

単純階層構造 n 次元図形クラスの 最適 1 ステージ学習アルゴリズム

The optimal 1-stage learning algorithm for the simple hierarchical class of n -dimensional figures

大槻正伸 (OHTSUKI Masanobu) *

平成 15 年 9 月

Abstract

We consider the learning algorithm for the simple hierarchical class of n -dimensional figures that has 1-1 correspondence with an interval. In a word such a class is $C_{[a,b]} = \{c(t); a \leq t \leq b\}$, the set of n -dimensional figures, each member of which is characterized by a parameter t ($a \leq t \leq b$), and $t_1 \leq t_2$ implies $c(t_1) \subset c(t_2)$.

We construct the optimal 1-stage algorithm for learning $C_{[a,b]}$, and show that the error of that algorithm is $(f(b) - f(a))/(2(m+1))$, where $f(t)$ is the n -dimensional volume of $c(t)$ and m is the number of sample points.

1 はじめに

1.1 概念学習とアルゴリズムの誤差

コンピュータに人間と同じように抽象概念の学習を行わせ、そうして獲得した概念をもとに、人間と同じような感覚・判断能力を持たせることを目標に、その基礎理論である「学習理論」が研究されている [1][2]。

その中でも、様々な概念を n 次元空間内の n 次元図形としてとらえ、それをサンプリングにより学習(推定)することにより、人間に近い感覚を持つコンピュータを構成しようという試みがなされている [3][6]。

ここでは「区間と 1 対 1 に対応する単純階層構造をもつ」 n 次元図形の集合(学習理論では図形の集合のことを「概念クラス」「図形のクラス」あるいは単に「クラス」とよぶこともある)の「サンプリングによる学習」について考える。

「区間と 1 対 1 に対応する単純な階層構造をもつ」の意味については次節で厳密に定義し、その性質について調べるが、直観的に言うと、この図形の集合とは、Figure.1 のように、同じような図形が 1 次元的に(年輪やタマネギのような)階層構造を持っているものである。

文献 [7] ではこの特別な場合である n 次元図形の集合 $C_{ncube[a,b]}$ (原点を含む n 次元立方体の集合) について最適な学習アルゴリズムを構成しその誤差を求めた。今回は、その意味で文献 [7] の一般化になっている。

さて、ここで n 次元図形の「サンプリングによる学習」について詳しく述べる。

学習者(学習アルゴリズム) A は、 n 次元空間 $X \subset \mathbf{R}^n$ 内の n 次元図形のクラス C から選ばれた未知の $c \in C$ の形を、サンプリングにより推定する。学習理論では、この「推定」を「学習」と同一視す

*福島工業高等専門学校 電気工学科 (いわき市平上荒川字長尾 30)

る。また、図形 c のことを「概念」、推定すべき図形の集合 C のことを「概念クラス」ともよぶ。また X のことを「学習領域」とよぶ。

A は X, C , および $c \in C$ である事実は知識として持っているが、具体的に学習すべき概念 c がどのようなものかは分からない。

A は、 c の形を推定するために、任意の点 $x \in X$ を選んで、教師機械 (オラクル) \mathcal{O} から x が c の内部の点か外部の点かの情報、すなわち、 $I_c(x)$ を得ることができる。

ただし、 $I_c(\cdot)$ は領域 c 上の特性関数であり、 $I_c(x) = 0 (x \notin c \text{ の場合}), I_c(x) = 1 (x \in c \text{ の場合})$ とする。教師機械 \mathcal{O} は、 A のこのような質問を受けると、1 ステップで $I_c(x)$ の値を A に返す。

A がある点 x について $I_c(x)$ の値を調べることを「サンプリング」という。

A は未知の $c \in C$ を推定するのに、まず「許容サンプル数」 $m \in \mathbf{Z}^+$ を受け取り、 m 回以内の範囲で上記の質問をし、それで得た情報をもとに c を推定し推定結果 $h \in C$ を出力する。

1 ステージアルゴリズムとは、(1) m を入力する。(2) m 個のサンプル点を選んでそれらの点について調べる。(3) (2) の結果をもとに計算をして仮説を出力する。という構造のアルゴリズムである。すなわち、ある点について調べて、その結果により次のサンプル点を決める、というようなことはしない静的サンプリングのアルゴリズムである。

h は「仮説」とよばれるが、当然ながら c に一致するとは限らない。そこで、 h と c の誤差は $vol_n(h\Delta c)$ で計量される。ここで、 $vol_n(\cdot)$ は n 次元の体積、 Δ は対称差であり、 $h\Delta c = (h - c) \cup (c - h)$ である。すなわち、 h と c の誤差は h と c の一致しない部分の大きさで計量される。

m 個のサンプル点 $S_m = [x_1, x_2, \dots, x_m]$ は、 $S_m \in X^m$ としておく。すなわち、 S_m は m 点の組 (m -tuple) であり、重複サンプリングも許すものとする。

