学習理論の練習 11 付録

このファイルでは学習理論の基本定理に関心がある人に
数学的な基礎を短く紹介しています。

学習理論という広い研究分野の中にたくさんの研究がありますが
その中のとても小さな一部分としてこのようなものがあるという例です。

もちろん関心のない人はお読みになる必要はありません。

学習理論の基本定理

学習理論のゼータ関数

真の分布 $q(x)$, 学習モデル $p(x|w)$, 事前分布 $\varphi(w)$ が与えられたとする。

$$L(w) = - \int q(x) \log p(x|w)\, dx.$$

の最小値を与えるパラメータを w_0 (ユニークとは限らない)、関数 $K(w)$ を

$$K(w) = \int q(x) \log \{ p(x|w_0)/p(x|w) \}\, dx$$

と定義する。学習理論のゼータ関数を

$$\zeta(z) = \int K(w)^z \varphi(w)\, dw \qquad (z \text{ in } \mathbf{C})$$

と定義すると、この関数は $\mathrm{Re}(z)>0$ で正則であり、全複素平面に有理型関数として一意に解析接続できることを証明できる(Atiyah)。その極はすべて負の有理数である(Kashiwara)。原点に一番近い極を $-\lambda$ としその位数を m とするとき λを実対数閾値といい、m を多重度という。

