

単純階層構造 n 次元図形クラスの 多ステージ最適学習アルゴリズム

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

(平成 16 年 9 月受理)

大槻正伸* (OHTSUKI Masanobu)

Abstract

We consider multi-stage learning algorithms with m sample points for the simple hierarchical class of n -dimensional figures $\mathcal{C}_{n[a,b]}$. $\mathcal{C}_{n[a,b]} = \{c(t) ; a \leq t \leq b\}$ is 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 will construct the the optimal k -stage learning algorithm and show that the error of it is $(f(b) - f(a))/2(m_1 + 1)(m_2 + 1) \dots (m_k + 1)$, where $f(t)$ is the n -dimensional volume of $c(t)$, and $m_i = \lfloor m/k \rfloor$ or $\lceil m/k \rceil$ is the number of sample points in the i -th stage ($i=1, 2, \dots, k$) such that $m_1 + m_2 + \dots + m_k = m$.

And then we will show that the optimal multi stage algorithm is obtained by putting $k = m$, and the error of which is $(f(b) - f(a))/2^{(m+1)}$.

1 はじめに—概念学習とアルゴリズムの誤差について

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][5]。

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（ n 次元図形概念クラス）の「サンプリングによる学習」、ただし「多ステージ」の学習について考える。

以下、 n 次元図形のサンプリングによる学習・学習アルゴリズムのステージ数・区間と 1 対 1 に対応する単純階層構造をもつ n 次元図形のクラスについて述べる。

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。

学習者（学習アルゴリズム） \mathcal{A} は、 n 次元空間 $X \subset \mathbf{R}^n$ 内の n 次元図形の集合 $\mathcal{C} \subset 2^X$ から選ばれた未知の $c \in \mathcal{C}$ の形を、「サンプリング」（後述）により推定する。学習理論では、この「推定」を「学習」と同一視する。また、図形 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。また X のことを「学習領域」とよぶ。

\mathcal{A} は X, \mathcal{C} , および $c \in \mathcal{C}$ 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。

\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。

*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）

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

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

A は未知の $c \in \mathcal{C}$ 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。

アルゴリズム A で、概念クラス C を m 個のサンプルで学習するときの誤差 $e_A(m)$ とは次で定義される。

定義 1 学習アルゴリズムの誤差 $e_A(m) \stackrel{def}{=} \sup\{vol_n(A(c, m)\Delta c) ; c \in C\}$

ここで $A(c, m)$ は、アルゴリズム A で概念 c を m 個のサンプルで学習したときの仮説 h のこととする。

また、概念クラス C の誤差 $e_C(m)$ とは次で定義される。

定義 2 概念クラスの誤差 $e_C(m) \stackrel{def}{=} \inf\{e_A(m) ; A \text{ は } C \text{ の学習アルゴリズム}\}$

すなわち、 $e_A(m)$ は学習アルゴリズム A で概念クラス C を (m 個のサンプルで) 学習する場合の最大 (最悪ケースの) 誤差、 $e_C(m)$ はどのようなよいアルゴリズムで学習しても出てしまう、 C の学習の際に避けられない誤差、すなわち、見方を変えれば、 C の学習しにくさを表す量となっている。文献 [6] [7] では、上記誤差の定義を、全て 1 ステージのアルゴリズムに限定して議論してきたが、本論文では、この限定をおかず、一般の多ステージのアルゴリズムを想定する (「ステージ数」については後述する)。

一般に、 $e_C(m)$ の正確な評価は非常に難しい。したがってよい上界、下界で評価していくことになるが、特によい下界で評価するのが難しい。

$C_{n[a,b]}$ のような単純なクラスの最適アルゴリズム、およびその誤差について解析することは、 $e_{C_{n[a,b]}}(m)$ が正確に評価でき、したがってこのクラスを部分集合として含むより複雑なクラス C に対する $e_C(m)$ の下界の評価に役立つ。[7]

単純なクラスの最適学習アルゴリズムの設計はこのような意義をもっている。

次に「ステージ」について説明する。「ステージ」とは、「(1) m を入力する。 (m がすでに明らかになっている場合は入力は省略できる。) (2) m 個 (以下) のサンプル点を選んでそれらの点について調べる。 (3) (2) の結果をもとに何か計算をする。」という一連の動作のことである。