上記のことから分かるように、アルゴリズム A は、 $A = (S, H)$ の組で表現される。ここで、 $S = (S_1, S_2, \dots, S_m, \dots)$ である。

また、 $H = (H_1, H_2, \dots, H_m, \dots)$ であり、 $H_m : S_m \times \{0, 1\}^m \rightarrow C$ である。すなわち、サンプル点に関する情報から仮説 h を得る関数である。

S を「サンプリング法」、 H を「仮説関数」とよぶことにする。

このような学習アルゴリズム A は、「1 ステージの確率現象を用いない」アルゴリズムであると言われる [6][7]。

1.2 学習アルゴリズムの誤差

いま、クラス C を学習する 1 ステージアルゴリズム A の誤差、クラス C の誤差を次のように定義する。

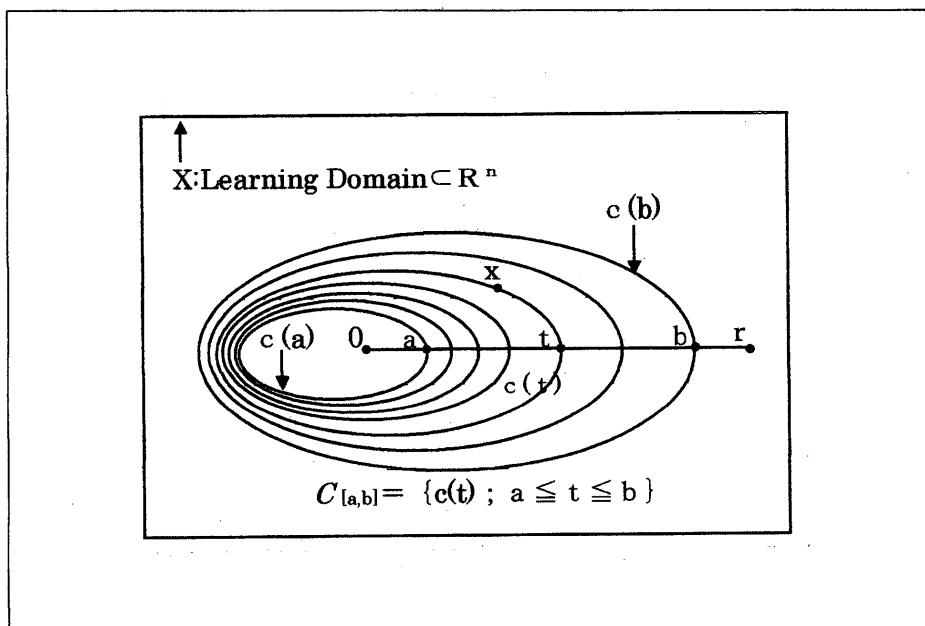
定義 1 (A の誤差、 C の誤差)

$$e_A(m) \stackrel{\text{def}}{=} \sup\{vol_n(A(m, c)\Delta c) ; c \in C\}$$

($A(m, c)$ は、概念 $c \in C$ をアルゴリズム A で、許容サンプル数 m で学習したときの仮説 h のこととする。)

$$e_C(m) \stackrel{\text{def}}{=} \inf\{e_A(m) ; A \text{ は } C \text{ に対する 1 ステージ学習アルゴリズム}\} \square \text{ (定義 1)}$$

$e_A(m)$ はアルゴリズム A で概念クラス C を学習したときの最大 (最悪) 誤差であり、 $e_C(m)$ は、最適 (最良) のアルゴリズムで C を学習したときの最大誤差である。(正確には、これらは「誤差の上限」であるが、今後は「最大誤差」とよぶことにする。) したがって、 $e_C(m)$ は、どのような良いアルゴリズムでこのクラスを学習しても、どうしても避け得ない C の構造の複雑さに起因する量であり、 C の学習しにくさを表すものとなっている。

Figure. 1: A simple hierarchical class of n -dimensional figures ($n = 2$)

一般的な図形の学習を行う際の最適なアルゴリズムを構成すること、およびそのときの誤差 $e_C(m)$ を求めることは非常に困難であり、通常は $e_C(m)$ を上界、下界で評価することになる。

