

# 初めての代数幾何学 ⑤

東京工業大学 渡辺澄夫



# 復習

- ①  $\{V; \text{代数多様体}\} \Leftrightarrow \{I(V); \text{定義イデアル}\}$  は全単射。
- ② グレブナー基底  $\langle \text{LT}(I) \rangle = \langle \text{LT}(f_1), \text{LT}(f_2), \dots, \text{LT}(f_k) \rangle$
- ③  $V$  の特異点は  $I(V)$  のヤコービ行列を使って判定できる。
- ④ 射影空間とブローアップ
- ⑤ 最終回 特異点解消定理と学習理論への応用

# 1 特異点解消定理

# 一般の代数多様体

- 1 アフィン空間の中の代数多様体
  - 2 射影空間の中の代数多様体
  - 3 ブローアップしてできる空間の中の代数多様体
- アフィン空間の代数多様体の貼りあわせでできるものを**一般の代数多様体**という。

**問題** 各局所座標ではアフィン代数多様体であるものを貼り合わせてできるが、できあがった全体のものはどんなぐあいになっているのだろうか？

# 一般のブローアップ

代数多様体  $V \subset \mathbf{R}^n$  とし、 $W = \mathbf{V}(x_1, x_2, \dots, x_r) \subset V$  とする。 $V$  の  $W$  を中心としたブローアップは