したがって、1 ステージ学習アルゴリズムとは、(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、1 ステージアルゴリズムとは、「ある点集合について調べて、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、どのような概念 c を学習する場合でも同じサンプル点について調べるという「静的サンプリング」のアルゴリズムである。

それに対して、多ステージアルゴリズム、例えば 2 ステージアルゴリズムとは、

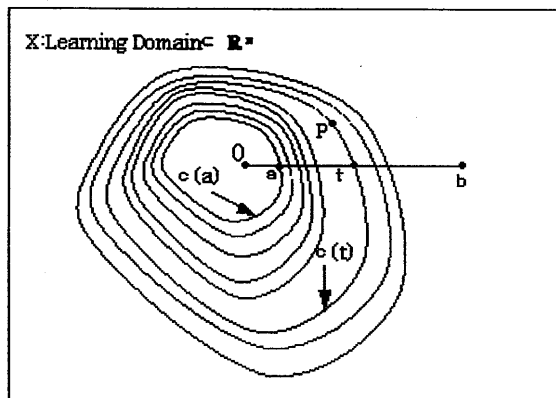
(1) m_1, m_2 を入力する。(あるいは m を入力し、 m_1, m_2 を計算する。)

<第 1 ステージ>

(2) m_1 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次の第 2 ステージで選ぶべき m_2 個のサンプル点を選ぶ。

<第 2 ステージ>

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

(4) (3) で選んだ m_2 個のサンプル点について調べる。

(5) 今までの結果から計算をして仮説を出力する。

というような構造が許される。すなわち、第2ステージ（以降）で調べるサンプル点は、学習すべき概念により異なる可能性がある。これは1ステージアルゴリズムではあり得ない、「動的サンプリング」を許容することを意味する。今回はこの多ステージアルゴリズムを考察対象とする。

「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」（以下、簡単に「単純階層クラス」とよぶことにする）とは直観的に言うと、Fig.1のように、同じような図形が1次的に（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。

これは、 $C_{n[a,b]} = \{c(t); a \leq t \leq b\}$ であり、各概念（図形） $c(t)$ は一つの実数 t ($a \leq t \leq b$) により特徴づけられていて、 $t_1 \leq t_2$ ならば $c(t_1) \subset c(t_2)$ となっているものである。^[7]

詳しく言うと、 $C_{n[a,b]}$ とは、 n 次元空間の図形（図形は閉集合とする）の集合であり、 n 次元空間中の座標づけされた線分 $[a, b]$ が存在するような状況で定義される。

そして、任意の点 $P \in X$ に対してユニークに概念（図形） $c(t)$ が定まり（あるいは $c(t)$ が存在しない場合も許されるが） $c(t)$ は P をその境界上にもつ、という性質のものと定義される。

このような $C_{n[a,b]}$ を学習する上では、サンプル点を線分 $[a, b]$ 上の点に限って議論してよいことが示されている。^[7]

さて、今までの $C_{n[a,b]}$ の学習アルゴリズムの研究では、次のことまで明らかにされている。

文献 [6] では $C_{n[a,b]}$ の特別な場合である $C_{ncube[a,b]} = \{[0, t]^n; a \leq t \leq b\}$ （原点を含む n 次元立方体の集合）について最適な1ステージ学習アルゴリズムを構成しその誤差を求めた。

文献 [7] では、一般的な単純階層クラス $C_{n[a,b]}$ の意味を厳密に定義し、その性質を調べ、それに対する最適な1ステージ学習アルゴリズムを構成しその誤差を求めた。

$f(t) \stackrel{\text{def}}{=} \text{vol}_n(c(t))$ ($\text{vol}_n(\cdot)$ は n 次元の体積) とする (f は単調増大関数であることを仮定しておく)。

この $C_{n[a,b]}$ に対する1ステージ学習アルゴリズムのサンプルは、許容サンプル数を m としたとき、

$$\text{線分 } [a, b] \text{ 上の座標の } t_i = f^{-1}\left(\frac{(m-i+1)f(a) + if(b)}{m+1}\right) \quad (i=1, 2, \dots, m)$$