以下、本論文では、区間と 1 対 1 に対応する単純階層構造をもつ n 次元図形のクラス (2 節参照) の、サンプリングによる 1 ステージ学習アルゴリズムについて考える。

上述のように、一般的な n 次元の図形のクラスについて、その最適な学習アルゴリズムを求めることは非常に困難であるが、このような単純な構造をもつ図形については、2 節以下で述べるように少々の解析で、最適な学習アルゴリズムおよびその誤差が求まるのである。

さて、以下 2 では、「区間と 1 対 1 に対応する単純階層構造をもつ図形のクラス」($C_{[a,b]}$) の意味を厳密に定義し、その性質について調べる。

3 で、この構造をもつクラスに対する最適なサンプリング法 $A = (S, H)$ を構成し、 $e_{C_{[a,b]}}(m)$ を正確に求める。

4 では、 $C_{[a,b]} \subset C$ となる C を学習する際の誤差の下界について考察する。ここで示すように、単純な概念クラスの最適アルゴリズムとその誤差が求まれば、そのクラスを含むようなより複雑なクラス C に対する $e_C(m)$ の下界が求まるこのような、単純な階層構造をもつ集合の最適学習アルゴリズムについて考察することは、これを含むより複雑なクラスの誤差の下界を評価する方法を提供することになる、という点で意義があるのである。

2 区間と 1 対 1 に対応する単純な階層構造をもつ n 次元図形のクラス

「区間と 1 対 1 に対応する単純な階層構造をもつ n 次元図形のクラス」とは、自然現象の中で言えば、2 次元では、ある点を中心に広がる波紋、年輪など、3 次元では、タマネギなどのように、同じような図形が単純に階層的な構造をもってすき間なく密集しているものである (Figure.1)。

以下では、この単純な階層構造を数学的に厳密に定義する。

定義 2 (座標づけられた線分 Q_1Q_2) 今 n 次元空間 \mathbf{R}^n 中にある長さ r の線分 $L = Q_1Q_2$ を考える。このうち点 Q_1 を基準にとれば、 L 上の点 x に対して、 Q_1x の長さ t を対応させると、自然に L に座標が

導入され、 L と $[0, r]$ は同一視される。

以下 $t \in [0, r]$ に対応する点を $P(t)$ ということにする。また、しばしば t と $P(t)$ を同一視する。

□ (定義 2)

定義 3 (単純な階層構造をもつクラス) まず、概念クラス C は、閉集合のクラスであることを仮定しておく。すなわち、 $\forall c \in C, c: \text{閉集合}$ 。

C が区間 $[a, b]$ と 1 対 1 対応する単純階層構造をもったクラスである。

$\stackrel{\text{def}}{\iff}$

\exists 長さ r の線分 L (区間 $[0, r]$ と同一視される線分 L) ;

(1) $[a, b] \subset [0, r]$

(2) 区間 $[a, b]$ と C との間には 1 対 1 対応がつく。

$t \in [a, b]$ に対応する概念 $c \in C$ を $c(t)$ と書くことにする。

(3) $\forall t \in [a, b], x = P(t)$ は概念 $c(t) \in C$ の境界上の点となっている。

(4) $a \leq t_1 \leq t_2 \leq b \Rightarrow c(t_1) \subset c(t_2)$

(5) $\forall x \in X$, 次のいずれか一つが成り立つ。

(5-1) $\forall c \in C, x \notin c$

(5-2) $\forall c \in C, x \in c$

(5-3) $\exists 1 c \in C; x \in c$ かつ x は c の境界上の点。 □ (定義 3)

以下では、「区間と 1 対 1 対応する単純階層構造をもったクラス」を単に「単純階層構造をもったクラス」、あるいは「単純階層クラス」ともよぶことにする。

原点を含む n 次元立方体のクラス [7] も単純階層クラスの特別な場合と考えられる。またある点を中心とする半径 a 以上、 b 以下の同心 n 次元球 (球面を含む) の集合なども単純階層クラスの一例である (Figure.2)。

さて、単純階層クラスのサンプリングによる学習を考える上では次の性質 1 が重要になる。性質 1 は、サンプリングをする場合、サンプル点は定義 3 の線分 $L = Q_1Q_2$ 上に限って考えてよいことを保証するものである。すなわち、これは最適なサンプリングを考察する場合、線分 L 上に限って議論しても十分であることを意味する。

性質 1 任意の m 個のサンプル点 x_1, x_2, \dots, x_m に対し、これと同等なサンプル点 z_1, z_2, \dots, z_m が存在する。ただし z_i ($i = 1, 2, \dots, m$) は全て $L = Q_1Q_2$ 上の点である。