---

$$B_W(W) = \{ (x, (x_1 : x_2 : \dots : x_r)) ; (x, y) \in V - W \}$$

これは第  $U_j$  座標で代入

$$x_1 = x_j' x_1'$$

$$x_2 = x_j' x_2'$$

$$x_j = x_j'$$

$$x_r = x_j' x_r$$

を行うことと等価になっている。

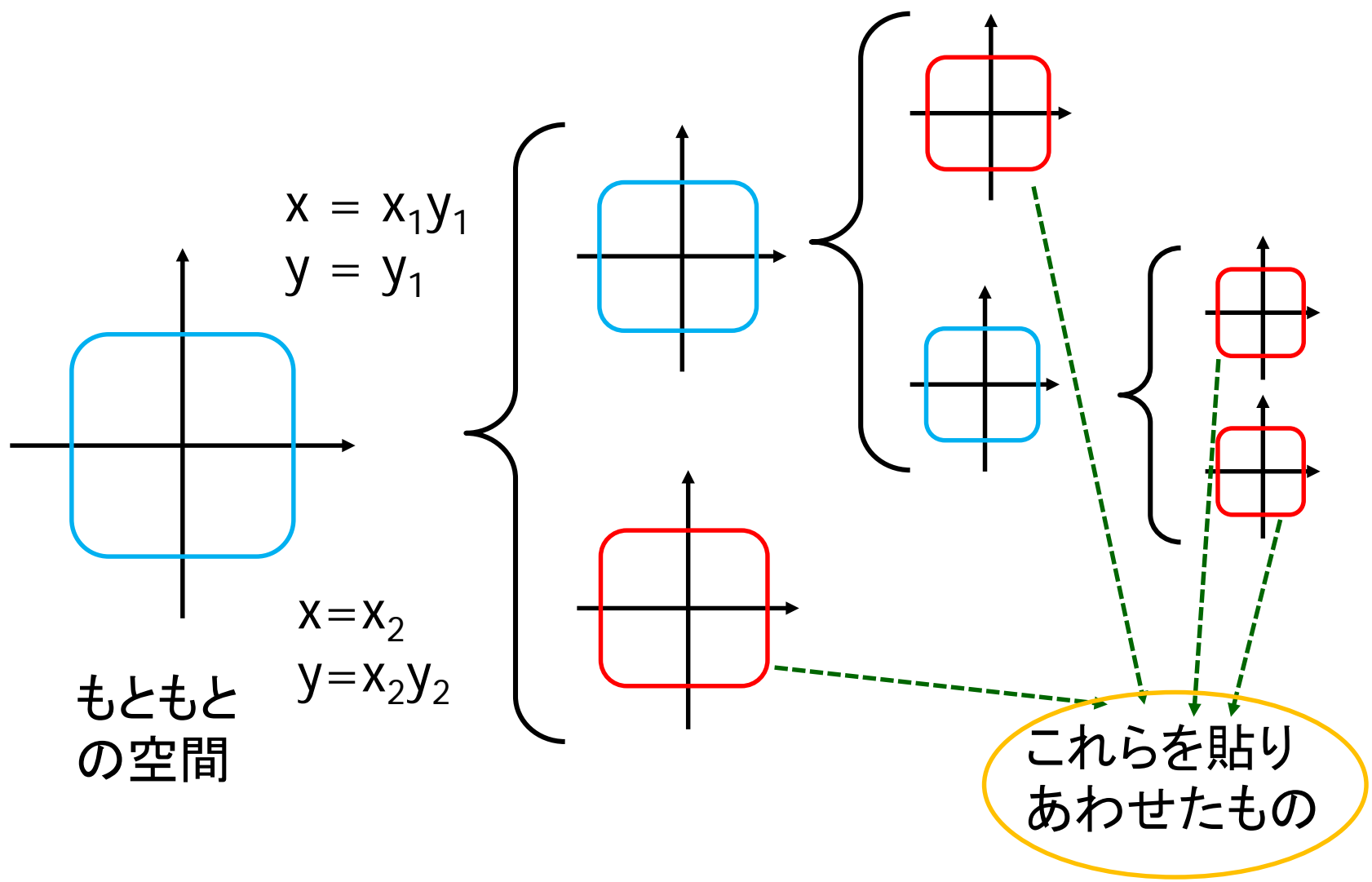
# 特異点解消定理

代数多様体  $V \subset \mathbf{R}^n$  について、 $V$  の特異点の集合  $\text{Sing}(V)$  に含まれる非特異集合  $W$  を見つけて、  
(微分同相な) 変数変換により  $W$  を  $W = \mathbf{V}(x_1, x_2, \dots, x_r)$  に変換して、 $V$  を  $W$  でブローアップし  $B_W(V)$  を求める

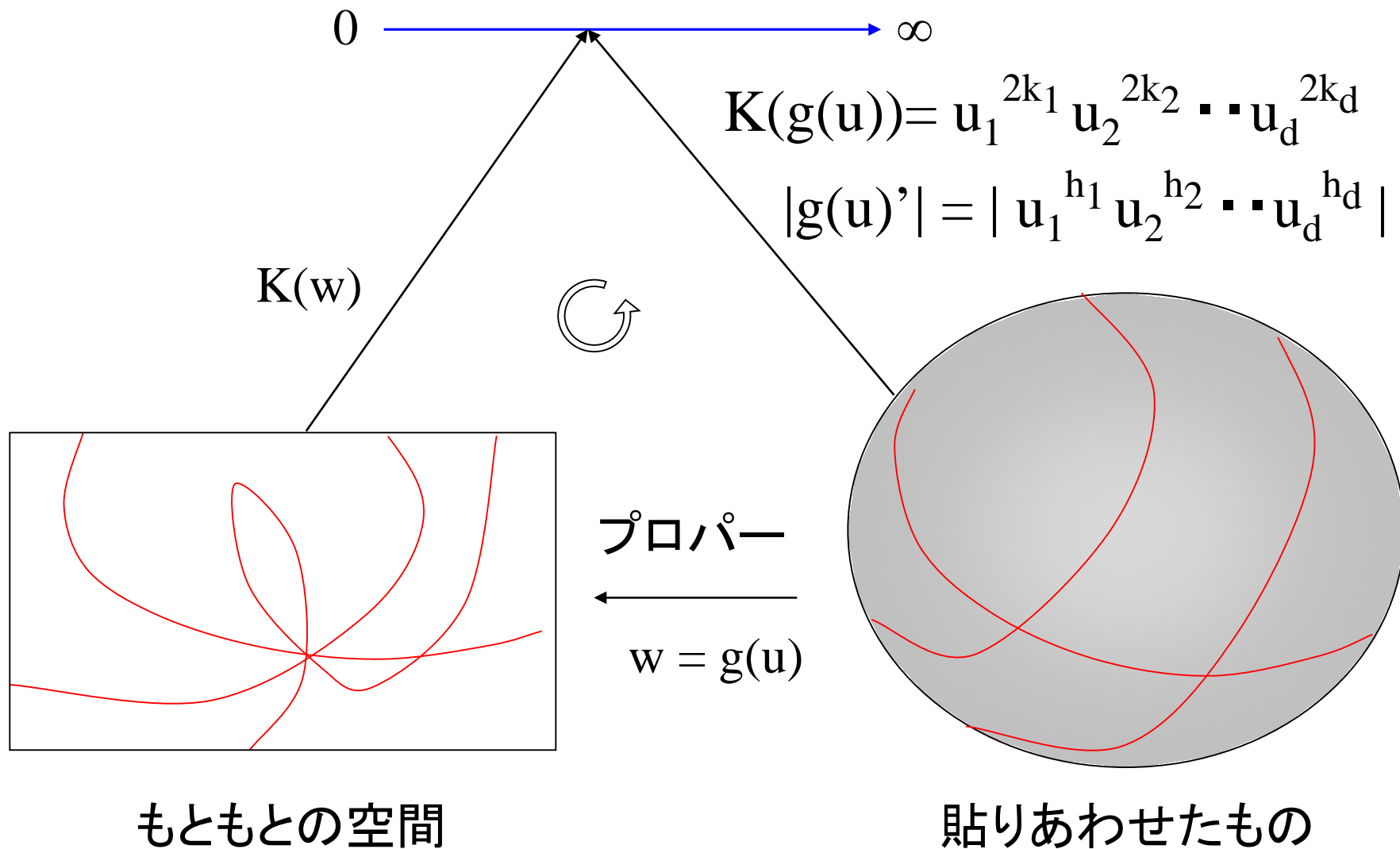
ことを繰り返していくと、その有限回の繰り返しの手続きの中に特異点が解消されるものが存在する。