に対応する線分上の点をサンプル点とするのが最適であり、そのときの誤差は

$$\frac{f(b) - f(a)}{2(m+1)}$$

であることが示されている。^[7]

今回は、この最適な1ステージアルゴリズムをもとにして、概念クラス $C_{n[a,b]}$ に対する k ステージの最適学習アルゴリズムを考える。それは、 $C_{n[a,b]}$ の最適な多ステージアルゴリズムについて考察したという意味で文献 [6]、文献 [7] の一般化になっている。

さて、本論文では以下で文献 [7] で得られた1ステージの最適学習アルゴリズムの結果、すなわち、サンプリング法およびそのときの誤差をもとに、一般の k ステージアルゴリズムで最適なものを構成し、その誤差を求める。

まず2節では、多ステージ学習アルゴリズムについて、また最適な k ステージアルゴリズムの構成方法について考察する。

次に最適な k ステージアルゴリズムは、トータル許容サンプル数を m 、各ステージのサンプル数を $m_i (i = 1, 2, \dots, k)$ とすると、最適化問題

$$(m_1 + 1)(m_2 + 1) \dots (m_k + 1) \rightarrow \max \text{ (sub.to } m_1 + m_2 + \dots + m_k = m)$$

を解くことに帰着されることを示す。

3節では、この最適化問題を解き、 k ステージの最適学習アルゴリズムを明らかにする。

結果は、 k ステージの学習アルゴリズムで最適なものは、次のようになることが示される。

$$m_i = \lfloor \frac{m}{k} \rfloor \text{ あるいは } m_i = \lceil \frac{m}{k} \rceil (i = 1, 2, \dots, k) \text{ ただし, } m_1 + m_2 + \dots + m_k = m$$

ここで、各ステージでは、前ステージまでで狭まった区間 $[a_s, b_s]$ に対して、最適な1ステージアルゴリズムのサンプリングを行うものとする。

これはすなわち、各ステージでなるべく均等な個数ずつ、かつ各ステージでも均等なサンプリングを行った方が有利であることを意味する。

このときの誤差は

$$e_{C_{n[a,b]}}^{k\text{-stage}}(m) = \frac{f(b) - f(a)}{2(m_1 + 1)(m_2 + 1) \dots (m_k + 1)}$$

また、多ステージのアルゴリズムの中で最適なものは、 $k = m$ ステージのものであり、そのときの誤差は、

$$e_{C_{n[a,b]}}^{m\text{-stage}}(m) = \frac{f(b) - f(a)}{2^{m+1}} \text{ であることも示される。}$$

2 多ステージアルゴリズムの考察

まず、1ステージの、 $C_{n[a,b]}$ に対する学習アルゴリズムについて概観する (詳しくは文献 [7] 参照)。

補題 1 サンプル点 0 個による最適な仮説

$C_{n[a,b]}$ をサンプル点 0 個で (すなわちサンプルなしで) 推定する場合、仮説は $c(t)$, $t = f^{-1}(\frac{f(a)+f(b)}{2})$ と選ぶのが最適であり、そのときの誤差は、 $\frac{f(b)-f(a)}{2}$ である。

<証明> $c(t)$ を仮説とした場合、(最悪ケースの) 誤差は $\max\{f(b) - f(t), f(t) - f(a)\}$ であり、 $f(b) - f(t)$ は単調減少関数、 $f(t) - f(a)$ は単調増大関数であるから、 $\max\{f(b) - f(t), f(t) - f(a)\}$ を最小にする t は、 $f(b) - f(t) = f(t) - f(a)$ より求まる。□ (補題 1)

この補題は、「学習すべき概念 $c(t)$ のパラメータ t の存在範囲が、区間 $[a, b]$ まで狭まった時点で仮説を出さなくてはならない場合、最良の仮説で推定しても誤差は $\frac{f(b)-f(a)}{2}$ が避けられない」ことを示している。このことはまた、「よいサンプル点の取り方は、 t の存在範囲を $f(b) - f(a)$ が小さい区間 $[a, b]$ に狭めるような方法である」ことを意味する。

