

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

The optimal 1-stage learning algorithm
for the class of n -dimensional cubes that include the origin.

(平成 14 年 9 月受理)

大槻 正伸* (OHTSUKI Masanobu)

Abstract

We construct the optimal 1-stage algorithm for learning $\mathcal{C}_{ncube[a,b]}$, the set of n -dimensional cubes $[0, r]^n$ such that $a \leq r \leq b$.

That is, we construct the optimal method of sampling and of calculating the hypothesis to learn that class, and show that the error of that optimal method is $(b^n - a^n)/2(m+1)$, where m is the number of sample points.

1 はじめに

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])。

今回は非常に単純な n 次元図形の集合 $\mathcal{C}_{ncube[a,b]}$ （定義は後出）をある意味で最適に学習するサンプリング法を構成し、そのときの誤差を求める。

まず、 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。

これは、学習者（学習アルゴリズム） \mathcal{A} が、 n 次元空間 $X \subset \mathbf{R}^n$ 内の n 次元図形の族 \mathcal{C} から選ばれた未知の $c \in \mathcal{C}$ の形を、サンプリングにより推定することである。学習理論では、この「推定」を「学習」と同一視する。また、 c のことを「概念」、推定すべき図形の集合 \mathcal{C} のことを「概念クラス」ともよぶ。

\mathcal{A} は $c \in \mathcal{C}$ であることは知識として持っているが、具体的に c がどのようなものかは分からない。

\mathcal{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、 $I_c(x) = 0$ ($x \notin c$ の場合)、 $I_c(x) = 1$ ($x \in c$ の場合) とする。

教師機械 \mathcal{O} は、 \mathcal{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 \mathcal{A} に返す。

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

\mathcal{A} は未知の $c \in \mathcal{C}$ 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 $h \in \mathcal{C}$ を出力する。

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。

さて、一般に \mathcal{A} は「多ステージ」あることが許される。「ステージ」とは、「いくつかの点のサンプリングと一連の計算」の動作を意味する。 \mathcal{A} は「いくつかの点のサンプリングをして、それで得た情報をもとに次にサンプリングすべき点を計算し、」次のステージで「今計算によって求めたサンプル点についてサンプリングし...」という構造のアルゴリズムも当然許される。

* 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）

大槻：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

また、 \mathcal{A} は一般にはサンプル点を決定するのに確率現象を用いることが許される。すなわち、モンテカルロ法のように、 X 上の点をランダムに選び、サンプル点とすることも許される。

今回は1ステージの確率現象を用いないアルゴリズムのみ考えることとする。

そうすると \mathcal{A} は

アルゴリズム \mathcal{A}

begin

 read (m);

 for $i:=1$ to m do

x_i を決定する。;

 for $i:=1$ to m do

$v_i := I_c(x_i)$;

 仮説 h の計算;

 write (h)

end.

という、単純な構造をしたアルゴリズムになる。

そうすると、1ステージで確率現象を用いないアルゴリズムでは、許容サンプル数が m のときのサンプル点 $S_m = [x_1, x_2, \dots, x_m]$ は学習すべき概念 $c \in \mathcal{C}$ によらず一定で、 m にのみ依存する。

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 は順序のついた m 組 (m -tuple) であり、重複サンプリングも許すものとする。

したがって、1ステージで確率現象を用いないアルゴリズム \mathcal{A} は、本質的に $\mathcal{A} = (S, H)$ の組で表現される。ここで、 $S = (S_1, S_2, \dots, S_m, \dots)$ で S_m は上記と同じ m 個のサンプル点の m -tuple である。また、 $H = (H_1, H_2, \dots, H_m, \dots)$ であり、 $H_m: S_m \times \{0,1\}^m \rightarrow \mathcal{C}$ 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。

以下では、学習アルゴリズム \mathcal{A} という場合、すべて1ステージの確率現象を用いないアルゴリズムであることを仮定する。

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

[定義 1] (\mathcal{A} の誤差、 \mathcal{C} の誤差)

$$e_{\mathcal{A}}(m) \stackrel{\text{def}}{=} \sup \{ \text{vol}_n(\mathcal{A}(m, c), c); c \in \mathcal{C} \}$$

($\mathcal{A}(m, c)$ は、概念 $c \in \mathcal{C}$ をアルゴリズム \mathcal{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。)

