黒木玄 **Gen Kuroki** @genkuroki on May 25
mathtod.online/@chijan/174433
式も書いてくれないと、何が疑問なのか、わからないのですが、勝手に補足。

以下、面倒なので $\beta = 1$.

『ベイズ統計の理論と方法』の2.2節では汎化損失の定義は

$$G_n = -E_X[\log E_w[p(X|w)]].$$

となっていて、

$$G_n = -\int q(x) \log p_*(x)$$

と書き直せます。$q$ は真の分布で $p_*$ は予測分布:

$$p_*(x) = \int dw\, p(x|w) p(w|X^n).$$

ここで $p(w|X^n)$ はサンプル $X^n$ から得られる事後分布:

$$p(w|X^n) \propto \varphi(w) \prod_{i=1}^n p(X_i|w).$$

続く

黒木玄 **Gen Kuroki** @genkuroki on May 25
続き

予測分布の真の分布に対する誤差を測るKL情報量と真の分布のShannon情報量はそれぞれ

$$D(q||p_*) = \int dx\, q(x) \log \frac{q(x)}{p_*(x)},$$
$$S(q) = -\int dx\, q(x) \log q(x)$$

なので

$$D(q||p_*) = G_n - S(q).$$

真の分布の情報量 $S(q)$ は真の分布で決まる定数なので、予測分布の真の分布に対する誤差 $D(q||p_*)$ がより小さいことと、汎化損失 $G_n$ がより小さいことは同値。

どの部分が「見慣れない」んだろうか?

私にはAIC以降では普通になった見慣れた式しか出て来ていないように見えた。

AICもWAICも $G_n$ の推定量になっている。

真の分布の情報量 $S(q)$ の推定は困難なので、KL情報量 $D(q\|p_*)$ の推定も困難。しかし、$G_n$ の推定ならば可能という話。

私はこの分野のど素人なので注意。

黒木玄 **Gen Kuroki** @genkuroki on May 25
続き

渡辺澄夫『ベイズ統計の理論と方法』は以上上に書いたようなことが書いてあるとみなせるので、Kullback-LeiblerおよびShannon情報量に関する基礎知識 (特にSanovの定理)に関する予備知識がないと、いきなりわからなくなる可能性が高いです。

統計力学の文脈では、KL情報量の $-1$ 倍は相対エントロピー、S情報量の $-1$ 倍はエントロピーと呼ばれているので、統計力学の素養がある人にとっては楽しい本だと思います。

統計力学を知らない人向けのその辺の事柄に関する解説が次の場所にある。

github.com/genkuroki/Sanov

統計力学的な素養は広く教養として広まった方が良いと思う。ベイズ統計や機械学習と独立に高い価値があると思います。

黒木玄 **Gen Kuroki** @genkuroki on May 25
mathtod.online/@chijan/175184

$f(x,w)$ の定義を単純に誤解しているのでは？

『ベイズ統計の理論と方法』のpp.35-36に $f(x,w)$ の定義が書いてあります。

$f(x,w)$ の定義は $\log(q(x)/p(x|w))$ ではなく、

$$f(x,w) = \log\frac{p_0(x)}{p(x|w)} = \log\frac{p(x|w_0)}{p(x|w)}.$$

黒木玄 **Gen Kuroki** @genkuroki on May 25
他人の記号法をそのまま受け入れて本を読むのは難しいので、全部自分でやり直した方が誤解が少なくなる場合が多いと思います。

本の筆者自身が書き間違えていることは普通によくあることなので、本に書いてある通りに忠実に読もうとすると、余計に苦労することが多い。

自分自身のスタイルと本の筆者のスタイルが違っているせいで、誤解が生じて苦労することも非常によくある。

他人のスタイルに合わせるとひどく苦労する場合が多いと思う。

だから、借りて来たスタイルじゃないやり方をしている人を見ると安心する。

黒木玄 **Gen Kuroki** @genkuroki on May 25

mathtod.online/@chijan/175345

定義5には、添付画像を見ればわかるように、$f(x,w)$ という記号を

$$f(x,w)=f(x,w_0,w)=\log\frac{p(x|w_0)}{p(x|w)}$$

という意味で使うとはっきり書いてあります。

引用していない箇所(p.33の定義4)には $p_0(x)$ は $p(x|w_0)$ が $w_0$ によらない場合の $p(x|w_0)$ を意味すると書いてあります。

「定義5からそう読み取るのは少し厳しい」と思った理由が不明。

この件で『ベイズ統計の理論と方法』の著者の渡辺澄夫さんは無実だと思いました。

mathtod.online/media/eBPfXu59U...

