theta \in \Theta(\Omega) := \left\{ (\theta_0 \cdots, \theta_{m-1}) \mid \theta_j \geq 0,\ j = 0, 1, \cdots, m{-}1,\ \sum_{j=0}^{m-1} \theta_j = 1 \right\}$$

$$P_\theta(z^n) := \prod_{j=0}^{m-1} \theta_j^{c_j} \tag{5}$$

W: Θ の測度 $(W(\Theta(\Omega)) = 1)$

$$Q_W(z^n) := \int_{\Theta(\Omega)} P_\theta(z^n)\, W(d\theta) \tag{6}$$

$l_\varphi(z^n) := -\log Q_W(z^n)$

$$D^n(P_\theta||Q_W) := El_\varphi(Z^n) - H(Z^n) = \sum_{z^n \in Z^n(\Omega)} P_\theta(z^n) \log \frac{P_\theta(z^n)}{Q_W(z^n)}$$

上界

$W(d\theta) = w(\theta)d\theta$、Direchlet 分布 $(a_0, a_1, \cdots, a_{m-1} = \frac{1}{2})$ を仮定

$$w_{a_0,\cdots,a_{m-1}}(\theta) = \frac{\Gamma\left(\sum_{j=0}^{m-1} a_j\right)}{\prod_{j=0}^{m-1}\Gamma(a_j)} \prod_{j=0}^{m-1} \theta_j^{a_j-1} \tag{7}$$

$$\begin{aligned}
Q_W(z^n) &= \int_{\Theta(\Omega)} w_{1/2,\cdots,1/2}(\theta) P_\theta(z^n)\, d\theta \\
&= \int_{\Theta(\Omega)} \frac{\Gamma(m/2)}{\Gamma(1/2)^m} \prod_{j=0}^{m-1} \theta_j^{c_j-1/2}\, d\theta \\
&= \frac{\Gamma(m/2)\prod_{j=0}^{m-1}\Gamma(c_j+1/2)}{\Gamma(1/2)^m\Gamma(n+m/2)} \int_{\Theta(\Omega)} w_{c_0+1/2,\cdots,c_{m-1}+1/2}(\theta)\, d\theta \\
&= \frac{\Gamma(m/2)\prod_{j=0}^{m-1}\Gamma(c_j+1/2)}{\Gamma(1/2)^m\Gamma(n+m/2)}. \qquad (8)
\end{aligned}$$

上界 (続)

$z_*^n := 0 \cdots 0$ (0 が n 回続く系列)

$$Q_W(z_*^n) = \frac{\Gamma(m/2)\Gamma(n+1/2)}{\Gamma(1/2)\Gamma(n+m/2)}, \tag{9}$$

$$\begin{aligned}
\frac{Q_W(z^n)}{Q_W(z_*^n)} &= \frac{\prod_{j=0}^{m-1}\Gamma(c_j+1/2)}{\Gamma(1/2)^{m-1}\Gamma(n+1/2)} && (10)\\
&= \frac{\prod_{j=0}^{m-1}(c_j+1)(c_j+2)\cdots(2c_j)}{(n+1)(n+2)\cdots(2n)} && (11)\\
&\geq \prod_{j=0}^{m-1}\left(\frac{c_j}{n}\right)^{c_j} && (12)\\
&\geq \prod_{j=0}^{m-1}\theta_j^{c_j} && (13)
\end{aligned}$$