復習：学習の漸近理論

データの数 n が無限大に近づくとき、ベイズ汎化損失と自由エネルギーの平均値 $E[G_n]$ と $E[F_n]$ は次の漸近挙動を持つことが証明されている。

$$E[G_n] = L(w_0) + \lambda/n + o(1/n),$$

$$E[F_n] = n\,L(w_0) + \lambda \log n - (m-1)\log\log n + O_n,$$

ここで $o(1/n)$ は $(1/n)$ よりも早く零に近づく数列であり, O_n は定数に収束する数列である。

研究課題の例：

(1) 統計学や機械学習に現れる（深層学習や混合正規分布などの）($q(x), p(x|w), \varphi(w)$) に対して実対数閾値と多重度を求めよ。

(2) その結果を統計的モデル選択やハイパーパラメータの設計に活用するアルゴリズムを考案し有用性を示せ。

(3) データから G_n や F_n を計算する方法を構成し理論と比較せよ。

ブローアップ

射影空間

実数の組 $(x,y) \in \mathbf{R}^2 - O$ の同値関係

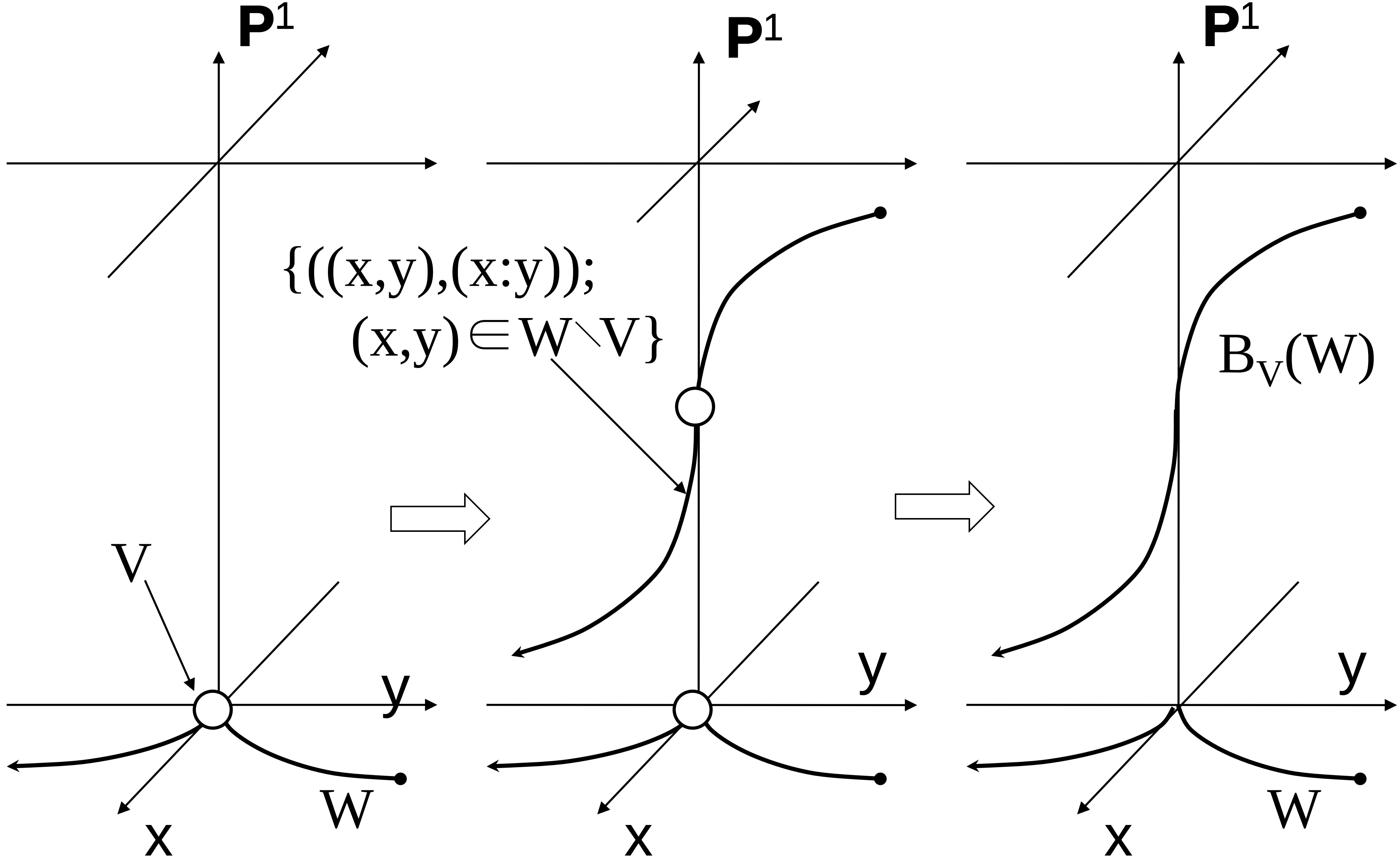
$(x,y) \sim (a,b) \Leftrightarrow \exists c \neq 0,\ (x,y)=c(a,b)$

1次元射影空間の定義 $\mathbf{P}^1 = \mathbf{R}^2/\sim$.

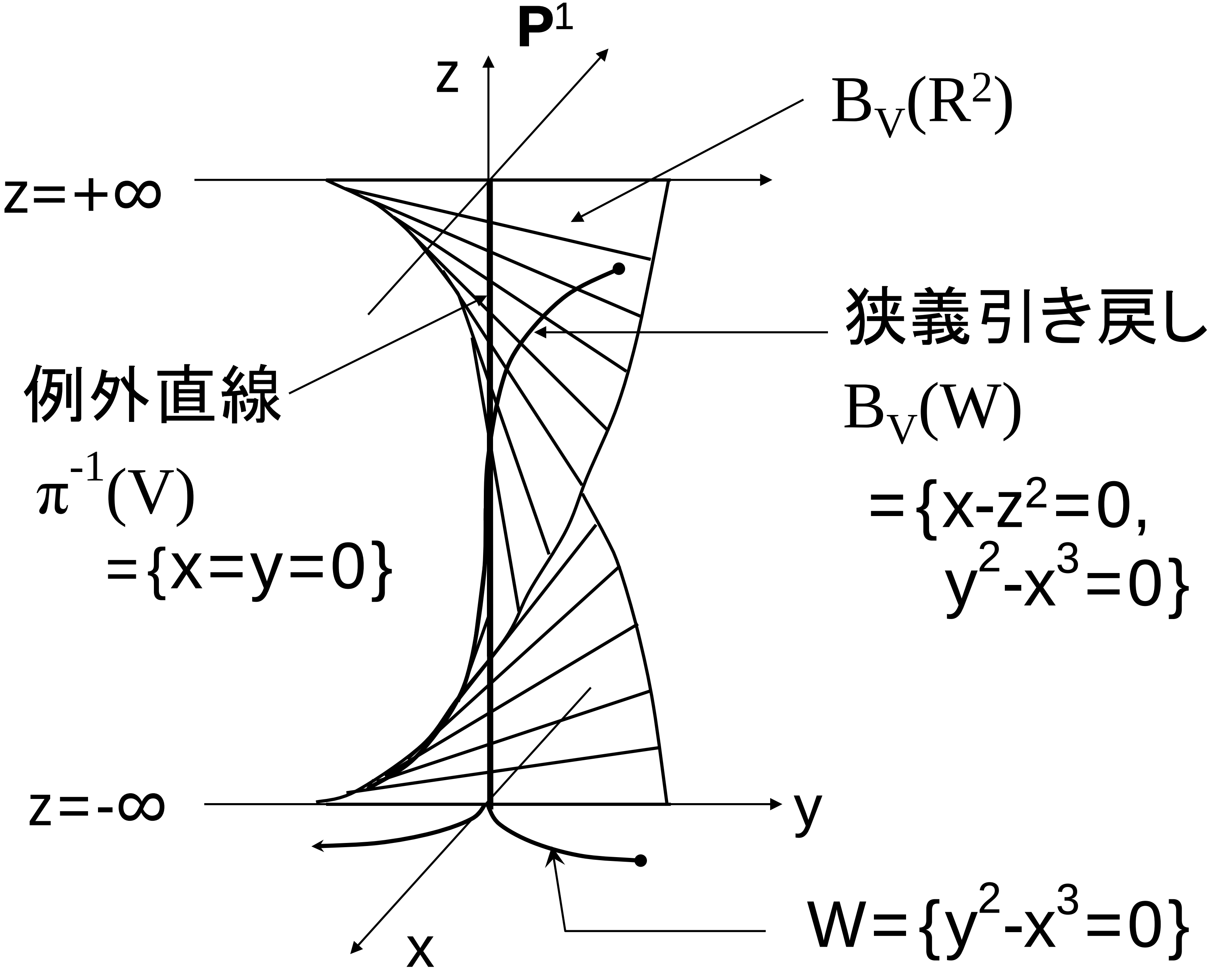
$(x,y) \in \mathbf{R}^2 - O$ の同値類を $(x:y)$ と書く。