---

黒木玄 **Gen Kuroki** @genkuroki on May 25

mathtod.online/@tmiya/175370

個人的な意見では『ベイズ統計の理論と方法』は3.2.3節「分配関数の主要項」の解説から読み始めれば読み易いと思います。

積分をガウス積分で近似するLaplaceの方法を知っていれば、Laplaceの方法を単純に適用する話をしていることがすぐにわかります。

Laplaceの方法を知らなければ、最初に次のリンク先のp.5にあるStirlingの公式のLaplaceの方法を使った証明を読んでおくとよいです。
github.com/genkuroki/Sanov

Laplaceの方法については
github.com/genkuroki/GenLaplac...
のpp.1-2にも簡潔な解説があります。

Laplaceの方法はまるで空気のごとくよく使われるので、早めに知っておいた方がお得。

Laplaceの方法を知っていれば、「Laplaceの方法が通用しない場合について第4章に書いてある」と理解できます。

黒木玄 **Gen Kuroki** @genkuroki on May 25

「"Gaussian" による近似で解決できない問題をどう扱うか」は結構基本的な問題です。

Atiyahさんは1970年の論文で広中の特異点解消を使えばGaussianでないケースも積分の漸近挙動がわかるという指摘をしました。

柏原さんも$b$函数に関する有名な仕事で広中の特異点解消を本質的な形で使いました。

渡辺澄夫さんは京大数理研にいたときに柏原さんの仕事について柏原さん自身から聞いている可能性があると思います。

"Gaussian"による近似＝Laplaceの方法は特異点解消定理を使った後の積分の形であればリンク先のように一般化されます。
github.com/genkuroki/GenLaplac…

---

黒木玄 **Gen Kuroki** @genkuroki on May 25

$L_i, a_i, b_i > 0$とし、$\lambda_i = b_i/a_i$ $(i = 1, \dots, d)$ の最小値を $\lambda$ と書き、$\lambda_i = \lambda$ となる $i$ の個数を $m$ と書く。$Z_n, F_n$ を

$$\begin{aligned} Z_n &= \int_0^{L_1} dx_1 \cdots \int_0^{L_d} dx_d \\ &\quad \times \exp(-n x_1^{a_1} \cdots x_d^{a_d}) x_1^{b_1 - 1} \cdots x_d^{b_d - 1}, \\ F_n &= -\log Z_n \end{aligned}$$

と定めると、$n \to \infty$ で

$$F_n = \lambda \log n - (m-1) \log\log n + O(1).$$

この結果は特別な飛び道具無しに大学1年性レベルの計算で直接示せる。$d = 2, 3$ の計算を実行すればすぐにわかる。

このことに気付けば渡辺澄夫著『ベイズ統計の理論と方法』の第4章を少し読みやすくなると思う。

第4章をあのまま読むのは結構難しい。

---

黒木玄 **Gen Kuroki** @genkuroki on May 26

$-\log Z_n$ の leading term $\lambda \log n$ が出て来る仕組みは、$t = u/n$ による積分変数の置換による

$$\begin{aligned} &\int_0^\infty e^{-nt} t^{\lambda - 1}\,dt \\ &= \int_0^\infty e^{-u} n^{-\lambda} u^{\lambda - 1}\,du \\ &= n^{-\lambda} \Gamma(\lambda) \end{aligned}$$

と同じです(この計算は頻出)。

$$-\log(n^{-\lambda}\Gamma(\lambda)) = \lambda \log n + \text{const.}$$

黒木玄 **Gen Kuroki** @genkuroki on May 26
$-\log Z_n$ の subleading term $-\log\log n$ の項が出て来る仕組みは、$a, b > 0$ に関する

$$\int_{a/n}^{b} x^{-1}\,dx \\ = \log n + \mathrm{const.} \\ = (\log n)(1+o(1))$$