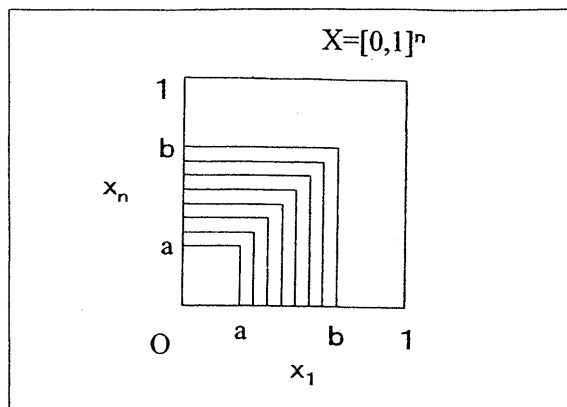
$$e_{\mathcal{C}}(m) \stackrel{\text{def}}{=} \inf \{ e_{\mathcal{A}}(m); \mathcal{A} \text{ は } \mathcal{C} \text{ に対する学習アルゴリズム} \}$$

$e_{\mathcal{A}}(m)$ は \mathcal{A} で概念クラス \mathcal{C} を学習したときの最大 (最悪) 誤差であり、 $e_{\mathcal{C}}(m)$ は、最良のアルゴリズムで \mathcal{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、 $e_{\mathcal{C}}(m)$ 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 \mathcal{C} の構造の複雑さに起因する量であり、 \mathcal{C} の学習しにくさを表すものとなっている。

通常は、 $e_{\mathcal{C}}(m)$ は明確にはならない。

例えば、これまで、 $X = [0,1]^n$ 内の軸平行な n 次元長方形 ($\mathcal{C} = \mathcal{C}_{n1} = \{c = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n]; 0 \leq a_i \leq b_i \leq 1, i = 1, 2, \dots, n\}$) の学習については、1ステージのもの、2ステージのもの、確率現象を用いたものなど様々なアルゴリズムが構成されているが ([3], [4], [5], [6])、このような比較的簡単な構造をした概念クラスでさえ、 $e_{\mathcal{C}}(m)$ 、またその最適サンプリング法などは明確になっていない。

$e_{\mathcal{C}}(m)$ は「どのようなアルゴリズムをもってきてもこれ以上の誤差が出ること、かつ、適当なアルゴリズム

図1: $C_{ncube[a,b]}$ の例 ($n=2$)

をもってこれ以内の最大誤差で学習できる」ことを示さなくてはならないため、これを求めるのが非常に困難だからである。

普通は、効率のよい学習アルゴリズム \mathcal{A} を構成することにより、 $(e_C(m) \leq e_{\mathcal{A}}(m))$ であるから $e_C(m)$ の上界を求め、また数学的議論により、その下界を求め、できるだけシャープに $e_C(m)$ を評価することになる。

さて、本論文の目的は C_{n1} の非常に単純な部分クラス $C_{ncube[a,b]} \stackrel{\text{def}}{=} \{[0,r]^n; a \leq r \leq b\}$ について、最適な1ステージの確率現象を用いないアルゴリズムを構成すること、すなわち最適なサンプリング法、仮説計算法を与えることである。 $(C_{ncube[a,b]}$ は原点を含む n 次元立方体の集合という非常に単純なクラスである (図1))。

またそうすると、同時に $e_{C_{ncube[a,b]}}(m)$ が明確になる。

ここまで単純なクラスになると、少々の解析により最適なサンプリング法、および、 $e_C(m)$ が求められるということになる。

以下、2では、サンプリングの同等性について述べる。これは、3で、 $C_{ncube[a,b]}$ に対する最適なサンプリング法を構成するが、構成する際の単純化に用いられる。

3では、最適な $\mathcal{A} = (S, H)$ を構成し、

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

であることを示す。

2 サンプリング法の同等性

n 次元空間 $X \subset \mathbf{R}^n$ 中の概念クラス $C \subset 2^X$ を考える。 C を学習するためのサンプリング法 $S = (S_1, S_2, \dots, S_m, \dots)$ と $S' = (S'_1, S'_2, \dots, S'_m, \dots)$ が同等であるとは次のこととする。