(注意) ブローアップの定義では、例外集合が取り除かれてしまうが、応用上は例外集合も大切である。

# 特異点が無くなるまでブローアップする



# 特異点解消定理 (広中の定理, 1964)





例  $K(a,b,c) = a^2b^2+b^2c^2+c^2a^2$

$a = a_1$   
 $b = a_1 b_1$   
 $c = a_1 c_1$  } *ブローアップ*  
 $K(a_1, b_1, c_1) = a_1^4 (b_1^2 + b_1^2 c_1^2 + c_1^2)$   
全部で3個のアフィン空間になるが  
対称性から他の座標も同じ

$b_1 = b_2$   
 $c_1 = b_2 c_2$  } *ブローアップ*  
 $K(a_2, b_2, c_2) = a_2^4 b_2^2 (1 + b_2^4 c_2^2 + c_2^2)$   
2個のアフィン空間になるが対称性から  
他の座標も同じ

6個のアフィン空間の貼り合わせで特異点解消できた。

# ヤコービ行列式を計算する

ふたつのブローアップ

$$\begin{array}{ll} a = a_1 & b_1 = b_2 \\ b = a_1 b_1 & c_1 = b_2 c_2 \\ c = a_1 c_1 & \end{array}$$

を合成すると

$$\begin{array}{l} a = a_1 \\ b = a_1 b_2 \\ c = a_1 b_2 c_2 \end{array}$$

ヤコービ行列式は  $|g'(u)| = |a_1^2 b_2|$

## 2 学習理論への応用

# 学習理論とは

パラメータ  $(a_1, \dots, a_k)$  を持つ条件つき確率

$p(y|x, a_1, \dots, a_k)$  がデータ  $\{X_i, Y_i; i=1, 2, \dots, n\}$  から

真の条件つき確率を推測する。

データが多くなれば推測は少しずつ正確に

なっていくが、それはどの程度だろうか。という

問題は、学習モデルの何によって決まるのか。

# 古典的な学習モデル

条件つき確率  $p(y|x, a_1, \dots, a_k)$

$$Y = a_1 + a_2 X + a_3 X^2 + \dots + a_k X^{k-1} + \text{雑音}$$

パラメータとモデルは1対1に対応する

$$\begin{aligned} p(y|x, a_1, \dots, a_k) &= p(y|x, b_1, \dots, b_k) \\ \Leftrightarrow (a_1, \dots, a_k) &= (b_1, \dots, b_k) \end{aligned}$$