$V \subseteq W$ とする。$V=O$ とするとき V を中心としたWの
ブローアップの定義

$$B_V(W) = \overline{\{(x,y,(x:y))\ ;\ (x,y) \in W - V\}}$$



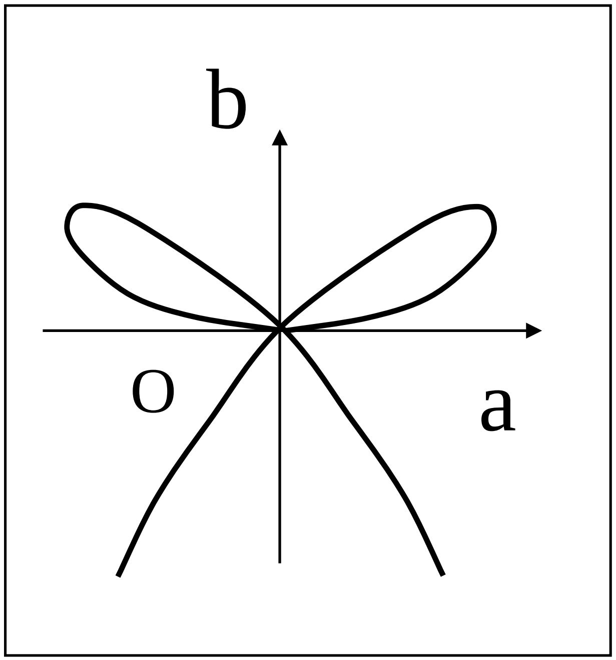
WからVを除く　　ブローアップ　　閉包をとる

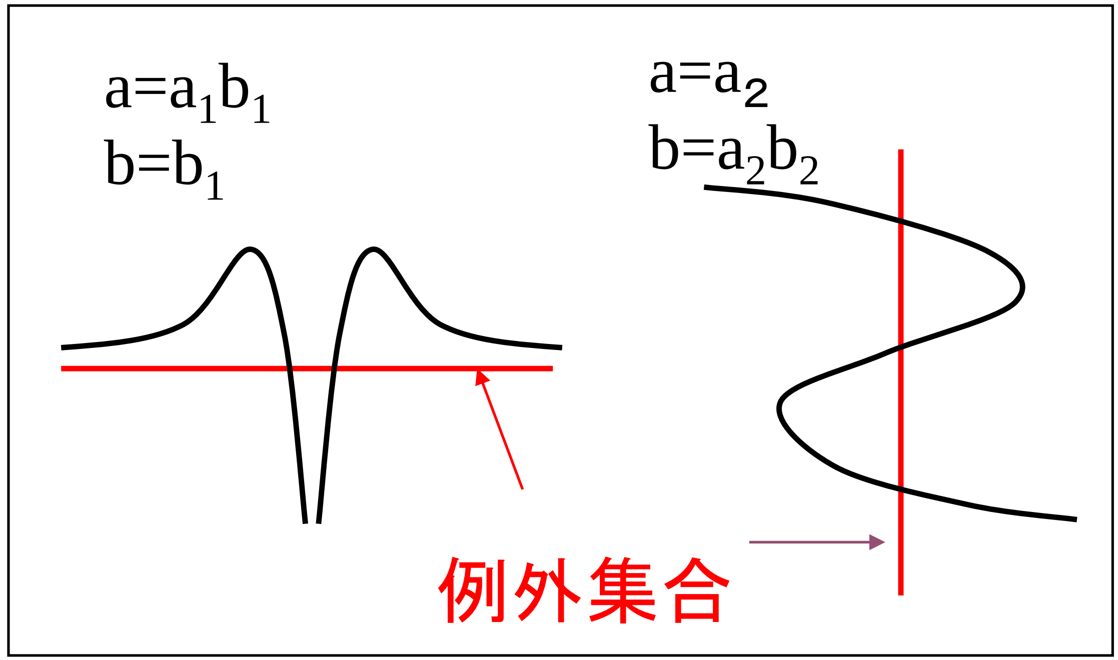


$\mathbf{P}^1$
z
$B_V(R^2)$
z=+∞
狭義引き戻し
例外直線
$B_V(W)$
$=\{x-z^2=0,$
$y^2-x^3=0\}$
$\pi^{-1}(V)$
$=\{x=y=0\}$
z=-∞
y
x
$W=\{y^2-x^3=0\}$