(10)⇒(11)

$$\begin{aligned}
\Gamma(n+1/2) &= (n-1/2)(n-3/2)\cdots(1/2)\cdot\Gamma(1/2) \\
&= \frac{(2n-1)(2n-3)\cdots 1}{2^n}\Gamma(1/2) \\
&= \frac{2n\cdot(2n-1)\cdot(2n-2)\cdot(2n-3)\cdots 2\cdot 1}{n\cdot(n-1)\cdots 1\cdot 2^{2n}}\Gamma(1/2) \\
&= \frac{(n+1)(n+2)\cdots 2n}{2^{2n}}\Gamma(1/2),
\end{aligned}$$

$$\begin{aligned}
\prod_{j=0}^{m-1}\Gamma(c_j+1/2) &= \prod_{j=0}^{m-1}\left\{\frac{(c_j+1)(c_j+2)\cdots 2c_j}{2^{2c_j}}\Gamma(1/2)\right\} \\
&= \frac{\prod_{j=0}^{m-1}[(c_j+1)(c_j+2)\cdots 2c_j]}{2^{2n}}\Gamma(1/2)^m
\end{aligned}$$

(11)⇒(12)

$$\left(1+\frac{1}{n}\right)\left(1+\frac{2}{n}\right)\cdots 2 \le \prod_{j=0}^{m-1}\left(1+\frac{1}{c_j}\right)\left(1+\frac{2}{c_j}\right)\cdots 2 \text{ と同値}$$

$$1+\frac{i}{n} \le 1+\frac{k}{c_j} \iff k \ge \frac{c_j}{n} i,$$

$$x_j \ge 0 \Longrightarrow \sum_{j=0}^{m-1} \lceil x_j \rceil \le \left\lceil \sum_{j=0}^{m-1} x_j \right\rceil + m - 1$$

S_i: $1+i/n \le 1+k/c_j,\ 1 \le k \le c_j$ なる (j,k) の個数

$$\sum_{j=0}^{m-1}\left\lceil \frac{c_j}{n} i \right\rceil \le \left\lceil i \sum_{j=0}^{m-1} \frac{c_j}{n} \right\rceil + m - 1 = m - 1 + i$$

$$S_i = \sum_{j=0}^{m-1}\left(c_j - \left\lceil \frac{c_j}{n} i \right\rceil + 1\right) \ge \sum_{j=0}^{m-1} c_j + 1 - i = n + 1 - i.$$

$1+\dfrac{i}{n} \le 1+\dfrac{k}{c_j},\ 1 \le k \le c_j$ なる i と 1 対 1 対応の (j,k) が存在

(12)⇒(13)

$$\sum_{j=0}^{m-1} \frac{c_j}{n} \log\left(\frac{c_j}{n}/n\right) \geq 0$$

と同値

上界 (続)

(5), (9), (13) より,

$$
\begin{aligned}
D^n(P_\theta||Q_W) &= E_\theta \log \frac{P_\theta(Z^n)}{Q_W(Z^n)} \\
&\le E_\theta[-\log Q_W(z_*^n)] = \log \frac{\Gamma(1/2)\Gamma(n+m/2)}{\Gamma(m/2)\Gamma(n+1/2)} \\
&\le \frac{1}{2}(m-1)\log n - \log \frac{\Gamma(m/2)}{\Gamma(1/2)} + O(1/n) \qquad (14)
\end{aligned}
$$

最後の変形: m が奇数のとき、

$$
\begin{aligned}
\log \frac{\Gamma(n+m/2)}{n^{(m-1)/2}\Gamma(n+1/2)} &= \log\left\{\left(1+\frac{m/2-1}{n}\right)\cdots\left(1+\frac{1/2}{n}\right)\right\} \\
&\le \sum_{k=0}^{\frac{m-1}{2}} \frac{k+1/2}{n} = \frac{(m+1)^2}{8n}
\end{aligned}
$$

$$
\frac{\Gamma(n+m/2)}{n^{\frac{m-1}{2}}\Gamma(n+1/2)} \le \frac{\Gamma(n+(m+1)/2)}{n^{\frac{m}{2}}\Gamma(n+1/2)}
$$