[定義2] (サンプリングの同等性) S と S' が同等 $\stackrel{\text{def}}{\iff}$

$$\begin{aligned} & \forall m \in \mathbf{Z}^+, \exists f_m: X^m \rightarrow X^m, \exists \sigma_m: \{1, 2, \dots, m\} \rightarrow \{1, 2, \dots, m\} (\sigma_m: 1-1, \text{上への写像 (i.e. } \sigma_m \text{ は置換)}); \\ & f_m(x_1, \dots, x_m) = (f_m^1(x_1), \dots, f_m^m(x_m)) = [x'_{\sigma(1)}, \dots, x'_{\sigma(m)}] = \sigma_m(S'_m) \text{ であり、かつ} \\ & \forall c \in C, I_c(x_i) = I_c(f_m^i(x_i)) (i = 1, 2, \dots, m) \end{aligned}$$

定義より、 $(\forall m \in \mathbf{Z}^+, f_m^i(x_i) = x_i)$ とすれば順序が異なり同じ m 個の点からなる m -tuple (for $\forall m$) のサンプリング法は互いに同等である。

また、サンプリングの同等性を直観的に理解するために、次の例をあげる。

まず、 $c(r)$ を $P(0,0)$ を中心とする半径 r の円とする。

$X = c(1)$ とし、 $C_{2Sphere} = \{c(r); c(r) (0 \leq r \leq 1)\}$ を考える (図2)。

いま $P(0,0)$ 始点とする半直線 L を考え、サンプル点の集合 $S_m = [x_1, \dots, x_m]$ の $x_i (i = 1, 2, \dots, m)$ に対し、 x_i をその円周に持つ、 P を中心とする円と L の交点を $x'_i = f_m^i(x_i)$ とする写像 f_m をとる。

そうすると、 $\forall c \in C_{2Sphere}$ に対し、 $I_c(x_i) = I_c(x'_i)$ が成り立ち、これにより同等のサンプリング法が得られ

大槻：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

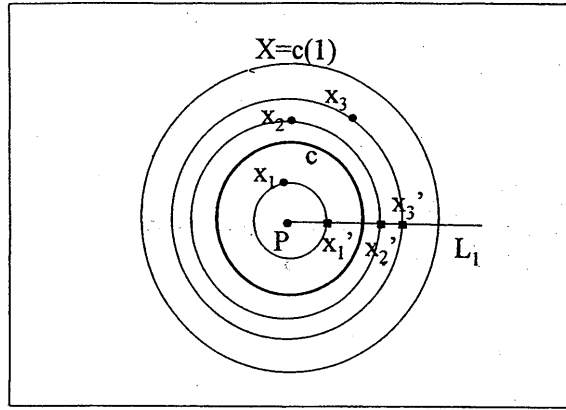


図2： $C_{2Sphere}$

る。(この場合 σ_3 は恒等置換とする。)

さて、図2の c を学習する際には、どちらのサンプリングでも同じ情報しか得られない。例えば図の場合、 $I_c(x_1) = 1, I_c(x_2) = 0, I_c(x_3) = 0$ という情報からも、 $I_c(x'_1) = 1, I_c(x'_2) = 0, I_c(x'_3) = 0$ という情報からも、学習すべき $c(r)$ は、 $\overline{Px_1} \leq r \leq \overline{Px_2}$ ということしか分からないからである。

このように、同等なサンプリング法とは、どちらのサンプリング法で点を調べても、未知の $c \in C$ を推定するための情報としては全く同じということである。

$S \neq S'$ かつ、これらが同等なサンプリングである、というようなことが起こりうるのは、 C が非常に単純な構造をしている場合のみである。

例えば、 $C_{21}([0,1]^2$ 内の長方形の集合)という比較的単純な構造のクラスでも、 $S \neq S'$ かつ同等なサンプリング法というものは存在し得ない。

さて、同等なサンプリング法が存在した場合、それらは、学習の上では全く同じ情報しか与えないということはほとんど直観的に明らかではあるが、ここでこの事実を証明しておく。

[補題1] S と S' は同等なサンプリング法とする。 $\mathcal{A} = (S, H)$ を C の学習アルゴリズムとする。このとき、 $\exists \mathcal{A}' = (S', H'); e_{\mathcal{A}'}(m) = e_{\mathcal{A}}(m)$

<証明> $S_m = [x_1, x_2, \dots, x_m], S'_m = [x'_1, x'_2, \dots, x'_m] (f'_m(x_i) = x'_i)$ とし、