$p(y|x, a_1, \dots, a_k)$  と  $p(y|x, b_1, \dots, b_k)$  の違いは局所的に2次式で近似できる。

# 古典的な誤差関数



$(a_1, \dots, a_k)$

# 現代の学習モデル

条件つき確率  $p(y|x, a_1, a_2, b_1, b_2)$

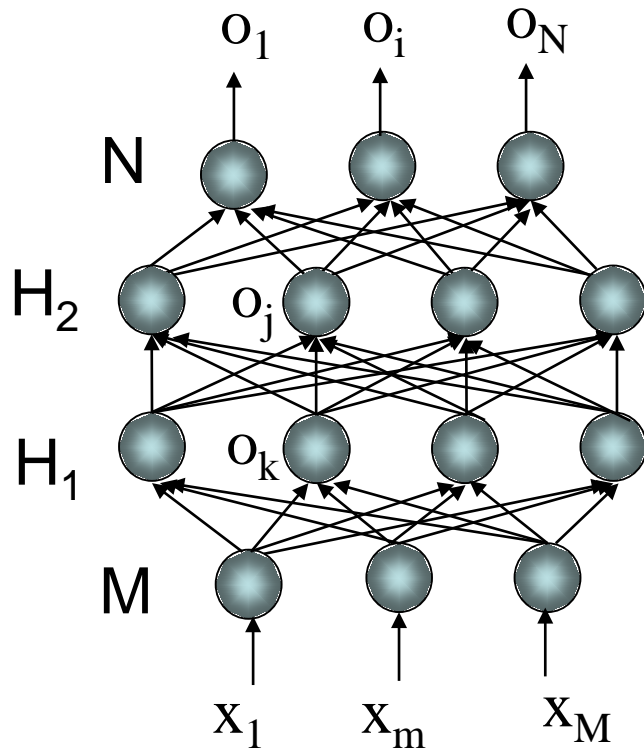
$$Y = a_1 \sin(b_1 x) + a_2 \sin(b_2 x) + \text{雑音}$$

$$p(y|x, a_1, a_2, b_1, b_2) = p(y|x, 0, 0, 0, 0)$$



$$\begin{aligned} & [ \{a_1 = 0\} \cap \{a_2 = 0\} ] \cup [ \{a_1 = 0\} \cap \{b_2 = 0\} ] \\ & \cup [ \{a_2 = 0\} \cap \{a_1 = 0\} ] \cup [ \{b_1 = 0\} \cap \{b_2 = 0\} ] \\ & \cup [ \{a_1 = a_2\} \cap \{b_1 = -b_2\} ] \\ & \cup [ \{a_1 = -a_2\} \cap \{b_1 = b_2\} ] \end{aligned}$$