より (m の帰納法), m が偶数でも成立。

下界

U: Θ の測度 ($U(\Theta(\Omega)) = 1$, W とは別もの)

$$I(\Theta, Z^n) := \int_{\Theta(\Omega)} D^n(P_\theta || Q_U)\, U(d\theta),$$

$$Q_U(z^n) := \int_{\Theta(\Omega)} P_\theta(z^n)\, U(d\theta)$$

$$\begin{aligned} \max_U I(\Theta, Z^n) &= \max_U \min_W \int D^n(P_\theta || Q_W)\, U(d\theta) \\ &\leq \min_W \max_U \int D^n(P_\theta || Q_W)\, U(d\theta) = \min_W \max_\theta D^n(P_\theta || Q_W). \end{aligned}$$

下界 (続)

最初の等式:

$$
\begin{aligned}
& \int_{\Theta(\Omega)} D^n(P_\theta || Q_W)\, U(d\theta) - \int_{\Theta(\Omega)} D^n(P_\theta || Q_U)\, U(d\theta) \\
= \; & \int_{\Theta(\Omega)} \sum_{z^n \in Z^n(\Omega)} P_\theta(z^n) \log \frac{P_\theta(z^n)}{Q_W(z^n)} U(d\theta) \\
& - \int_{\Theta(\Omega)} \sum_{z^n \in Z^n(\Omega)} P_\theta(z^n) \log \frac{P_\theta(z^n)}{Q_U(z^n)} U(d\theta) \\
= \; & \sum_{z^n \in Z^n(\Omega)} Q_U(z^n) \log \frac{Q_U(z^n)}{Q_W(z^n)} \geq 0
\end{aligned}
$$

微分エントロピー

確率変数 X の確率密度関数 f_X が存在するとき

X の微分エントロピー

$$h(X) := \int_{x \in X(\Omega)} -f_X(x) \log f_X(x)\, dx$$

確率変数 Y は離散 (確率分布: P_Y) で条件付密度関数 $f_{X|Y}$ が存在するとき

Y のもとでの X の条件付微分エントロピー

$$h(X|Y) := \sum_{y \in Y(\Omega)} \int_{x \in X(\Omega)} -f_{X|Y}(x|y) P_Y(y) \log \frac{f_{X|Y}(x|y) P_Y(y)}{P_Y(y)}\, dx$$

$h(x) \geq 0$, $h(X|Y) \geq 0$ は成立しない

$g : Y(\Omega) \to \mathbb{R}$ で、$g(Y)$ が確率変数のとき、$I(X, Y) \geq I(X, g(Y))$

$$
\begin{aligned}
I(X, Y) &:= h(X) - h(X|Y) \\
= &\int_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} f_{X|Y}(x|y) P_Y(y) \log \frac{f_{X|Y}(x|y) P_Y(y)}{f_X(x) P_Y(y)} dx \geq 0
\end{aligned}
$$

$$
\sum_{y \in Y(\Omega)} \int_{x \in X(\Omega)} f_{X|Y}(x|y) P_Y(y) \log \frac{f_{X|Y}(x|y) P_Y(y)}{f_{X|g(Y)}(x|g(y)) P_Y(y)} dx \geq 0
$$

$$
I(X, Y) = h(X) - h(X|Y) \geq h(X) - h(X|g(Y)) = I(X, g(Y)) \tag{15}
$$

下界 (続)

推定量 $\widehat{\theta}_j(Z^n) := c_j(Z^n)/n$, $\widehat{\theta}(Z^n) = (\widehat{\theta}_0(Z^n), \cdots, \widehat{\theta}_{m-1}(Z^n))$ は,
確率変数 Z^n の関数であるので,

$$\begin{aligned} I(\Theta, Z^n) &\geq I(\Theta, \widehat{\theta}(Z^n)) \\ &= h(\Theta) - h(\Theta|\widehat{\theta}(Z^n)) \\ &= h(\Theta) - h(\Theta - \widehat{\theta}(Z^n)|\widehat{\theta}(Z^n)) \\ &\geq h(\Theta) - h(\Theta - \widehat{\theta}(Z^n)). \end{aligned} \tag{16}$$