$H = (H_1, H_2, \dots, H_m, \dots)$ で $H_m = H_m(x_1, \dots, x_m, I_c(x_1), \dots, I_c(x_m)) \in C$ とする。

$\mathcal{A}' = (S', H')$ の $H' = (H'_1, H'_2, \dots, H'_m, \dots)$ を次のようにする。

$H'_m = H'_m(x'_1, \dots, x'_m, I_c(x'_1), \dots, I_c(x'_m)) = H_m(x_1, \dots, x_m, I_c(x_1), \dots, I_c(x_m))$ とすれば、どのような $c \in C$ を学習するときでも、 \mathcal{A}' は \mathcal{A} と全く同じ仮説を出力する。

これにより、 \mathcal{A} と全く同一の誤差の学習アルゴリズム \mathcal{A}' が得られる。ここでは、サンプリング法と、それから得られる情報から仮説を出力する関数関係にのみ着目し、 x'_i から x_i の対応の計算や H_m, H'_m の計算の計算量は考慮しないことに注意する。□ (補題1)

次に、 $C_{ncube[a,b]}$ というクラスは非常に単純な構造であり、任意のサンプリング法 S に対して、 $S \neq S'$ かつ S に同等なサンプリング法 S' が存在することを示す。

[補題2] $C_{ncube[a,b]}$ を学習するための、任意のサンプリング法、 $S = (S_1, S_2, \dots, S_m, \dots)$ を考える。

n 次元空間内の座標軸 x_1 上の区間 $[0,1]$ を L_1 とする。すなわち、

$L_1 = \{x = (x_1, 0, 0, \dots, 0) : 0 \leq x_1 \leq 1\}$ とする。

このとき、

$\forall S, \exists S' = (S'_1, S'_2, \dots, S'_m, \dots);$

- (1) $S'_m = [x'_1, \dots, x'_m]$ とすると、 $x'_i \in L_1 (i = 1, 2, \dots, m \text{ for } \forall m)$
 (2) S と S' は同等なサンプリング法

<証明> 任意の $m \in \mathbf{Z}^+$ をとり、 $S_m = [x_1, \dots, x_m]$

$x_i = (x_{i1}, x_{i2}, \dots, x_{in}) (i = 1, 2, \dots, m)$ とする。

いま、 $f_m^i(x_i) = (\max\{x_{i1}, x_{i2}, \dots, x_{in}\}, 0, 0, \dots, 0) \in L_1 = x'_i$ とすると、

$\forall c = [0, r]^n \in \mathcal{C}_{ncube[a,b]}$ (i.e. $\forall c = [0, r]^n, a \leq r \leq b$),

$$(1) I_c(x_i) = 1 \Rightarrow \forall j (j = 1, 2, \dots, n), 0 \leq x_{ij} \leq r$$

$$\Rightarrow 0 \leq \max\{x_{i1}, x_{i2}, \dots, x_{in}\} \leq r$$

$$\Rightarrow x'_i \in c = [0, r]^n \Rightarrow I_c(x'_i) = 1$$

$$(2) I_c(x_i) = 0 \Rightarrow \exists j (j = 1, 2, \dots, n); r < x_{ij}$$

$$\Rightarrow r \leq \max\{x_{i1}, x_{i2}, \dots, x_{in}\}$$

$$\Rightarrow x'_i \notin c = [0, r]^n \Rightarrow I_c(x'_i) = 0$$

したがって、 $\forall c \in \mathcal{C}_{ncube[a,b]}, I_c(x_i) = I_c(f_m^i(x_i)) (i = 1, 2, \dots, m)$ □ (補題2)

補題1、補題2により、 $\mathcal{C}_{ncube[a,b]}$ の最良の学習アルゴリズムを考える際には、サンプル点を L_1 上、すなわち、一つの座標軸上のサンプリング法に限って考えてもよいことが保証される。