です。

$$-\log((\log n)(1+o(1))) \\ = -\log\log n + o(1).$$

ガンマ函数のよくある使い方や高校レベルの対数の使い方から、渡辺澄夫さんが示したベイズ統計における自由エネルギーの漸近挙動

$$-\log Z_n = \lambda \log n - (m-1)\log\log n + O(1)$$

が出て来るわけ(これが私の指摘)。

そしてそういう平易な計算に帰着するための数学的大道具が広中の特異点解消定理。認めて使って良いとみなされている大定理の典型例。

黒木玄 **Gen Kuroki** @genkuroki on May 26
『ベイズ統計の理論と方法』の定式化のスタイルが統計力学に由来することは明らか。

統計力学について学ぶと「分配函数さえ計算できればすべてが分かる」というような説明によく出会います。

十分にパラメーターを入れた分配函数(母函数)が計算できれば他の量はそこから得られる。

『ベイズ統計の理論と方法』の第2章の後半は「十分にパラメーターを入れた分配函数を計算できればすべてがわかる」という主張をベイズ統計の場合に正当化するために費やされています。

十分にパラメーターを入れた($\beta$ だけではなく、$\alpha$ も入れた)分配函数で他の量を書くという視点で読み直すと、一般的な構造が見えて来て読み易くなると思います。

分配函数の様子がよくわかるというようなことは滅多にありません。「絵に描いた餅」になる可能性がある。

しかし、最初に3.2.3節のLaplaceの方法を使える場合の計算を見れば、ベイズ統計における分配函数の漸近挙動の解析が「絵に描いた餅」ではないことがわかり、全体のストーリーを受け入れ易くなるはず。ベイズ情報量規準BICがそれだけで理解できてしまうこともうれしい。

黒木玄 **Gen Kuroki** @genkuroki on May 26
続き

分配函数の典型例は

$$Z = \int dx\, e^{-\alpha f(x) - \beta g(x)} \varphi(x)$$

のような形をしていて、これを用いて $x$ に関する確率分布(密度函数)が

$$p(x|\alpha, \beta) = \frac{1}{Z} e^{-\alpha f(x) - \beta g(x)} \varphi(x)$$

のように定義されます(統計力学でのカノニカル分布)。この確率分布 $p(x|\alpha, \beta)$ に関する $f(x)^k g(x)^l$ の平均は

$$\frac{1}{Z}\left(-\frac{\partial}{\partial\alpha}\right)^k \left(-\frac{\partial}{\partial\beta}\right)^l Z$$

と分配函数だけで書けます。

黒木玄 **Gen Kuroki** @genkuroki on May 26

$L_i, a_i, b_i > 0$ とし、$M_i = L_i^{a_i}, \lambda_i = b_i/a_i$ とおき、

$$\lambda := \lambda_1 = \cdots = \lambda_m < \lambda_{m+1}, \ldots, \lambda_d$$

と仮定する。

$$Z_n = \int_0^{L_1} dx_1 \cdots \int_0^{L_d} dx_d \\ \times \exp(-n x_1^{a_1} \cdots x_d^{a_d}) x_1^{b_1 - 1} \cdots x_d^{b_d - 1}$$

とおく。このとき、$x_i^{a_i}$ を $x_i$ で置換すると、定数倍の違いを除いて

$$Z_n = \int_0^{M_1} dx_1 \cdots \int_0^{M_d} dx_d \\ \times \exp(-n x_1 \cdots x_d) x_1^{\lambda_1 - 1} \cdots x_d^{\lambda_d - 1}$$

続く

黒木玄 **Gen Kuroki**
@genkuroki

続き

$$x_1 = \frac{t}{n x_2 \ldots x_d}$$

によって積分変数を $t, x_2, \ldots, x_d$ に置換すると主要因子の

$$n^{-\lambda}$$

が出ます。

そして、積分のネストを外側から

$$x_{m+1}, \dots, x_d, t, x_2, \dots, x_m$$

の順序になるように書き直します。

一番内側の $x_2, \dots, x_m$ に関する積分を実行すると

$$(\log n)^{m-1}(1 + o(1))$$

の形になり、二番目の因子が得られます。

$t$ に関する積分は本質的に定数である $\Gamma(\lambda)$ 倍の寄与になり、一番外側の $x_{m+1}, \dots, x_d$ に関する積分も本質的に定数倍の寄与に過ぎなくなります。

その結果

$$Z_n = n^{-\lambda}(\log n)^{m-1}(\mathrm{const.} + o(1))$$

が得られます。これの対数の $-1$ 倍が自由エネルギーです。

ベイズ統計の自由エネルギーの漸近挙動の解析は以上のような計算になります。

2017年05月26日 15:27 · Web · 1 · 2 · Webで開く