したがって、 t の存在範囲が区間 $[a, b]$ まで狭まってあと m 個のサンプリングが 1 回だけ許される場合、それはすなわち、1 ステージアルゴリズムで $C_{n[a,b]}$ を学習する場合になるが、サンプル点 $\{t_1, t_2, \dots, t_m\}$ (ここで $t_1 < t_2 < \dots < t_m$ とする) は $\max\{f(t_1) - f(a), f(t_2) - f(t_1), \dots, f(b) - f(t_m)\}$ を最小にするようにするのが最適なサンプリング法である。

なぜなら、このサンプリングで、 t の存在範囲が、 $[a, t_1], [t_1, t_2], \dots, [t_m, b]$ のいずれかまで狭まり、あとは 0 個のサンプリングで仮説を出力しなくてはならないから、最悪ケース $\frac{f(t_1) - f(a)}{2}$, または、 $\frac{f(t_2) - f(t_1)}{2}$ または、 $\dots, \frac{f(b) - f(t_m)}{2}$ の誤差が出るからである。

そして 1 ステージの最適なサンプル点 $\{t_1, t_2, \dots, t_m\}$ は $f(t_1) - f(a) = \dots = f(t_2) - f(t_1) = \dots = f(b) - f(t_m) = \frac{f(b) - f(a)}{m+1}$ となるようにするのが最適であり、その場合の誤差は $\frac{f(b) - f(a)}{2(m+1)}$ である。[7]

さて今 k を固定し、 $C_{n[a,b]}$ を k ステージのアルゴリズムで、合計 m 個のサンプリングが許容されて学習する場合を考える。第 i ($i = 1, 2, \dots, k$) ステージで m_i 個のサンプルをとるものとすれば、 $m_1 + m_2 + \dots + m_k = m$ であるが、次のようにサンプリングするのが最適であることは明らかである。

第 1 ステージ 区間 $[a, b]$ に対し、最適な 1 ステージアルゴリズムにより m_1 個のサンプルで t の存在範囲を $[a_1, b_1]$ に狭める。ただし $f(b_1) - f(a_1) = \frac{f(b) - f(a)}{m_1 + 1}$

第 2 ステージ 区間 $[a_1, b_1]$ に対し、最適な 1 ステージアルゴリズムにより m_2 個のサンプルで t の存在範囲を $[a_2, b_2]$ に狭める。ただし $f(b_2) - f(a_2) = \frac{f(b_1) - f(a_1)}{m_2 + 1}$

⋮

第 k ステージ 区間 $[a_{k-1}, b_{k-1}]$ に対し、最適な 1 ステージアルゴリズムにより m_k 個のサンプルで t の存在範囲を $[a_k, b_k]$ に狭める。ただし $f(b_k) - f(a_k) = \frac{f(b_{k-1}) - f(a_{k-1})}{m_k + 1}$

さて、上記のように狭めていくと、最終的に t の存在範囲は、

$$[a_k, b_k] \text{ まで狭まる。ただし、} f(b_k) - f(a_k) = \frac{f(b) - f(a)}{(m_1 + 1)(m_2 + 1) \dots (m_k + 1)}$$

そしてこのときのアルゴリズム \mathcal{A} の最終的な誤差は

$$e_{\mathcal{A}}(m) = \frac{f(b) - f(a)}{2(m_1 + 1)(m_2 + 1) \dots (m_k + 1)} \text{ となる。}$$

以上のことにより、 $C_{n[a,b]}$ の最適多ステージアルゴリズム (サンプリング法) を求めるためには、次の問題が自然に提起される。

$k, m_1, m_2, \dots, m_k, m$ を全て自然数とする。

問題 1. $(m_1 + 1)(m_2 + 1) \dots (m_k + 1)$ を条件「 $m_1 + m_2 + \dots + m_k = m$ 」のもとで最大にする。 $(k$ を固定してこのような m_1, m_2, \dots, m_k を求める。)