3 $\mathcal{C}_{ncube[a,b]}$ に対する最適な1ステージアルゴリズムの構成

次元 n が大きくなると、 n 次元図形の質量は表面に集中する。なぜなら、 $\forall \Delta > 0$ に対し、

$$vol_n([0, r]^n) = r^n, vol_n([0, r-\Delta]^n) = (r-\Delta)^n \text{ から、}$$

$$\lim_{n \rightarrow \infty} \frac{(r-\Delta)^n}{r^n} = \lim_{n \rightarrow \infty} \left(1 - \frac{\Delta}{r}\right)^n = 0$$

$$\lim_{n \rightarrow \infty} \frac{r^n - (r-\Delta)^n}{r^n} = 1$$

が言えるからである。

このことにより、 $\mathcal{C}_{ncube[a,b]}$ の学習を効率よく行うには、 $[0, 1]^n$ 内に均等にサンプル点をとるよりは、 $c = [0, r]^n$ の r の少しの違いで誤差が大きくなる $[0, 1]^n$ の表面付近を細かく調べた方がよいということが直観的に分かる。

以下で、最適アルゴリズム $\mathcal{A} = (S, H)$ を構成する。

まず次の2つの補題を用意する。

[補題3] 区間 $[a, b]$ 上で定義される連続関数 $f(x), g(x)$ があり、 $f(x)$ は単調増大、 $g(x)$ は単調減少とする。かつ、点 $x_0 (a \leq x_0 \leq b)$ で $f(x_0) = g(x_0)$ とする。

この区間 $[a, b]$ で、 $\max\{f(x), g(x)\}$ を最小にする $x (a \leq x \leq b)$ は、 $x = x_0$ である。

<証明> ほとんど明らかである。□ (補題3)

[補題4] 学習すべき概念が $c = [0, r]^n \in \mathcal{C}_{ncube[a,b]}$ である場合、ここでサンプリングせずに (i.e. $m = 0$ 個のサンプリングで) 仮説を出すとすれば、最適な仮説は

$$h = [0, \left(\frac{a^n + b^n}{2}\right)^{1/n}]^n \text{ である。}$$

<証明> $h = [0, y]^n$ とする。

$c = [0, r]^n$ が $\mathcal{C}_{ncube[a,b]}$ を動いたとき、それを学習する場合の最大誤差を最小にすればよい。

すなわち、 $\sup\{vol_n(h\Delta c); c \in \mathcal{C}_{ncube[a,b]}\}$ を最小にする y を求めればよい。

大槻：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

(1) $a \leq y \leq b$ の場合 学習すべき $c = [0, r]^n$ は、 $a \leq r \leq b$ をとり得るが、

$$a \leq r \leq y \Rightarrow \text{vol}_n(h\Delta c) = y^n - r^n$$

$$y \leq r \leq b \Rightarrow \text{vol}_n(h\Delta c) = r^n - y^n$$

したがって、 y を固定したときの最大誤差は $\max \{y^n - a^n, b^n - y^n\}$ となる。

上記の補題3より、最大誤差を最小にする y は、 $y^n - a^n = b^n - y^n$ の関係から求められる。

$$\text{これより、} y = \left(\frac{a^n + b^n}{2}\right)^{1/n} \text{ となる。}$$

(2) $y < a$ あるいは $b < y$ の場合 すでに $c = [0, r]^n$ で、 $a \leq r \leq b$ であることは分かっているから、仮説 $h = [0, y]^n$ で $y < a$ あるいは $b < y$ とすることはいわゆる「仮定に矛盾した仮説」である。しかし、一般的には矛盾した仮説を出力することも許される。

しかし、 $y < a$ の場合、最大誤差は $b^n - y^n > b^n - a^n$ であり、 $b < y$ の場合は最大誤差は $y^n - a^n > b^n - a^n$ である。

$a \leq y \leq b$ では $\max \{y^n - a^n, b^n - y^n\} \leq (b^n - a^n)$ であるから、

いずれも、上記(1)の最大誤差を超えてしまう。

当然の結果ではあるが、最大誤差を最小にする際には、仮定に矛盾した仮説は考えなくてもよい。

(1)(2)より、 $y = \left(\frac{a^n + b^n}{2}\right)^{1/n}$ の仮説が最大誤差を最小にし、そのときの誤差は、 $\frac{b^n - a^n}{2}$ となる。

<注>これはいわば、 $m = 0$ の場合であり、 $e_{C_{\text{ncube}[a,b]}}(0) = \frac{b^n - a^n}{2}$ であることが示されたことになる。

3.1 $m = 1$ の場合

上記の補題4は、いわば、 $m = 0$ の場合の最適なアルゴリズムであった。ここでは $m = 1$ の場合の最適なアルゴリズム、すなわち、 $m = 1$ のときのサンプリング法と仮説計算法を構成する。