ブローアップによる特異点解消

座標を分けて計算

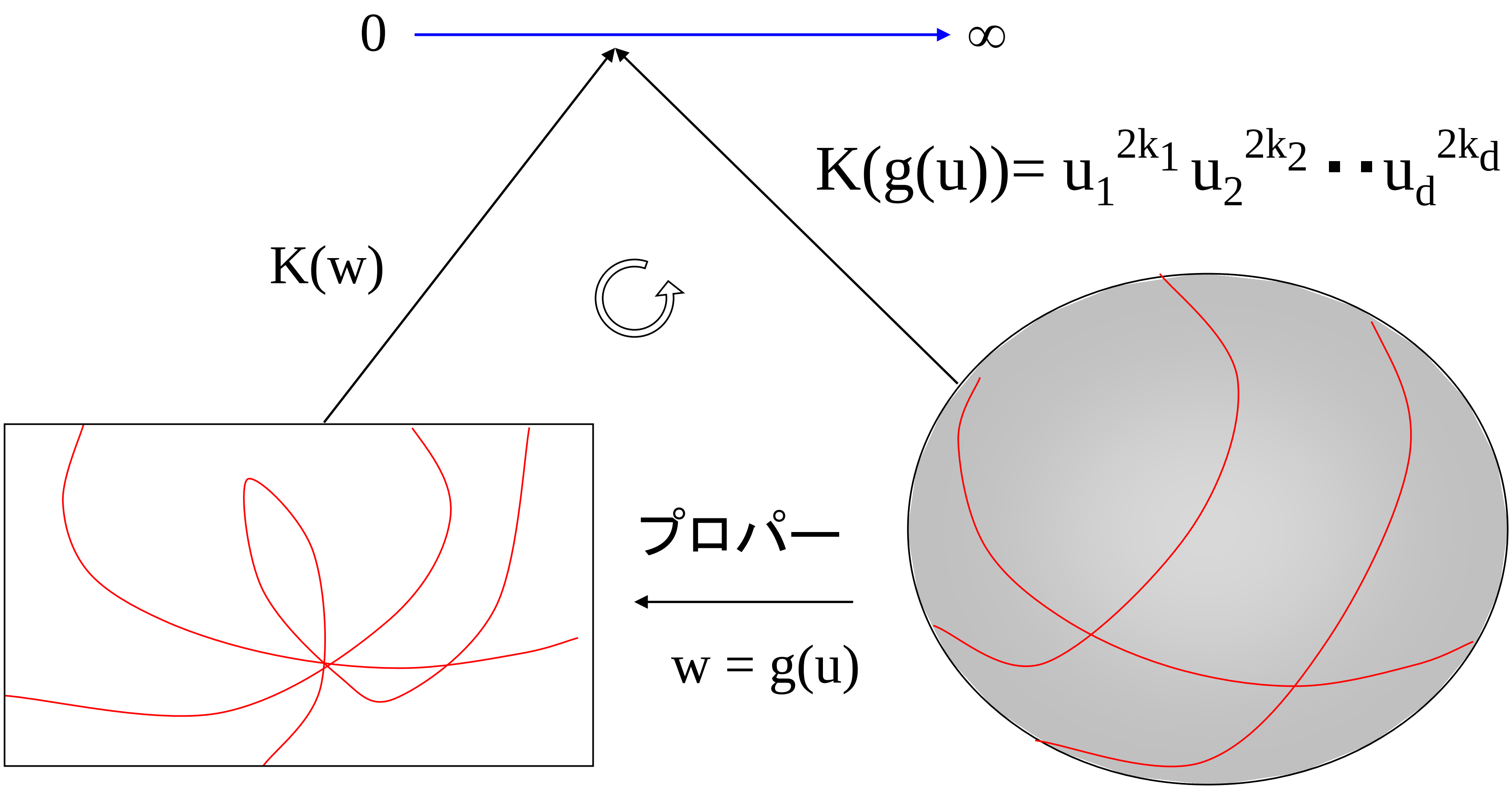




$K(a,b)= a^4-a^2b+b^3$

$K=b_1{}^3(a_1{}^4b_1-a_1{}^2+1)$ $K=a_2{}^3(a_2-b_2+b_2{}^3)$

特異点解消定理 (広中の定理,1964)



ゼータ関数と自由エネルギー

ゼータ関数の定義

$$\zeta(z) = \int K(w)^z\ \varphi(w)\, dw \quad (\mathrm{Re}(z)>0)$$

(例) $\zeta(z) = \int_0^1 \int_0^1 (a^2 b^4)^z\, da\, db \quad (\mathrm{Re}(z)>0)$