直観的には、「同等なサンプル点」 [7] とは、任意の概念 $c \in C$ を学習する場合、どちらのサンプル点について調べても得られる情報は全く同じであることを意味する。

<性質 1 の証明> 任意の点 $x \in X$ を考える。またこの x に対して、次の 2 つの概念クラスを考える。

$C_0(x) \stackrel{\text{def}}{=} \{c \in C; I_c(x) = 0\}, C_1(x) \stackrel{\text{def}}{=} \{c \in C; I_c(x) = 1\}$

さて、定義 3(5) により、次の 3 つの場合が考えられる。(5-1) $\forall c \in C, x \notin c$ の場合。(5-2) $\forall c \in C, x \in c$ の場合。(5-3) $\exists 1 c \in C; x \in c$ かつ x は c の境界上の点である場合。

(5-1) の場合は、どのような $c \in C$ の学習に際しても、点 x について調べることにより何の情報も得られない。なぜなら、 $C_0(x) = C, C_1(x) = \emptyset$ であるから、この点を調べた後でも、概念 c を学習する際には $I_c(x) = 0$ が得られ、まだ全ての $c \in C$ が学習すべき概念の候補として残るからである。

(5-2) の場合も同様に何の情報も得られない。

したがって、(5-1)(5-2) のような \mathbf{x} はサンプル点としては採用せずに省いてもよいことになる。

問題は (5-3) の場合である。

この場合は、点 \mathbf{x} を境界上にもつ唯一の c が存在しその c に対し、定義 3(2) により、 $t \in [a, b]$ が対応する。この t に対して L 上の点 $\mathbf{z} = P(t)$ を考える。さて、そうすると、

$$(1) C_1(\mathbf{x}) = C_1(\mathbf{z}) = \{c(t_1); t \leq t_1 \leq b\}$$

$$(2) C_0(\mathbf{x}) = C_0(\mathbf{z}) = \{c(t_1); a \leq t_1 < t\}$$

となる。なぜなら、

$$(1) t \leq t_1 \implies c(t) \subset c(t_1) \text{ であり、したがって、} \mathbf{x} \in c(t_1), \mathbf{z} \in c(t_1) \text{ となる。}$$

$$(2) t_1 < t \implies c(t_1) \subset c(t) \text{ であり、} \mathbf{x} \notin c(t_1), \mathbf{z} \notin c(t_1) \text{ となる。}$$

なぜなら、 $\mathbf{x} \in c(t_1)$ とすると、まず、 $c(t)$ の境界点 \mathbf{x} は $c(t_1)$ の境界点ではあり得ない。もしも $c(t_1)$ の境界点とすると、それは、定義 3(5-3) により、境界点を含む $c \in C$ のユニーク性に矛盾するからである。

そうすると、点 \mathbf{x} は $c(t_1)$ の内点となる。定義 3(4) より、 $c(t_1) \subset c(t)$ であるから、 \mathbf{x} が $c(t)$ の境界点になることができなくなり矛盾。点 \mathbf{z} についても同様である。

以上のことにより、概念 c を学習する場合、サンプル点 \mathbf{x} をとり、 $I_c(\mathbf{x})$ という情報を得ても、上記の線分 L 上の点 \mathbf{z} をとり $I_c(\mathbf{z})$ という情報を得ても、 $I_c(\mathbf{x}) = I_c(\mathbf{z})$ であり、その値が 0 であっても、1 であっても、それにより選ばれた学習すべき概念 $c \in C$ の候補の範囲は全く同等であることが分かる。

さて、サンプル点が m 個になった場合でも全く同様である。任意の m 個のサンプル $S_1 = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m]$ に対し、上と同様の対応での、線分 L 上の点集合 $S_2 = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m]$ を考える。

$c \in C$ を学習するのに、 S_1 をサンプリングして調べた結果は

$(\langle \mathbf{x}_1, I_c(\mathbf{x}_1) \rangle, \dots, \langle \mathbf{x}_m, I_c(\mathbf{x}_m) \rangle)$ である。

これを、 $(\langle \mathbf{x}_1, I_1 \rangle, \langle \mathbf{x}_2, I_2 \rangle, \dots, \langle \mathbf{x}_m, I_m \rangle)$ と書くことにし、これを満たす $c \in C$ の集合を $C_{(I_1 I_2 \dots I_m)}(S_1)$ とすれば、 $C_{(I_1 I_2 \dots I_m)}(S_1) = C_{I_1}(\mathbf{x}_1) \cap \dots \cap C_{I_m}(\mathbf{x}_m)$ である。