$m = 1$ 個のサンプル点 $x_1 = (z, 0, \dots, 0)$ をとったものとする。しばらく z を固定して考える。

この z であるが、補題4(2)と同様に $a \leq z \leq b$ の場合のみ考えれば十分である。(以下、無駄な矛盾するサンプル点、仮説については、煩わしさをなくするため特に言及しないこととする。)

補題4により、 $I_c(x) = 0$ の場合、

$$I_c(x) = 0 \text{ の場合、} h = \left[0, \left(\frac{a^n + z^n}{2}\right)^{1/n}\right]^n \text{ とし、}$$

$$I_c(x) = 1 \text{ の場合、} h = \left[0, \left(\frac{z^n + b^n}{2}\right)^{1/n}\right]^n \text{ とするのが、最適な仮説である。}$$

このアルゴリズムの \mathcal{A} の最大誤差は、

$$\max \left\{ \frac{z^n - a^n}{2}, \frac{b^n - z^n}{2} \right\} \text{ である。}$$

これを最小にする z は、補題3により、

$$\frac{z^n - a^n}{2} = \frac{b^n - z^n}{2} \text{ より、} z = \left(\frac{a^n + b^n}{2}\right)^{1/n} \text{ である。}$$

以上をまとめると、 $m = 1$ のとき、 $C_{\text{ncube}[a,b]}$ を学習する最良のアルゴリズムは、

サンプル点 $x_1 = \left(\left(\frac{a^n + b^n}{2}\right)^{1/n}, 0, \dots, 0\right)$ で、仮説計算法は、補題4より、

$$I_c(x_1) = 0 \text{ のときは } h = \left[0, \left(\frac{3a^n + b^n}{4}\right)^{1/n}\right]^n$$

$$I_c(x_1) = 1 \text{ のときは } h = \left[0, \left(\frac{a^n + 3b^n}{4}\right)^{1/n}\right]^n$$

そのときの最大誤差 $e_{\mathcal{A}}(1)$ は $e_{\mathcal{A}}(1) = \frac{b^n - a^n}{4}$ となる。

これはすなわち、 $e_{C_{ncube[a,b]}}(1) = \frac{b^n - a^n}{4}$ であることが示されたことになる。

3.2 $m = 2$ の場合

サンプル点を $\mathbf{x}_1 = (z_1, 0, \dots, 0)$ 、 $\mathbf{x}_2 = (z_2, 0, \dots, 0)$ ($z_1 < z_2$) とする。

いま、 z_1 を固定しておく。補題4により、

(1) $I_c(\mathbf{x}_1) = 0$ のとき、 $h = [0, (\frac{a^n + z_1^n}{2})^{1/n}]^n$ が最良の仮説計算法であり、

そのときの最大誤差は、 $\frac{z_1^n - a^n}{2}$ である。

(2) $I_c(\mathbf{x}_1) = 1$ の場合、 $C_{ncube[z_1, b]}$ の学習となり、 $m = 1$ の場合に帰着される。そしてこの場合、

$z_2 = (\frac{z_1^n + b^n}{2})^{1/n}$ として、

(2-1) $I_c(\mathbf{x}_1) = 1, I_c(\mathbf{x}_2) = 0$ のとき、 $h = [0, (\frac{3z_1^n + b^n}{4})^{1/n}]^n = [0, (\frac{z_1^n + z_2^n}{2})^{1/n}]^n$

(2-2) $I_c(\mathbf{x}_1) = 1, I_c(\mathbf{x}_2) = 1$ のとき、 $h = [0, (\frac{z_1^n + 3b^n}{4})^{1/n}]^n = [0, (\frac{z_2^n + b^n}{2})^{1/n}]^n$

とするのが最良の仮説計算法であり、その誤差は、 $\frac{b^n - z_1^n}{4}$ である。

したがって、 $\max\{\frac{z_1^n - a^n}{2}, \frac{b^n - z_1^n}{4}\}$ が、 \mathbf{x}_1 を固定し、 \mathbf{x}_2 を最良にし、

最良の仮説を出力したときの最大誤差である。

これを最小にするには、補題3より、

$\frac{z_1^n - a^n}{2} = \frac{b^n - z_1^n}{4}$ から求められる。すなわち、 $2z_1^n - 2a^n = b^n - z_1^n$ より、