問題 2. $G\max_k(m) \stackrel{\text{def}}{=} \max\{(m_1 + 1)(m_2 + 1) \dots (m_k + 1) ; m_1 + m_2 + \dots + m_k = m, m_i \in \mathbf{Z}^+\}$ として、 $\max\{G\max_k(m) ; k = 1, 2, \dots, m\}$ を最大にする k およびそのときの $G\max_k(m)$ の値を求める。

次節でこの 2 つの問題を解き、 $C_{n[a,b]}$ に対する多ステージの最適な学習アルゴリズム (i.e. サンプリング法) を求める。

3 多ステージ最適学習アルゴリズム

まず、 m, k を固定しておき問題 1. $(m_1+1)(m_2+1)\dots(m_k+1) \rightarrow \max$ (Sub.to $m_1+m_2+\dots+m_k = m$) を解く。

ここで、各 $m_i (i = 1, 2, \dots, k)$ は $m_i \geq 0$ としておく。 $m_i \geq 1$ は当然であるが、後で問題 2. を解く際にこの結果を使用するために拡張して 0 以上としておくのである。ここで次の関数を定義する。

定義 3 $\phi_{ij}, g_k, Gmax_k(m)$

$\mathbf{Z}^+ = \{0, 1, 2, 3, \dots\}$ (非負整数の集合) とする。

$\phi_{i,j} : (\mathbf{Z}^+)^k \rightarrow (\mathbf{Z}^+)^k$ および、 $g_k : (\mathbf{Z}^+)^k \rightarrow \mathbf{Z}^+$ を次で定義する。

$\phi_{ij}(m_1, m_2, \dots, m_k) \stackrel{\text{def}}{=} (m_1, \dots, m_i - 1, \dots, m_j + 1, \dots, m_k)$ (i, j は $1 \leq i, j \leq k$)

$g_k(m_1, m_2, \dots, m_k) \stackrel{\text{def}}{=} (m_1 + 1)(m_2 + 1)\dots(m_k + 1)$

$Gmax_k(m) \stackrel{\text{def}}{=} \max\{g_k(m_1, m_2, \dots, m_k) ; m_1 + m_2 + \dots + m_k = m, m_i \in \mathbf{Z}^+ (i = 1, 2, \dots, k)\}$

補題 2 $g_k(m_1, \dots, m_k) < g_k(\phi_{i,j}(m_1, \dots, m_k)) \iff m_i - m_j > 1$ (i.e. $m_i - m_j \geq 2$)

<証明>

$(m_1 + 1)\dots(m_i + 1)\dots(m_j + 1)\dots(m_k + 1) < (m_1 + 1)\dots(m_i)\dots(m_j + 2)\dots(m_k + 1)$

$\iff (m_i + 1)(m_j + 1) < m_i(m_j + 2) \iff m_i m_j + m_i + m_j + 1 < m_i m_j + 2m_i$

より明らか。□ (補題 2)

補題 3 $m = kq + r$ ($0 \leq r < k$) のとき、 $g_k(m_1, \dots, m_k)$ が最大になるのは、 $m_1 = q + 1, m_2 = q + 1, \dots, m_r = q + 1, m_{r+1} = q, \dots, m_k = q$ のときである。

<証明> $\phi_{ij}(m_1, m_2, \dots, m_k)$ は、第 i 成分から第 j 成分への 1 の「移動」である。

この移動により、 $g_k(m_1, m_2, \dots, m_k)$ の値が大きくなるのは、 m_i と m_j にたくさん (2 以上の) 差があり、その差を均等にするような方向の移動のときのみである、というのが補題 2 の主張するところである。したがって、なるべく均等に m_1, m_2, \dots, m_k をとった方が $g_k(m_1, m_2, \dots, m_k)$ が大きくなるから、この補題の結果は直感的にはほとんど明らかである。

以下でより詳しく証明する。

$S_k(m) \stackrel{\text{def}}{=} \{(m_1, m_2, \dots, m_k) \in (\mathbf{Z}^+)^k ; m_1 + m_2 + \dots + m_k = m \in \mathbf{Z}^+\}$ とする。