# 深層学習のネットワーク



中間層2から出力へ

$$o_i = \sigma\left(\sum_{j=1}^{H_2} w_{ij} o_j + \theta_i\right)$$

中間層1から中間層2へ

$$o_j = \sigma\left(\sum_{k=1}^{H_1} w_{jk} o_k + \theta_j\right)$$

入力から中間層1へ

$$o_k = \sigma\left(\sum_{m=1}^M w_{km} x_m + \theta_k\right)$$

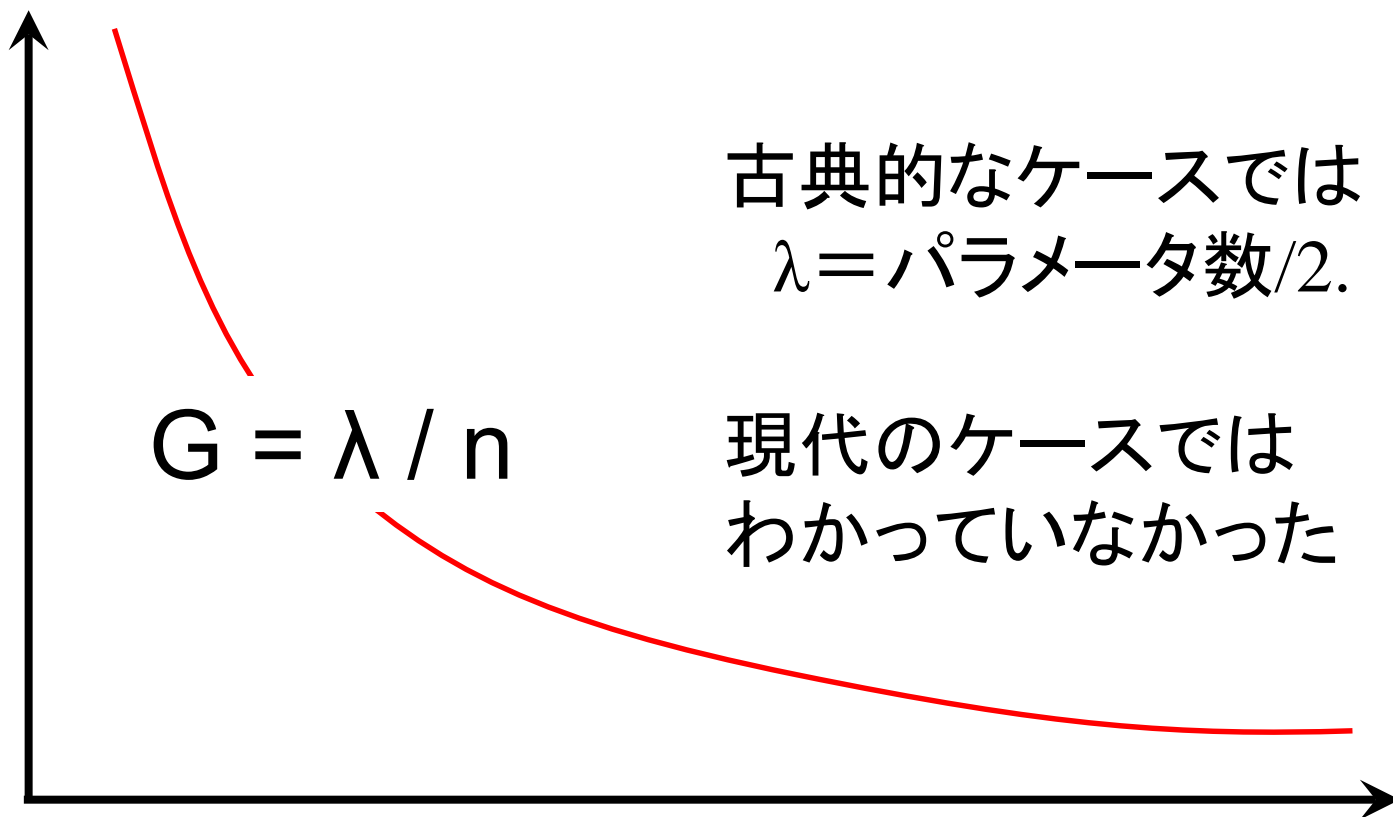
パラメータ  $w_{ij}$ ,  $w_{jk}$ ,  $w_{km}$ ,  $\theta_k$ ,  $\theta_j$ ,  $\theta_i$  がパラメータ