$$= 1/\{(2z+1)(4z+1)\}$$

$$= 2/(4z+1) - 1/(2z+1)$$

定理

$\zeta(z)$ は**C**全体に有理型関数として解析接続できる。

$\zeta(z)$ の極を　$0 > -\lambda_1 > -\lambda_2 > -\lambda_3 >$ ・・・とし、

それぞれの位数を　$m_1 \quad m_2 \quad m_3$　・・・ とするとき、

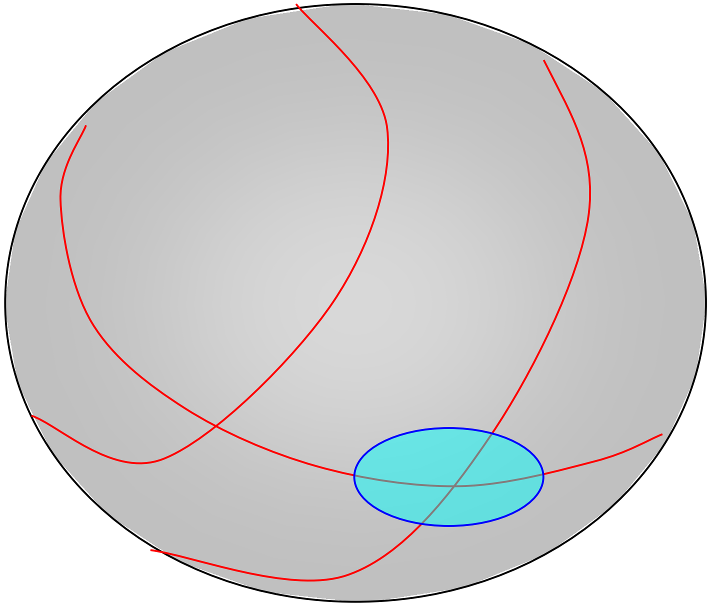
自由エネルギーの漸近展開

$$F(n) = -\log \int e^{-nK(w)} \varphi(w)dw$$

$$= \lambda_1 \log n - (m_1 - 1) \log\log n + \text{定数} + \cdots$$

が成り立つ。

関数 K(w) の分解



各局所座標で

$$K(g(u)) = \prod_{i=1}^{d} u_i^{2k_i}$$

ヤコビアン

$$|g(u)'| = b(u)\prod_{i=1}^{d} |u_i^{h_i}|$$

ゼータ関数の展開

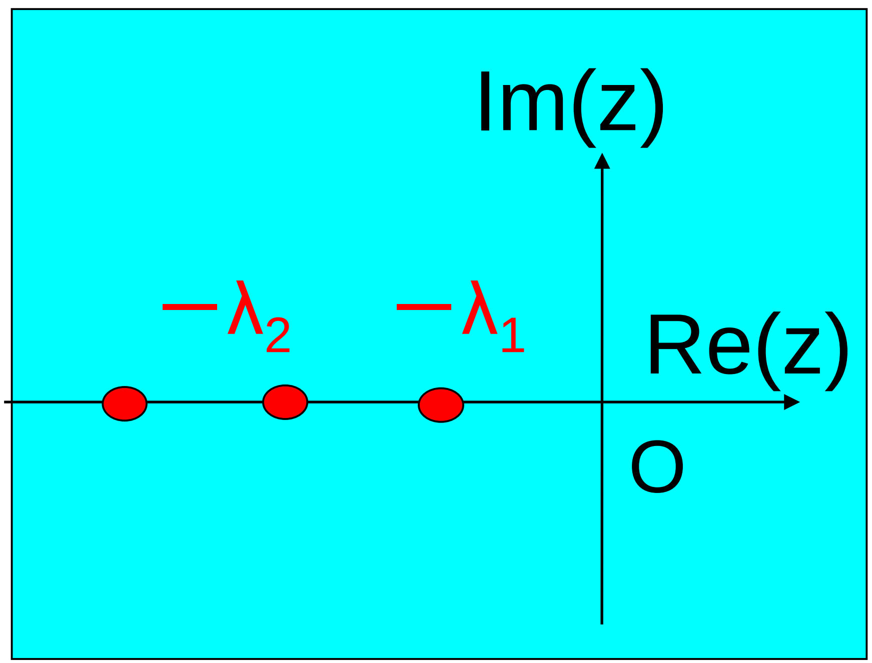
$$\zeta(z) = \int K(w)^z \ \varphi(w)\, dw$$

$$= \sum_{\substack{\text{局所座標} \\ \text{（有限個）}}} \int_{[0,1]^d} K(g(u))^z \ \varphi(g(u))\ |g'(u)|\ du$$

$$= \sum_{\substack{\text{局所座標} \\ \text{（有限個）}}} \int_{[0,1]^d} \Pi\ du_j \ \ u_j^{2k_j z} \ \ u_j^{h_j} \varphi(g(u))\ b(u)$$

ゼータ関数の極

$$\zeta(z) = \frac{C_1}{(z+\lambda_1)^{m_1}} + \frac{C_2}{(z+\lambda_2)^{m_2}} + \cdots$$



ゼータ関数と状態密度関数

$$\zeta(z) = \int K(w)^z \ \varphi(w)\, dw$$

逆 Mellin 変換 ↓ ↑ Mellin 変換

$$\zeta(z) = \int_0^1 v(t)\ t^z\, dt$$

$$v(t) = \int \delta(t-K(w))\ \varphi(w)\, dw$$

解析接続 → 漸近展開

$$\zeta(z) = \sum_{k=1}^{\infty} \sum_{m=1}^{m_k} \frac{C_{km}}{(z+\lambda_k)^m}$$

Re(z)⟶ -∞
に向かう
解析接続

↓

$$v(t) = \sum_{k=1}^{\infty} \sum_{m=1}^{m_k} C'_{km}\, t^{\lambda_k - 1} (-\log t)^{m-1}$$

t ⟶ +0
における
漸近展開

状態密度 → 分配関数

$$Z(n) = \int e^{-nK(w)}\,\varphi(w)\,dw$$

$$= \int dt \int dw\,\delta(t-K(w))\,e^{-nt}\,\varphi(w)$$

$$= \int dt\,e^{-nt}\,v(t)$$

$$= \int dt/n\;\;e^{-t}\,v(t/n)$$

$$= \sum_{k=1}^{\infty}\;\sum_{m=1}^{m_k} C'_{km}\;\int dt\,e^{-t}\,(t/n)^{\lambda_k}\{-\log(t/n)\}^{m-1}$$