また S_2 をサンプリングして調べても、得られる情報は、 $(\langle \mathbf{z}_1, I_1 \rangle, \langle \mathbf{z}_2, I_2 \rangle, \dots, \langle \mathbf{z}_m, I_m \rangle)$ で、 $C_{(I_1 I_2 \dots I_m)}(S_2) = C_{I_1}(\mathbf{z}_1) \cap \dots \cap C_{I_m}(\mathbf{z}_m)$ であり、 $C_{(I_1 I_2 \dots I_m)}(S_1) = C_{(I_1 I_2 \dots I_m)}(S_2)$ が成り立つ。すなわち、どちらのサンプル点を調べても、それにより選ばれた学習すべき概念 $c \in C$ の候補の範囲は全く同等であることが分かる。

以上のことにより、どのようなサンプル点 $[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m]$ に対しても、上と同様にこれに同等な、線分 L 上のサンプル点 $[\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m]$ をとれることは明らかである。□ (性質 1)

3 単純階層クラスの 1 ステージ最適学習アルゴリズム

ここでは、区間 $[a, b]$ と 1 対 1 に対応する単純な階層構造をもつ n 次元図形のクラスの最適な 1 ステージアルゴリズムを構成する。アルゴリズムを構成する際には、2 節の性質 1 から、サンプル点は線分 L 上の点に限って (より詳しく言えば、区間 $[a, b]$ の点に対応する線分 L 上の点に限って) 議論してもよいことが保証されている。

以下では、 $C_{[a,b]}$ を区間 $[a, b]$ と 1 対 1 対応をもつ単純階層の n 次元図形のクラスとする。2 節の定義 3 により、このようなクラス $C_{[a,b]}$ と区間 $[a, b]$ には 1 対 1 対応がつく。また、 $c(t) \in C_{[a,b]}$ の n 次元の体積を $f(t)$ とおく。すなわち

定義 4 $f(t) \stackrel{def}{=} \text{vol}_n(c(t))$ ($t \in [a, b]$) ただし $\text{vol}_n(\cdot)$ は n 次元体積。□ (定義 4)

そして、以下では f は区間 $[a, b]$ で連続な単調増大関数であることを仮定しておく。
 まず次の2つの補題を用意する。

補題 1 $q_1, q_2, \dots, q_k (q_i \geq 0, i = 1, 2, \dots, k)$ は $q_1 + q_2 + \dots + q_k = Q (Q: \text{定数})$ という制約のもとで自由な変数とする。

このとき、 $\max\{q_1, q_2, \dots, q_k\}$ を最小にするのは、 $q_1 = q_2 = \dots = q_k = \frac{Q}{k}$ のときである。

<証明>

(1) $\max\{q_1, q_2, \dots, q_k\}$ は、 $\frac{Q}{k}$ より小さくなることはできない。

(2) また $q_1 = q_2 = \dots = q_k = \frac{Q}{k}$ とすると、制約条件と (1) の下界を満たす (q_1, q_2, \dots, q_k) が構成できる。□ (補題 1)

補題 2 学習すべき概念 $c \in C_{[a,b]}$ が、 $C_{[a,b]}$ の部分集合 $C_{[a_1,b_1]} \stackrel{\text{def}}{=} \{c(t); t \in [a_1, b_1] \subset [a, b]\}$ のメンバーであるとき、これだけの情報から仮説を出す場合、(すなわち、 $C_{[a_1,b_1]}$ を 0 個のサンプルで学習する場合) 最も (最大) 誤算の少ないものは、

$$h = c(t) = c\left(f^{-1}\left(\frac{f(a_1) + f(b_1)}{2}\right)\right) \text{ である。またそのときの誤差は、} \frac{f(b_1) - f(a_1)}{2} \text{ である。}$$

<証明>

仮説を $c(t) (t \in [a_1, b_1])$ とした場合、(この仮説 $c(t)$ を出力するアルゴリズムを A とすると) 誤差は $e_A(0) = \max\{f(t) - f(a_1), f(b_1) - f(t)\}$ である。

さて、 $f(t) - f(a_1) + f(b_1) - f(t) = f(b_1) - f(a_1) = \text{定数}$ であるから、補題 1 により、

$$f(t) - f(a_1) = f(b_1) - f(t) = \frac{f(b_1) - f(a_1)}{2} \text{ のときに誤差は最小になる。すなわち、}$$

$$f(t) = \frac{f(a_1) + f(b_1)}{2} \text{ より、} t = f^{-1}\left(\frac{f(a_1) + f(b_1)}{2}\right) \text{ の仮説 } c(t) \text{ が最も誤差の少ない仮説となる。}$$