$z_1 = (\frac{2a^n + b^n}{3})^{1/n}$ またこのとき、

$z_2 = (\frac{\frac{2a^n + b^n}{3} + b^n}{2})^{1/n} = (\frac{2a^n + b^n + 3b^n}{6})^{1/n} = (\frac{a^n + 2b^n}{3})^{1/n}$

以上をまとめると、 $m = 2$ のとき $C_{ncube[a,b]}$ を学習する最良のアルゴリズム、すなわち、サンプリング法と仮説計算法は、

サンプル点 $\mathbf{x}_1 = ((\frac{2a^n + b^n}{3})^{1/n}, 0, \dots, 0)$

サンプル点 $\mathbf{x}_2 = ((\frac{a^n + 2b^n}{3})^{1/n}, 0, \dots, 0)$ で、仮説計算法は、補題4より、

$I_c(\mathbf{x}_1) = 0$ のときは $h = [0, (\frac{5a^n + b^n}{6})^{1/n}]^n$

$I_c(\mathbf{x}_1) = 1, I_c(\mathbf{x}_2) = 0$ のときは $h = [0, (\frac{3a^n + 3b^n}{6})^{1/n}]^n$

$I_c(\mathbf{x}_1) = 1, I_c(\mathbf{x}_2) = 1$ のときは $h = [0, (\frac{a^n + 5b^n}{6})^{1/n}]^n$

が最良であり、

そのときの最大誤差 $e_{\mathcal{A}}(2)$ は $e_{\mathcal{A}}(2) = \frac{b^n - a^n}{6}$ となる。

これは、 $e_{C_{ncube[a,b]}}(2) = \frac{b^n - a^n}{6}$ であることが示されたことになる。

3.3 一般の m の場合

3.2 と全く同様に（正確には、 m に関する数学的帰納法により）、次の定理が導ける。

[定理 1] 許容サンプル数が m のとき、 $C_{ncube[a,b]}$ を学習する最良の 1 ステージのアルゴリズムは次のとおりである。

サンプル点は、

$$\begin{aligned}
 x_1 &= \left(\left(\frac{ma^n + b^n}{m+1} \right)^{1/n}, 0, \dots, 0 \right) \\
 &\vdots \\
 x_k &= \left(\left(\frac{(m-k+1)a^n + kb^n}{m+1} \right)^{1/n}, 0, \dots, 0 \right) \quad (k = 1, 2, \dots, m) \\
 &\vdots \\
 x_m &= \left(\left(\frac{a^n + mb^n}{m+1} \right)^{1/n}, 0, \dots, 0 \right)
 \end{aligned}$$

ここで、 $z_k \stackrel{\text{def}}{=} \left(\frac{(m-k+1)a^n + kb^n}{m+1} \right)^{1/n}$ とする。

最良の仮説計算法は、

$$I_c(x_1) = 0 \text{ のときは } h = \left[0, \left(\frac{a^n + z_1^n}{2} \right)^{1/n} \right]^n = \left[0, \left(\frac{(2m+1)a^n + b^n}{2(m+1)} \right)^{1/n} \right]^n$$

$I_c(x_{k-1}) = 1, I_c(x_k) = 0 \quad (k = 2, 3, \dots, m)$ のときは、

$$h = \left[0, \left(\frac{z_{k-1}^n + z_k^n}{2} \right)^{1/n} \right]^n = \left[0, \left(\frac{(2m-2k+3)a^n + (2k-1)b^n}{2(m+1)} \right)^{1/n} \right]^n$$

$$I_c(x_m) = 1 \text{ のときは } h = \left[0, \left(\frac{z_m^n + b^n}{2} \right)^{1/n} \right]^n = \left[0, \left(\frac{a^n + (2m+1)b^n}{2(m+1)} \right)^{1/n} \right]^n$$

そしてそのときの最大誤差は、 $e_{\mathcal{A}}(m) = e_{C_{ncube[a,b]}}(m) = \frac{b^n - a^n}{2(m+1)}$ となる。

<証明> m に関する数学的帰納法による。□ (定理 1)

定理の与える $C_{ncube[0,1]}$ を学習するための最良のサンプル点と最良の仮説を与える点 ($h = [0, r]^n$ の r) を、いくつかの n について図示すると図 3 のようになる。(図の「×」がサンプル点、「I」が仮説を与える点である。)