二つの深層学習がほぼ同じ推論でもパラメータはまったく違う



# 学習曲線

予測時における誤差  $G$  の平均



学習したデータの個数  $n$

# 学習曲線

現代的なモデルにおいては

$$K(g(u)) = u_1^{2k_1} u_2^{2k_2} \cdots u_d^{2k_d}$$

$$|g(u)'| = |u_1^{h_1} u_2^{h_2} \cdots u_d^{h_d}|$$

のときには 学習曲線の係数は下記のものになる。

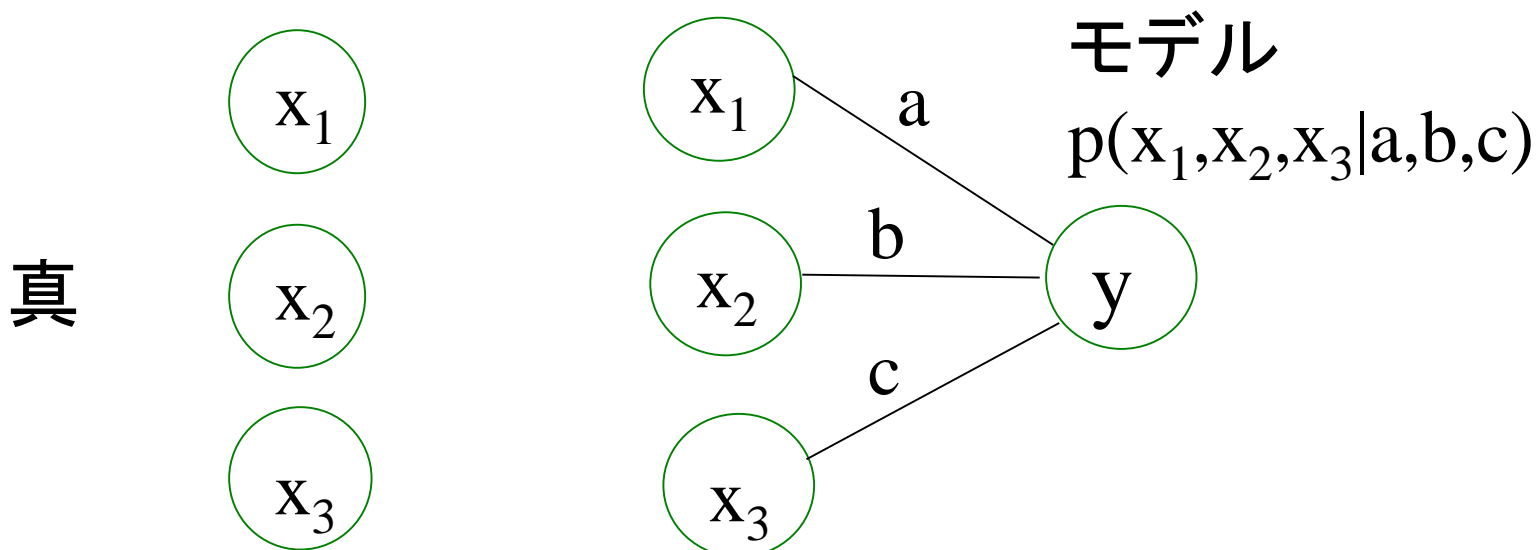
$$\lambda = \min_{j=1}^d [ (h_j+1) / (2k_j) ]$$

特異点解消により 学習曲線を求めることができた。

# 3 具体例

# 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



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

# ボルツマンマシン

$t(a) = \tanh(a)$  とおくと

$$p(x_1, x_2, x_3 | a, b, c)$$

$$= (1/8) \{ 1 + t(a)t(b)x_1x_2 + t(b)t(c)x_2x_3 + t(c)t(a)x_3x_1 \}$$

$$\sum_{x_1, x_2, x_3 = \pm 1} \{ p(x_1, x_2, x_3 | a, b, c) - p(x_1, x_2, x_3 | 0, 0, 0) \}^2$$

$$\propto (t(a)t(b))^2 + (t(b)t(c))^2 + (t(c)t(a))^2$$

$$= a^2b^2 + b^2c^2 + c^2a^2 + \dots$$

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

$$\left. \begin{array}{l} a = a_1 \\ b = a_1 b_1 \\ c = a_1 c_1 \end{array} \right\} K(a_1, b_1, c_1) = a_1^4 (b_1^2 + b_1^2 c_1^2 + c_1^2)$$

$$\left. \begin{array}{l} b_1 = b_2 \\ c_1 = b_2 c_2 \end{array} \right\} K(a_2, b_2, c_2) = a_2^4 b_2^2 (1 + b_2^4 c_2^2 + c_2^2)$$

ヤコービ行列式は  $|g'(u)| = |a_1^2 b_2|$

$$\lambda = \min(3/4, 2/2, 1/0) = 3/4$$

学習曲線は  $G = 3/4n$

# 4 まとめ

# まとめ

- ① 現代の学習モデルは特異点があるため学習曲線がわからなかった。
- ② 特異点解消定理に基づいて学習曲線を求められる。
- ③ ブローアップの繰り返しで特異点を解消しそのときのヤコービ行列式を求めればよい。