□ (補題 2)

上記補題から、 $C_{[a,b]}$ に対する m 個のサンプル点で学習する場合の最適な 1 ステージアルゴリズムは次のように構成できる。

定理 1 $C_{[a,b]}$ を m 個のサンプル点で (1 ステージで) 学習する場合、次のようにサンプル点、仮説点をとるのが最適である。

サンプル点 = $\{P(t_1), P(t_2), \dots, P(t_m)\}$ 、仮説点 = $\{P(t_{h0}), P(t_{h1}), \dots, P(t_{hm})\}$

$$t_k = f^{-1}\left(\frac{(m-k+1)f(a) + kf(b)}{m+1}\right) (k = 1, 2, \dots, m)$$

$$t_{hk} = f^{-1}\left(\frac{f(t_k) + f(t_{k+1})}{2}\right) = f^{-1}\left(\frac{(2m-2k+1)f(a) + (2k+1)f(b)}{2(m+1)}\right) (k = 0, 1, \dots, m)$$

そして、そのときの誤差は、

$$e_{C_{[a,b]}}(m) = \frac{f(b) - f(a)}{2(m+1)}$$

ただし、仮説点とは、 $I_c(P(t_k)) = 1, I_c(P(t_{k+1})) = 0$ の場合、 $c(t_{hk})$ を仮説とする場合の $t_{hk} \in [t_k, t_{k+1}]$ 、あるいはこの t_{hk} に対する $P(t_{hk})$ のことである。

サンプル点も t_k と $P(t_k) (k = 1, 2, \dots, m)$ を同一視し、 t_k も $P(t_k)$ もサンプル点とよぶことにする。

<証明> $a = t_0 < t_1 < t_2 < \dots < t_m < t_{m+1} = b$ で、 $t_k (k = 1, 2, \dots, m)$ をサンプル点とする。
ある t_k で、 $I_c(P(t_k)) = 1, I_c(P(t_{k+1})) = 0$ であった場合、補題 2 より仮説点は、

$$t_{hk} = f^{-1}\left(\frac{f(t_k) + f(t_{k+1})}{2}\right) \text{ とするのが最適である。}$$

そして、そのときの誤差は $\frac{f(t_{k+1}) - f(t_k)}{2}$ となる。

そうすると、 m 個のサンプル点に対し、次の $m+1$ 個の数（誤差）が対応する。

$$q_1 = \frac{f(t_1) - f(t_0)}{2}, q_2 = \frac{f(t_2) - f(t_1)}{2}, \dots, q_{m+1} = \frac{f(t_{m+1}) - f(t_m)}{2}$$

このアルゴリズム（サンプル点と仮説点）を A とすると、その誤差は、

$e_A(m) = \max\{q_1, q_2, \dots, q_{m+1}\}$ ただし、必ず、 $q_1 + q_2 + \dots + q_{m+1} = \frac{f(b) - f(a)}{2}$
したがって、これを最小にするには、補題 1 より、最小値は $e_{C_{[a,b]}}(m) = \frac{f(b) - f(a)}{2(m+1)}$
またこれを実現するサンプル点は、

$$q_1 = \frac{f(t_1) - f(t_0)}{2} = \frac{f(b) - f(a)}{2(m+1)} \text{ より、}$$

$$f(t_1) = \frac{mf(a) + f(b)}{(m+1)}, \text{ したがって、} t_1 = f^{-1}\left(\frac{mf(a) + f(b)}{(m+1)}\right)$$

以下同様に（正確には k に関する数学的帰納法により）

$$t_k = f^{-1}\left(\frac{(m-k+1)f(a) + kf(b)}{m+1}\right) (k = 1, 2, \dots, m) \text{ が示される。}$$

仮説点は、 $t_{hk} = f^{-1}\left(\frac{f(t_k) + f(t_{k+1})}{2}\right) (k = 0, 1, \dots, m)$ より、

$$t_{hk} = f^{-1}\left(\frac{(2m-2k+1)f(a) + (2k+1)f(b)}{2(m+1)}\right) (k = 0, 1, \dots, m) \text{ が示される。}$$

□（定理 1）

ここで、定理 1 で求めた、最適な 1 ステージ学習アルゴリズムのサンプル点、仮説点集合についての性質を観ておく。この点集合は次の性質をもっている。

性質 2（サンプル点、仮説点に関する性質）

(A)（サンプル点、仮説点の同一性）サンプル点、仮説点の $M_A = (2m+1)$ 個の点集合

$SH = \{t_{h0}, t_1, t_{h1}, t_2, \dots, t_m, t_{hm}\}$ は、 $C_{[a,b]}$ を $M_A = (2m+1)$ 個のサンプルで学習する際の、最適なサンプル点になっている。