予想通り、最適なサンプル点は、次元 n が上がるにしたがって、 $[0, 1]^n$ の表面近くに集中することが分かる。

3.4 必ず $h \subset c$ となる場合のサンプル点

学習アルゴリズムでは、どのような c を学習する場合でも仮説 h が、 $h \subset c$ となることが要求される場合がある。この場合も上記定理と全く同様に考えて、次の定理が成り立つ。

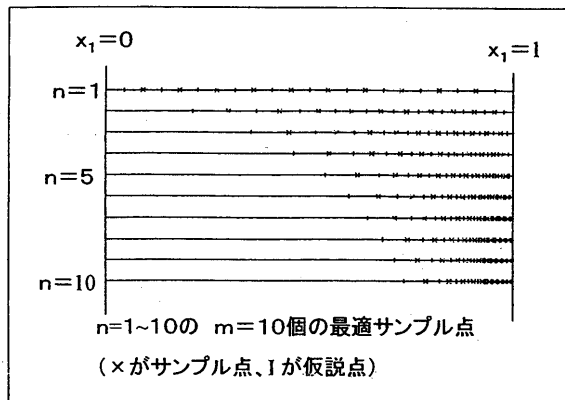


図 3 : $C_{ncube[0,1]}$ を学習するための最適サンプル点と仮説点

[定理2] 許容サンプル数が m のとき、 $C_{ncube[a,b]}$ を学習する最良の1ステージのアルゴリズムは次のとおりである。(ただし、 $\forall c \in C_{ncube[a,b]}, h \subset c$ なる仮説 h を出力するものとする。)

サンプル点は、定理1と全く同じく、

$$x_k = \left(\left(\frac{(m-k+1)a^n + kb^n}{m+1} \right)^{1/n}, 0, \dots, 0 \right) (k = 1, 2, \dots, m)$$

ここで、 $z_k \stackrel{def}{=} \left(\frac{(m-k+1)a^n + kb^n}{m+1} \right)^{1/n}$ とすると最良の仮説計算法は、

$$I_c(x_1) = 0 \text{ のときは } h = [0, a]^n$$

$$I_c(x_{k-1}) = 1, I_c(x_k) = 0 (k = 1, 2, \dots, m) \text{ のときは、}$$

$$h = [0, z_{k-1}]^n$$

$$I_c(x_m) = 1 \text{ のときは、 } h = [0, z_m]^n$$

そしてそのときの最大誤差は、 $e_{\mathcal{A}}(m) = \frac{b^n - a^n}{(m+1)}$ となる。

<証明> m に関する数学的帰納法による。□ (定理2)

このように、「必ず、学習すべき概念に含まれる仮説を出力する」という条件をつけた場合は、サンプル点はこの条件をつけない場合の最適なサンプル点と同じで、仮説計算法は「サンプルにより得られた情報に無矛盾な最小の n 次元立方体とする」になり、その最大誤差は、この条件をつけない場合の2倍となることが分かった。

4 結言および今後の課題

$C_{ncube[a,b]}$ という非常に単純なクラスをサンプリングによる学習する、最良の、確率現象を用いない1ステージアルゴリズム \mathcal{A} を構成し、その誤差 $e_{\mathcal{A}}(m)$ を求めた。そしてそれは、 $C_{ncube[a,b]}$ の構造の複雑さ(学習しにくさ)を表す $e_{C_{ncube[a,b]}}(m)$ でもある。

今後の課題として、次のことがあげられる。

- (1) これを応用して、よい C_{n1} (n 次元長方形の集合) の1ステージ学習アルゴリズムを構成すること。
例えば、 C_{21} の1ステージアルゴリズムでは、 $e_{\mathcal{A}}(m) = O(1/m^{2/3})$ のものが知られている ([4]) が、これよりも誤差の小さい1ステージアルゴリズムを構成すること。
- (2) 例えば、 $C_{nSphere} \stackrel{def}{=} \{c(r); c(r) \text{ は原点を中心とする } n \text{ 次元球の集合 } 0 \leq r \leq 1\}$ (ただし、学習領域 X は、 $X = c(1)$) というクラスも、この議論と同様に最良の1ステージアルゴリズムが構成できる。ここで最良アルゴリズムが適用できる概念クラスの性質を明らかにすること。

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, pp113-121