$\mathbf{y} \stackrel{\text{def}}{=} (q + 1, q + 1, \dots, q + 1, q, q, \dots, q)$ ($(q + 1)$ が r 個、 q が $(k - r)$ 個)

このとき以下を示せば十分である。

(*) $\forall \mathbf{x} = (m_1, m_2, \dots, m_k) \in S_k(m), \exists \phi_{i_1 j_1}, \phi_{i_2 j_2}, \dots, \phi_{i_p j_p}$

(1) $\phi_{i_p j_p} \circ \dots \circ \phi_{i_2 j_2} \circ \phi_{i_1 j_1}(\mathbf{x}) = \mathbf{y}$ かつ

(2) $g_k(\mathbf{x}) \leq g_k(\phi_{i_1 j_1}(\mathbf{x})) \leq g_k(\phi_{i_2 j_2} \circ \phi_{i_1 j_1}(\mathbf{x})) \leq \dots \leq g_k(\phi_{i_p j_p} \circ \dots \circ \phi_{i_2 j_2} \circ \phi_{i_1 j_1}(\mathbf{x}))$

以下 (*) の証明。

任意の \mathbf{x} を $\mathbf{x} = (m_1, m_2, \dots, m_k)$ とする。ただし、 $m_1 + m_2 + \dots + m_k = m, m_i \geq 0 (i = 1, 2, \dots, k)$
 $m = kq + r (0 \leq r < k)$

この m_1, m_2, \dots, m_k を一般性を失わずに次のようにおく。(Fig.2)

$m_1 = q + 1 + a_1, m_2 = q + 1 + a_2, \dots, m_u = q + 1 + a_u (a_i > 0, i = 1, 2, \dots, u)$

$m_{u+1} = q + 1 + a_{u+1}, m_{u+2} = q + 1 + a_{u+2}, \dots, m_v = q + 1 + a_v (a_i < 0, i = u + 1, \dots, v)$

$m_{v+1} = q + 1 + a_{v+1}, m_{v+2} = q + 1 + a_{v+2}, \dots, m_r = q + 1 + a_r (a_i = 0, i = v + 1, \dots, r)$

$m_{r+1} = q + b_{r+1}, m_{r+2} = q + b_{r+2}, \dots, m_w = q + b_w (b_i > 0, i = r + 1, \dots, w)$

$m_{w+1} = q + b_{w+1}, m_{w+2} = q + b_{w+2}, \dots, m_z = q + b_z (b_i < 0, i = w + 1, \dots, z)$

$m_{z+1} = q + b_{z+1}, m_{z+2} = q + b_{z+2}, \dots, m_k = q + b_k (b_i = 0, i = z + 1, \dots, k)$

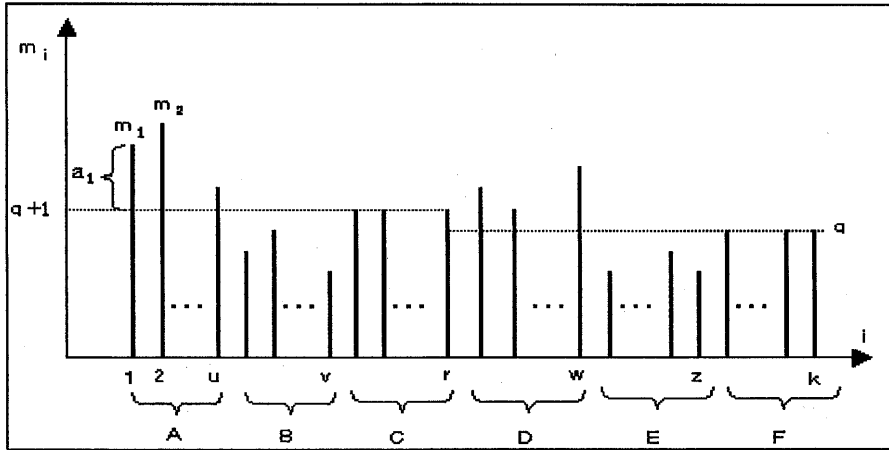


Fig. 2: Aspect of $x = (m_1, m_2, \dots, m_k) \in S_k(m)$

また、以下の添え字の集合を定めておく。(Fig. 2)