(B)（サンプル点の相似性）任意の $k_1 < k_2 (k_1, k_2 = 1, 2, \dots, m)$ に対するサンプル点 t_{k_1}, t_{k_2} と、 $C_{[t_{k_1}, t_{k_2}]} \subset C_{[a,b]}$ を考える。このとき、 $[t_{k_1+1}, t_{k_1+2}, \dots, t_{k_2-1}]$ という $(k_2 - k_1 - 1)$ 個の $(C_{[a,b]}$ を m 個のサンプル点で学習する際の最適な) サンプル点は、 $C_{[t_{k_1}, t_{k_2}]}$ を $M_B = ((k_2 - k_1 - 1))$ 個のサンプル点で学習する際の最適なサンプル点そのものになっている。

<証明 (A)>

$C_{[a,b]}$ を $M_A = (2m+1)$ 個のサンプルで学習する際の最適なサンプル点は定理 1 より、

$$t_K = f^{-1}\left(\frac{(M_A - K + 1)f(a) + Kf(b)}{M_A + 1}\right) (K = 1, 2, \dots, M_A) \text{ すなわち、}$$

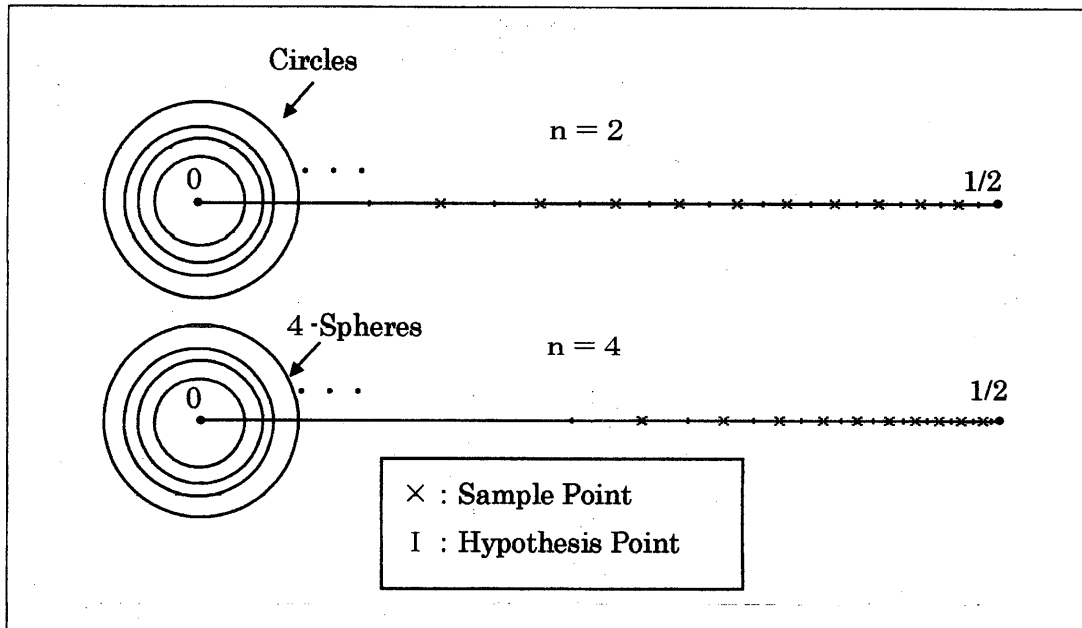


Figure. 2: The examples of optimal sample points (for the class of 2, 4-dimensional spheres)

$$t_K = f^{-1}\left(\frac{(2m+2-K)f(a) + Kf(b)}{2(m+1)}\right) \quad (K = 1, 2, \dots, 2m+1)$$

これは、 $K = 2k$ ($k = 1, 2, \dots, m$) のときには、 $\{t_1, t_2, \dots, t_m\}$ になっており、また、 $K = 2k - 1$ ($k = 1, \dots, m+1$) のときには、 $\{t_{h0}, t_{h1}, \dots, t_{hm}\}$ になっている。

