「ベイズ統計の理論と方法」
Chapter 2, p.40-45

@tmiya_

June 28, 2017

定義 6

平均対数損失関数

$$L(w) = -\mathbb{E}_X[\log p(X|w)] = \mathbb{E}[L_n(w)]$$

経験対数損失関数

$$L_n(w) = -\frac{1}{n}\sum_{i=1}^{n} \log p(X_i|w)$$

自由エネルギー

$$F_n(\beta) = -\frac{1}{n}\log Z_n(\beta)$$

汎化損失

$$G_n = -\mathbb{E}_X[\log \mathbb{E}_w[p(X|w)]]$$

経験損失

$$T_n = -\frac{1}{n}\sum_{i=1}^{n} \log \mathbb{E}_w[p(X_i|w)]$$

定義 7

定義 7: **真の分布** $q(x)$ **と確率モデル** $p(x|w)$ **から定まる対数尤度比関数を** $f(x,w)$ **とする。（と、書かれているが、真の分布と確率モデルが決まると** W_0 **が決まるので、** f **は** w_0 **を使って定める。）**

$$
\begin{aligned}
f(x,w) &= \log \frac{p(x|w_0)}{p(x|w)} \\
K(w) &= \mathbb{E}_X[f(X,w)] \quad \textbf{平均誤差関数} \\
K_n(w) &= \frac{1}{n}\sum_{i=1}^{n} f(X_i,w) \quad \textbf{経験誤差関数}
\end{aligned}
$$

定義より

$$
\begin{aligned}
L(w) &= L(w_0) + K(w) \\
\because -\mathbb{E}_X[\log p(X|w)] &= -\mathbb{E}_X[\log p(X|w_0)] + \mathbb{E}_X[\log \frac{p(x|w_0)}{p(X|w)}] \\
L_n(w) &= L_n(w) + K_n(w) \\
\because -\frac{1}{n}\sum_i \log p(X_i|w) &= -\frac{1}{n}\sum_i \log p(X_i|w_0) + \frac{1}{n}\sum_i \log \frac{p(X_i|w_0)}{p(X_i|w)}
\end{aligned}
$$

定義 7

KL 情報量の性質から (?)$K(w) \geq 0$ であり、
$K(w) = 0 \Leftrightarrow w \in W_0$ である。

$$K(w) = L(w) - L(w_0) \geq 0 \text{ (等号成立は } w \in W_0)$$

$$\begin{aligned}
K(w) &= \mathbb{E}_X[\log \frac{p(x|w_0)}{p(x|w)}] = \int q(x) \log \frac{p(x|w_0)}{p(x|w)} dx \\
&= -\int q(x) \log \frac{q(x)}{p(x|w_0)} dx + \int q(x) \log \frac{q(x)}{p(x|w)} dx \\
&= -\mathrm{KL}(q \| p_{w_0}) + \mathrm{KL}(q \| p_w)
\end{aligned}$$

定義 7

$$Z_n^{(0)}(\beta) = \int \exp\left(-n\beta K_n(w)\right)\varphi(w)dw.$$ (**正規化された分配関数**)

$$\prod_i p(X_i|w) = \left(\prod_i p(X_i|w_0)\right)\exp(-nK_n(w)).$$

$$\therefore Z_n(\beta) = \exp(-n\beta L_n(w_0)) \cdot Z_n^{(0)}(\beta).$$ (**分配関数**)

$$p(w|X^n) = \frac{1}{Z_n^{(0)}}\exp(-n\beta K_n(w))\varphi(w).$$ (**事後分布**)

定義 8

正規化された自由エネルギー

$$F_n^{(0)}(\beta) = -\frac{1}{\beta}\log\int \exp(-n\beta K_n(w))\varphi(w)dw.$$

正規化された汎化誤差 $G_n^{(0)}$ と経験誤差 $T_n^{(0)}$

$$\begin{aligned}
G_n^{(0)} &= -\mathbb{E}_X[\log \mathbb{E}_w[\exp(-f(X,w))]], \\
&= \mathbb{E}_X\left[\log \frac{q(X)}{\mathbb{E}_w[p(X|w)]}\right] \quad \Leftarrow ??? \\
T_n^{(0)} &= -\frac{1}{n}\sum_{i=1}^{n}\log \mathbb{E}_w[\exp(-f(X_i,w))]. \\
&= \frac{1}{n}\sum_{i}\log \frac{q(X_i)}{\mathbb{E}_w[p(X_i|w)]}. \quad \Leftarrow ???
\end{aligned}$$

（ここから数ページ、$q(x)$ と $p(x|w_0)$ と混乱しているのではないかという疑いあり？ ）

補題 5

$$
\begin{aligned}
F_n(\beta) &= -\frac{1}{\beta}\log Z_n(\beta) \\
&= -\frac{1}{\beta}\log \int \exp(-n\beta \underbrace{L_n(w)}_{L_n(w_0)+K_n(w)})\varphi(w)dw \\
&= nL_n(w_0) - \frac{1}{\beta}\log \int \exp(-n\beta K_n)\varphi(w)dw \\
&= nL_n(w_0) + F_n^{(0)}(\beta)
\end{aligned}
$$