$\Theta = \theta$ を任意に固定すると、
$(\Theta_1 - \widehat{\theta}_1(Z^n), \cdots, \Theta_{m-1} - \widehat{\theta}_{m-1}(Z^n))$ は, 平均 0,

$$\sigma_{i,j}^2 := E[(\Theta_i - \widehat{\theta}_i(Z^n))(\Theta_j - \widehat{\theta}_j(Z^n))]$$

を成分とする共分散行列 $\Sigma = (\sigma_{ij}^2)_{1 \leq i,j \leq m-1}$ をもつ。

正規分布: 同じ分散で微分エントロピー最大の分布

f_X: 平均 0，共分散行列 Σ の正規分布の確率密度関数
g_X: 平均 0，共分散行列 Σ の任意の確率密度関数
$\log f_X(x)$ の平均が確率変数 X の分散のみに依存するので，

$$
\begin{aligned}
0 &\leq \int g_X(x) \log \left(\frac{g_X(x)}{f_X(x)} \right) dx \\
&= \int g_X(x) \log g_X(x) dx - \int g_X(x) \log f_X(x)\, dx \\
&= \int g_X(x) \log g_X(x) dx - \int f_X(x) \log f_X(x)\, dx
\end{aligned}
$$

f_X の微分エントロピー:

$$
\begin{aligned}
&\int -f_X(\mathbf{x}) \log f_X(\mathbf{x})\, d\mathbf{x} \\
&\quad = \int f_X(\mathbf{x}) \cdot \frac{1}{2}\, {}^t\mathbf{x}\Sigma^{-1}\mathbf{x}\, d\mathbf{x} + \int f_X(\mathbf{x}) \log(\sqrt{2\pi \det(\Sigma)})\, d\mathbf{x} \\
&\quad = \frac{1}{2} \log(2\pi e \det(\Sigma))
\end{aligned}
$$

下界 (続)

D: $\sigma_{jj}^{-1/2}$ を対角成分にもつ対角行列
$D\Sigma D$ の対角成分が 1 で，Σ は正定値対称行列なので，
その固有値を $\lambda_1, \cdots, \lambda_{m-1}$ とおくと，それらは正で

$$\{\det(D\Sigma D)\}^{1/(m-1)} = \left(\prod_{j=1}^{m-1} \lambda_j\right)^{1/(m-1)}$$

$$\leq \quad \frac{1}{m-1}\sum_{j=1}^{m-1} \lambda_j = \frac{1}{m-1}\mathrm{trace}(D\Sigma D) = 1$$

$$\det \Sigma \leq \prod_{j=1}^{m-1} \sigma_{jj}$$

同じ共分散行列をもつ分布で、正規分布が $h(\Theta - \widehat{\theta}(Z^n))$ を最大。

$$h(\Theta - \widehat{\theta}(Z^n)) \leq \frac{1}{2}\log[(2\pi e)^{m-1}\det(\Sigma)] \leq \sum_{j=1}^{m-1}\frac{1}{2}\log(2\pi e \sigma_{jj}^2).$$

下界 (続)

$$\Lambda := \{(\theta_1 \cdots, \theta_{m-1}) \,|\, 0 \le \theta_j \le 1,\ j = 1, \cdots, m{-}1,\ \theta_1 + \cdots + \theta_{m-1} \le 1\}$$

$$\int_\Lambda d\theta_1 \cdots d\theta_{m-1} = \frac{1}{(m-1)!}$$

$u(\theta) := (m-1)!$

$$\sigma_{jj}^2 = \int_\Lambda E_\theta \left(\frac{c_j(Z^n)}{n} - \theta_j \right)^2 U(d\theta) = \int_\Lambda \frac{\theta_j(1-\theta_j)}{n} U(d\theta) = \frac{1}{6n},$$

$$h(\Theta - \widehat{\theta}(Z^n)) \le \frac{1}{2}(m-1) \log \frac{\pi e}{3n}, \tag{17}$$

$$h(\Theta) = -\log(m-1)!\ . \tag{18}$$

(17), (18) を (16) に代入すると,

$$\begin{aligned} & \min_W \max_\theta D^n(P_\theta || Q_W) \\ \ge\ & \frac{1}{2}(m-1)\log n - \log(m-1)! - \frac{1}{2}(m-1)\log\frac{\pi e}{3} \end{aligned} \tag{19}$$