<証明 (B)>

$C_{[t_{k_1}, t_{k_2}]}$ を $M_B = ((k_2 - k_1 - 1))$ 個のサンプル点で学習する際の最良のサンプル点は、定理 1 より

$$t_K = f^{-1}\left(\frac{(M_B - K + 1)f(t_{k_1}) + Kf(t_{k_2})}{M_B + 1}\right) \quad (K = 1, 2, \dots, M_B)$$

$$M_B = k_2 - k_1 - 1, \quad f(t_{k_i}) = \frac{(m - k_i + 1)f(a) + k_i f(b)}{(m + 1)} \quad (i = 1, 2) \text{ を代入して計算すると、}$$

$$t_K = f^{-1}\left(\frac{(m - (k_1 + K) + 1)f(a) + (k_1 + K)f(b)}{(m + 1)}\right) \quad (K = 1, 2, \dots, M_B)$$

となり、 $C_{[a,b]}$ を m 個のサンプル点で学習する際の最適なサンプル点のうち、区間 $[t_{k_1}, t_{k_2}]$ 内のものと一致している。□ (性質 2)

$C_{[a,b]}$ の特別な場合のサンプル点と仮説点の様子をしてみる。

ここでは、 $C_{n\text{Sphere}[0,1/2]}$ (半径 r が $0 \leq r \leq 1/2$ である n 次元球のクラス) を例にする。 $n = 2, 4$ の場合の定理 1 による最適なサンプル点、仮説点は Figure.2 に示すとおりである。

4 誤差の下界についての考察

単純階層クラスの最適 1 ステージアルゴリズムおよびその誤差を調べてきたが、このような単純なクラスについて調べることは次のような意義がある。

1 節でも述べたが、一般に概念クラス C の誤差、 $e_{C(m)}$ を求めるのは非常に困難である。そこで、この上界、下界を求めて誤差を評価することになるが、特にそのうちの下界は求めるのが難しい。

しかし、 $C_1 (C)$ の誤差、 $e_{C_1}(m)$ の下界が求まれば、それは $e_{C_1}(m) \leq e_C(m)$ は明らかであるから、自動的に $e_C(m)$ の下界の一つが求まることになる。

今回は、非常に単純なクラスではあるが、 $e_{C_{[a,b]}}(m)$ が正確に求まったから、これを含む n 次元図形のクラスの誤差の下界が評価できることになる。

例えば、 $[0, 1]^n$ 内の軸平行な n 次元長方形のクラス C_{n1} は、原点を含む n 次元立方体のクラス $C_{ncube[0,1]}$ を部分クラスとして含むから、 $e_{C_{ncube[0,1]}}(m) = \frac{1}{2(m+1)} \leq e_{C_{n1}}(m)$ が示せることになる。

5 結言および今後の課題

本論文では、区間と 1 対 1 に対応する単純階層構造の n 次元図形のクラス $C_{[a,b]}$ の、最適な 1 ステージ学習アルゴリズムおよびその誤差を正確に求めた。

また、この $C_{[a,b]}$ を含むクラスの (1 ステージアルゴリズムの) 誤差の下界について考察した。

今後の課題としては次があげられる。

- ・ より複雑な n 次元図形のクラスの最適 1 ステージアルゴリズム、およびその誤差を求めること。また、その結果により、それを含むさらに複雑なクラスの誤差の下界を求めること。
- ・ 単純階層クラス $C_{[a,b]}$ の多ステージのアルゴリズムについて解析すること。例えば、 m を許容サンプル数としたとき、 m ステージの最適な学習アルゴリズム (すなわち、サンプル点を調べるごとに、その調べた結果をもとに次のサンプル点を選んでよいアルゴリズム) について調べること。

参考文献

- [1] P.D.Laird: Learning from Good and Bad Data. Kluwer Academic Publishers, 1988. (邦訳「例からの学習」横森 貴訳 オーム社 1992 年)
- [2] NATARAJAN, B: MACHINE LEARNING, Morgan Kaufmann Publishers, Inc. San Mateo, California, 1991
- [3] M.Ohtsuki, A.Minato, S.Ozawa: Mathematical Study of Conceptualization in N-Dimensional Space, Forma, 14, pp303-307, 1999.
- [4] M.Ohtsuki, A.Minato, S.Ozawa: A Good Sampling Method for Guessing Rectangles in $[0, 1]^2$, Forma, 15, pp339-345, 2000.
- [5] 大槻 正伸: 高次元長方形の効率よい学習アルゴリズム, 福島工業高等専門学校紀要第 40 号, pp30-35, 2001 年,
- [6] 大槻 正伸: 近似問題と高次元ポリゴンの学習およびその応用に関する研究, 博士学位論文, 茨城大学, 2001 年,
- [7] 大槻 正伸: 原点を含む n 次元立方体のクラスの最適 1 ステージ学習アルゴリズム, 福島工業高等専門学校紀要第 43 号, pp16-24, 2003 年,