$$A = \{i; a_i > 0\} \quad B = \{i; a_i < 0\} \quad C = \{i; a_i = 0\}$$

$$D = \{i; b_i > 0\} \quad E = \{i; b_i < 0\} \quad F = \{i; b_i = 0\}$$

さて、上記の定義から

$$m_1 + m_2 + \dots + m_k = (q+1)r + q(k-r) + a_1 + \dots + a_r + b_{r+1} + \dots + b_k = m \text{ より、}$$

$$a_1 + \dots + a_r + b_{r+1} + \dots + b_k = 0$$

したがって、次の4パターンに分類されるいくつかの ϕ_{ij} により、 $\phi_{i_p j_p} \circ \dots \circ \phi_{i_2 j_2} \circ \phi_{i_1 j_1}(\mathbf{x}) = \mathbf{y}$ とすることができる。

(パターン1) $i \in A$ かつ $j \in B$ (パターン2) $i \in A$ かつ $j \in E$

(パターン3) $i \in D$ かつ $j \in B$ (パターン4) $i \in D$ かつ $j \in E$

しかも、 $g_k(\mathbf{x}) \leq g_k(\phi_{i_1 j_1}(\mathbf{x})) \leq g_k(\phi_{i_2 j_2} \circ \phi_{i_1 j_1}(\mathbf{x})) \dots$

というように、パターン3のみ ϕ_{ij} を施す前後で g_k の値が「大きくなるまたは等しくなる」ということもあり得るが、その他は ϕ_{ij} を施すことにより g_k の値を増大させるように、 $\phi_{i_1 j_1}, \dots, \phi_{i_p j_p}$ を選ぶことができるのは、補題2より明らかである。□ (補題3)

補題4 $k+1 \leq m$ のとき $Gmax_k(m) < Gmax_{k+1}(m)$

<証明> $m = kq + r$ ($0 \leq r < k$) とする。

$$Gmax_k(m) = g_k(q+1, \dots, q+1, q, \dots, q) = g_{k+1}(q+1, \dots, q+1, q, \dots, q, 0) < Gmax_{k+1}(m) \quad \square$$

(補題4)

これまでの議論により、次の定理が証明されたことになる。

定理1 $C_{n[a,b]}$ を学習する多ステージアルゴリズムについて次のことが言える。

1. k ステージの最適アルゴリズムは、許容サンプル数が $m = kq + r$ ($0 \leq r < k$) のとき次の通りである。

$m_1 = q+1, m_2 = q+1, \dots, m_r = q+1, m_{(r+1)} = q, \dots, m_k = q$ で、各ステージともそれまで狭まった区間で最適1ステージアルゴリズムを用いて概念を表すパラメータ t の存在範囲を狭めていくものである。そしてそのときの誤差は

$$e_{C_{n[a,b]}}^{k\text{-stage}}(m) = \frac{f(b) - f(a)}{2(\lceil \frac{m}{k} \rceil + 1)^r (\lfloor \frac{m}{k} \rfloor + 1)^{(k-r)}} \text{ である。}$$

2. 多ステージのアルゴリズムの中で最も誤差の少ないものは許容サンプル数が m のとき、 $k = m$ ステージのアルゴリズムであり、そのときの誤差は、

$$e_{C_{n[a,b]}}^{m-stage}(m) = \frac{f(b) - f(a)}{2^{(m+1)}} \text{である。}$$

4 結言

単純階層構造 n 次元図形クラス $C_{n[a,b]}$ に対して、多ステージの最適学習アルゴリズムとその誤差を求めた。

m 個のサンプルで学習する多ステージの最適アルゴリズムは、1つのサンプルを取るごとに、次のサンプルを計算するという2分探索方式の m ステージアルゴリズムであり、その誤差が許容サンプル数 m の指数関数の逆数に比例して少なくなる、非常に効率のよいものであることが分かった。

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: 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,
- [6] 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,
- [7] 大槻 正伸: 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, 福島工業高等専門学校紀要第44号, pp15-23, 2004年,