$\min_W \max_\theta D^n(P_\theta || Q_W)$ の上界下界

(14), (19) より、

定理 3.5 (Davisson, 1981)

$$
\begin{aligned}
&\frac{1}{2}(m-1)\log n - \log(m-1)! - \frac{1}{2}(m-1)\log\frac{\pi e}{3} \\
&\quad \le \min_W \max_\theta D^n(P_\theta || Q_W) \\
&\quad \le \frac{1}{2}(m-1)\log n - \log\frac{\Gamma(m/2)}{\Gamma(1/2)} + O(1/n) \qquad (20)
\end{aligned}
$$

$w = w_{a_0,\cdots,a_{m-1}}$ で $a_j = 1$ とおいた場合，w が一様分布になるが，
定理 18 より精密な上界は得られていない。

Jeffreys の事前確率

各 $\theta \in \Theta(\Omega)$ の近傍にある θ' について，その確率測度 μ_θ が θ の近傍にある各 θ' で $\mu_\theta \ll \mu_{\theta'}$ のとき，

確率変数 Θ に関する Fisher 情報行列

$I_{i,j} := \left[\left. \frac{\partial D(\mu_\theta || \mu_{\theta'})}{\partial \theta'_i \partial \theta'_j} \right|_{\theta'=\theta} \right]$ を成分とする行列 $I = (I_{i,j})$

$c^* := \int_\Lambda \sqrt{\det I(\theta)}\, d\theta < \infty$ のとき、

Jeffreys の事前確率

$$w^*(\theta) = \frac{\sqrt{\det I(\theta)}}{c^*} \tag{21}$$

Jeffreys の事前確率 (続)

例: $X(\Omega)=\{0,1\}$, $\mu_X\{1\}=\theta$, $\mu_X\{0\}=1-\theta$ について,

$$D(\mu_\theta||\mu_{\theta'})=\theta\log\frac{\theta}{\theta'}+(1-\theta)\log\frac{1-\theta}{1-\theta'},$$

$$\frac{\partial^2}{\partial(\theta')^2}D(\mu_\theta||\mu_{\theta'})=\frac{\theta}{\theta'^2}+\frac{1-\theta}{(1-\theta')^2}.$$

$\theta'=\theta$ を代入すると，Fisher 情報量行列はスカラーとなり,

$$I(\theta)=\frac{1}{\theta(1-\theta)}$$

$$c^*=\int_0^1\frac{d\theta}{\sqrt{\theta(1-\theta)}}=\pi$$

$$w^*(\theta)=\frac{1}{\pi\sqrt{\theta(1-\theta)}}$$

Clarke-Barron, 1994

- Jeffreys の事前確率が最適となるための十分条件
- $$\lim_{n\to\infty}\left[\min_W \max_\theta D^n(P_\theta||Q_W) - \frac{d(\Theta(\Omega))}{2}\log\frac{n}{2\pi e}\right]$$
$$= \log\int_\Lambda \sqrt{\det I(\theta)}\,d\theta$$
- 確率変数 Z の値域が有限集合となる場合に関しては，$a_j = 1/2$ の場合が Jeffreys の事前確率。
- $w_{1/2,\cdots,1/2}$ は Clarke-Barron の十分条件を満足しない
- 最適な W は絶対連続ではない (Xie-Barron, 1997)。

定理 3.5 の意味で十分最適に近いことから，Θ の事前確率 Q_W が未知である場合に，$w_{1/2,\cdots,1/2}$ が準最適な事前確率として用いられている。