$n \to \infty$

における

漸近展開

分配関数 → 自由エネルギー

$$F(n) = -\log Z(n)$$

$$= \lambda_1 \log n - (m_1 - 1)\log\log n + \text{定数} + \cdots$$

自由エネルギーの漸近形の導出法

(1) 真の分布と学習モデルの距離 K(w) を計算

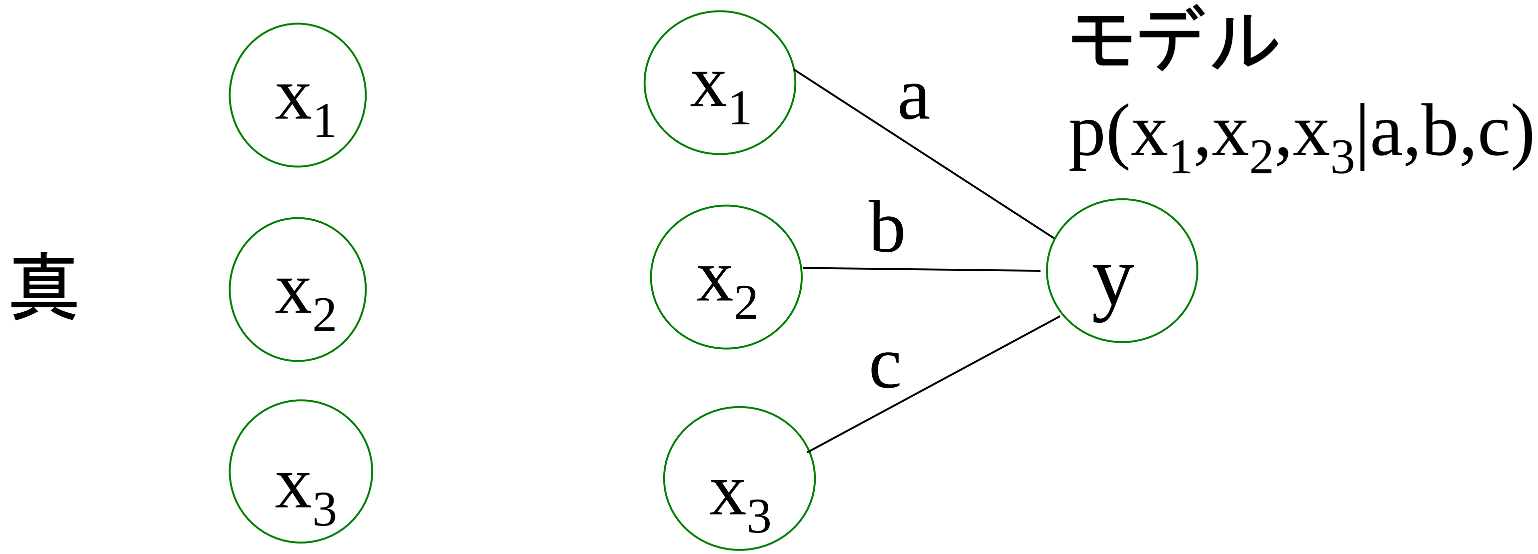
(2) 集合 K(w)=0 の特異点を解消

(3) 実対数閾値と多重度を求める

(4) 上式を用いて自由エネルギーを得る。

例　簡単なボルツマンマシン

真の分布が独立、1個の隠れユニットを持つボルツマンマシンで学習したとき



$$p(x_1,x_2,x_3|a,b,c)=\frac{1}{Z}\sum_{y=\pm 1}\exp(ax_1y+bx_2y+cx_3y)$$

例　$K(a,b,c) = a^2b^2+b^2c^2+c^2a^2 + \cdots$

$$\left.\begin{array}{l} a = a_1 \\ b = a_1 b_1 \\ c = a_1 c_1 \end{array}\right\}$$

ブローアップ

$$K(a_1,b_1,c_1) = a_1^{\,4}(b_1^{\,2}+b_1^{\,2}c_1^{\,2}+c_1^{\,2})$$

$$\left.\begin{array}{l} b_1 = b_2 \\ c_1 = b_2 c_2 \end{array}\right\}$$

ブローアップ

$$K(a_2,b_2,c_2) = a_2^{\,4}\, b_2^{\,2}(1+b_2^{\,4}c_2^{\,2}+c_2^{\,2})$$

→ *正規交差 $\lambda=3/4$, $m=1$*

答　$F(n) = (3/4) \log n$ + 定数$+\cdots$