補題 5

$$
\begin{aligned}
G_n(\beta) &= -\mathbb{E}_X[\log \mathbb{E}_w[\underbrace{p(X|w)}_{p(X|w_0)e^{-f(X,w)}}]] \\
&= \underbrace{-\mathbb{E}_X[\log p(X|w_0)]}_{L(w_0)} \underbrace{-\mathbb{E}_X[\log \mathbb{E}_w[e^{-f(X,w)}]]}_{G_n^{(0)}(\beta)} \\
&= L(w_0) + G_n^{(0)}. \\
T_n(\beta) &= -\frac{1}{n}\sum_i \log \mathbb{E}_w[\underbrace{p(X_i|w)}_{p(X_i|w_0)e^{-f(X,w)}}] \\
&= \underbrace{-\frac{1}{n}\sum_i \log p(X_i|w_0)}_{L_n(w_0)} \underbrace{-\frac{1}{n}\sum_i \log \mathbb{E}_w[e^{-f(X_i,w)}]}_{T_n^{(0)}} \\
&= L_n(w_0) + T_n^{(0)}.
\end{aligned}
$$

注意 15

以上で述べたことより、自由エネルギー、汎化損失、経験損失の挙動を知るためには、正規化された自由エネルギー、汎化損失、経験損失の挙動が判れば良い。

注意 16

対数尤度比関数が相対的に有限な分散を持つときは、ある定数 $\lambda, m > 0$ が存在して $n \to \infty$ のとき、

$$\hat{Z}_n^{(0)}(\beta) = \frac{n^\lambda}{(\log n)^{m-1}} Z_n^{(0)}(\beta).$$

は確率変数として法則収束する。事後分布が正規分布で近似できる場合は、$\beta = 1$ のとき、$\lambda = d/2,\ m = 1$ である。

$$Z_n^{(0)}(1) = \int \prod_i \frac{p(X_i|w)}{q(X_i)} \varphi(w) dw \quad \Leftarrow ???\ q(x)?$$

と書けるので、任意の n について $\mathbb{E}[Z_n^{(0)}(1)] = 1$ がなりたつ。すなわち、$\hat{Z}_n^{(0)}(1)$ は $n \to \infty$ で法則収束するが平均 $\mathbb{E}[\hat{Z}_n^{(0)}(1)]$ は発散している。分配関数は、確率変数としての挙動と平均値の挙動が同じオーダーではない。
一方、確率変数 $-\log \hat{Z}_n^{(0)}(\beta)$ は法則収束し、平均値も収束する。よって挙動を見るときは、分配関数より自由エネルギーが良い。

定義 9

補題 6

$$
\begin{aligned}
\mathcal{G}_n^{(1)}(\alpha) &= \frac{d}{d\alpha}\mathbb{E}_X[\log \mathbb{E}_w[p(X|w)^\alpha]] \\
&= \mathbb{E}_X\left[\frac{\frac{d}{d\alpha}\mathbb{E}_w[p(X|w)^\alpha]}{\mathbb{E}_w[p(X|w)^\alpha]}\right] \\
&= \mathbb{E}_X\left[\frac{\mathbb{E}_w[p(X|w)^\alpha(\log p(X|w))]}{\mathbb{E}_w[p(X|w)^\alpha]}\right] = \mathbb{E}_X[\ell_1(X)]. \\
\mathcal{G}_n^{(2)}(\alpha) &= \frac{d^2}{d\alpha^2}\mathbb{E}_X[\log \mathbb{E}_w[p^\alpha]] \\
&= \frac{d}{d\alpha}\mathbb{E}_X\left[\frac{\mathbb{E}_w[p^\alpha \log p]}{\mathbb{E}_w[p^\alpha]}\right] \\
&= \mathbb{E}_X\left[\frac{\frac{d}{d\alpha}\mathbb{E}_w[p^\alpha \log p]\mathbb{E}_w[p^\alpha] - \mathbb{E}_w[p^\alpha \log p]\frac{d}{d\alpha}\mathbb{E}_w[p^\alpha]}{\mathbb{E}_w[p^\alpha]^2}\right] \\
&= \mathbb{E}_X\left[\frac{\mathbb{E}_w[p^\alpha(\log p)^2]\mathbb{E}_w[p^\alpha] - \mathbb{E}_w[p^\alpha \log p]^2}{\mathbb{E}_w[p^\alpha]^2}\right] \\
&= \mathbb{E}_X[\ell_2(X) - \ell_1(X)^2].
\end{aligned}
$$

補題 6

$\mathcal{G}_n^{(k)}(\alpha)$ を求めるには、

$$g(X,\alpha) = \mathbb{E}_w[p(X|w)^\alpha] = \int p(X|w)^\alpha p(w|X^n)dw$$

に対して、$k \geq 1$ で

$$\ell_k(X) = \frac{g^{(k)}(X,\alpha)}{g(X,\alpha)}$$

であり、あとは商の微分法の公式より

$$\begin{aligned}\left(\frac{d}{d\alpha}\right)\left(\frac{g^{(k)}(X,\alpha)}{g(X,\alpha)}\right) &= \frac{g^{(k+1)}(X,\alpha)}{g(X,\alpha)} - \frac{g^{(k)}(X,\alpha)}{g(X,\alpha)}\frac{g^{(1)}(X,\alpha)}{g(X,\alpha)} \\ &= \ell_{k+1}(X) - \ell_k(X)\ell_1(X)\end{aligned}$$

別ファイル参照

See p.46-47 **の資料**

補題 9

サンプルの現れ方に関して平均したものは更に次の関係がなりたつ。ベイズ推論における汎化損失と経験損失の間の特別な関係である。4 章で使う。

$$\mathbb{E}[\mathcal{G}_{n-1}(\beta)] = -\mathbb{E}[\mathcal{T}_n(-\beta)]$$

. 従って 3 次以上のキュムラントが $1/n$ より小さければ、

$$\mathbb{E}\left[\mathcal{G}'_{n-1}(0) + \frac{\beta}{2}\mathcal{G}''_{n-1}(0)\right] = \mathbb{E}\left[\mathcal{T}'_n(0) - \frac{\beta}{2}\mathcal{T}''_n(0)\right] + o\left(\frac{1}{n}\right).$$

証明

後半は $\beta = 0$ のところで $\mathcal{G}_{n-1}$, $\mathcal{T}_n$ を 2 次の項までテーラー展開して得られる。

補題 9

前半を示す

$$
\begin{aligned}
&\mathbb{E}[\mathcal{G}_{n-1}(\beta)] \\
&= \mathbb{E}[\mathbb{E}_X[\log \mathbb{E}_w[p(X|w)^\beta]]] \\
&= \mathbb{E}\left[\int q(x)\log\left(p(x|w)^\beta \frac{\varphi(w)\prod_{i=1}^{n-1} p(X_i|w)^\beta}{\int \varphi(w)\prod_{i=1}^{n-1} p(X_i|w)^\beta dw} dw\right) dx\right].
\end{aligned}
$$

真の分布についての平均ゆえ $p(x|w)$ を $p(X_n|w)$ と思っても良い。

$$
\begin{aligned}
&= \mathbb{E}\left[\int q(x)\log\left(\frac{\int \varphi(w)\prod_{i=1}^{n} p(X_i|w)^\beta dw}{\int \varphi(w)\prod_{i=1}^{n-1} p(X_i|w)^\beta dw} dw\right) dx\right]. \\
&= \mathbb{E}\left[\int q(x)\log\frac{Z_n(\beta)}{Z_{n-1}(\beta)} dx\right] = \mathbb{E}\left[\log\frac{Z_n(\beta)}{Z_{n-1}(\beta)}\right] \\
&= -\mathbb{E}\left[\log\frac{Z_{n-1}(\beta)}{Z_n(\beta)}\right].
\end{aligned}
$$

補題 9

$$
\begin{aligned}
&= -\mathbb{E}\left[\log \frac{Z_{n-1}(\beta)}{Z_n(\beta)}\right] \\
&= -\mathbb{E}\left[\log\left(\frac{1}{Z_n(\beta)}\int p(X_n|w)^{-\beta}\varphi(w)\prod_{i=1}^{n} p(X_i|w)^{\beta}dw\right)\right] \\
&= -\mathbb{E}\left[\log\left(\int p(X_n|w)^{-\beta}\frac{\varphi(w)\prod_{i=1}^{n} p(X_i|w)^{\beta}}{Z_n(\beta)}dw\right)\right] \\
&= -\mathbb{E}\left[\log \mathbb{E}_w[p(X_n|w)^{-\beta}]\right].
\end{aligned}
$$

ここで、サンプルの出方についての平均であるから

$$
\mathbb{E}\left[\log \mathbb{E}_w[p(X_1|w)^{-\beta}]\right] = \cdots = \mathbb{E}\left[\log \mathbb{E}_w[p(X_n|w)^{-\beta}]\right]
$$

と考えて良い。よって

$$
\begin{aligned}
\mathbb{E}[\mathcal{G}_{n-1}(\beta)] &= -\mathbb{E}\left[\log \mathbb{E}_w[p(X_n|w)^{-\beta}]\right] \\
&= -\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\log \mathbb{E}_w[p(X_i|w)^{-\beta}]\right] \\
&= -\mathbb{E}[\mathcal{T}_n(-\beta)].
\end{aligned}
$$