数理解析研究所講究録
第 90 巻 1970 年 124-142

学習によるパターン分類について

九大理新大理　田　中　謙輔

§1. まえがき

最近パターン認識の問題に学習による方法が多く用いられるようになった。これらの方面の研究は，良く知られている様に Ya. Z. Tsypkin, M. A. Aizerman, E. M. Braverman, J. M. Schumpert, S. S. Yau, J. D. Patterson, K. S. Fu, etc などの人々によっていろいろの type でなされた。
ここではいろいろの type の中で，特に H. Robbins, S. Monro, によって導入されている，J. R. Blum によって多次元に拡張された確率近似法（Stochastic Approximation）の応用として学習のアルゴリズムが組まれている type に注目し，次のような２つの問題を考えたい。

(I) 次の (1) (2) (3) をみたすパターン認識の問題

(1) θ_1, θ_2（category と呼ばれる）なる R^n におけるパターンの２つの class が存在するとする.

(2) θ_1, θ_2 の a priori distribution はそれぞれ g_1, g_2 とする

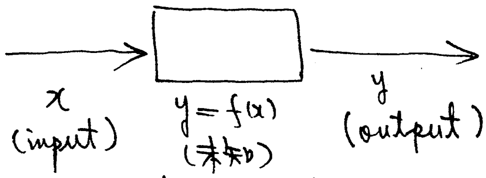
(3) θ_1, θ_2 が真のときパターンの probability density function

はそれぞれ $f_1(x)$, $f_2(x)$ とする．

ここに $q_1, q_2, f_1(x), f_2(x)$ は未知とする．もしこの問題で $q_1, q_2, f_1(x), f_2(x)$ が知られていれば Bayes の意味で optimal な判定関数を用いることができる。我々の問題では $q_1, q_2, f_1(x), f_2(x)$ が未知であるが training sequence $(x^1, \theta^1), \cdots$ を用いることができる．これを用いて Bayes の意味で optimal な判定関数 $D(x)$ を近似するアルゴリズムを求める．ただし

$$D(x) = \begin{cases} \dfrac{q_1 f_1(x) - q_2 f_2(x)}{p(x)} & , \ p(x) = q_1 f_1(x) + q_2 f_2(x) > 0 \\ 0 & , \ p(x) = 0 \end{cases} .$$

(II) 未知関数を構成する問題



上図の system においては，関数 $f(x)$ は未知であるが，input x と output y は観測できる．

このとき互に独立な同じ分布にしたがう input の列と，それに対応する output の列から未知関数 $f(x)$ を近似するアルゴリズムを求める．

§2. 準備

この章では以下の主な結果を証明するために3つの補助定理を述べる．$\{y^n\}_{n=1}^{\infty}$ は実数値 stochastic process とし，$\{U_n\}_{n=1}^{\infty}$, $\{V_n\}_{n=1}^{\infty}$, $\{\zeta_n\}_{n=1}^{\infty}$ は非負なる実数値可測関数の3つの列とする

ただし，各々の U_n, V_n, ζ_n は R^{nN} の上に定義されている。
このとき $\{U_n(y^1, y^2, \cdots, y^n)\}_{n=1}^{\infty}$，$\{V_n(y^1, y^2, \cdots, y^n)\}_{n=1}^{\infty}$，$\{\zeta_n(y^1, y^2, \cdots, y^n)\}_{n=1}^{\infty}$ は再び stochastic process となる。以下では記号を簡単にするために，$U_n = U_n(y^1, y^2, \cdots, y^n)$，$V_n = V_n(y^1, y^2, \cdots, y^n)$，$\zeta_n = \zeta_n(y^1, y^2, \cdots y^n)$ と書くことにする．更に $\{\gamma_n\}_{n=1}^{\infty}$，$\{\mu_n\}_{n=1}^{\infty}$ は実数の2つの列とする。

今3つの stochastic process $\{U_n\}_{n=1}^{\infty}$，$\{V_n\}_{n=1}^{\infty}$，$\{\zeta_n\}_{n=1}^{\infty}$ に対して次の様な基本条件を導入する：

(A1) $E[U_1]$ と $E[V_1]$ は存在する．

(A2) 全ての n について

$$E[U_{n+1} \mid y^1, y^2, \cdots, y^n] \leqq (1+\mu_n)U_n - \gamma_n V_n + \zeta_n$$ が成立する

(A3) $\displaystyle\sum_{n=1}^{\infty} \gamma_n = \infty$.

(A4) $\displaystyle\sum_{n=1}^{\infty} |\mu_n| < \infty$.

(A5) 次の様な条件を満す正数の列 $\{M_n\}_{n=1}^{\infty}$ が存在する

全ての n について $P[\zeta_n \leqq M_n] = 1$,

$$\sum_{n=1}^{\infty} M_n < \infty .$$

<u>補助定理 1</u>．3つの stochastic process $\{U_n\}_{n=1}^{\infty}$，$\{V_n\}_{n=1}^{\infty}$，$\{\zeta_n\}_{n=1}^{\infty}$ が次の様な条件を満しているとする：(i) 条件 (A1) ～(A5) が成立する，(ii) $\lim_{n\to\infty} \gamma_n = 0$ ，(iii) もし $P[\lim_{k\to\infty} V_{n_k} = 0] = 1$ なる部分列 $\{n_k\}$ が存在すれば，$P[\lim_{k\to\infty} U_{n_k} = 0] = 1$ となる．
このとき次の様な事柄が成立する

$$P[\lim_{n\to\infty} U_n = 0] = 1 \quad , \quad \lim_{n\to\infty} E[U_n^{\beta}] = 0 \ \text{for all } 0<\beta<1 .$$

補助定理2. 次の様な条件を満している非負の実数列 $\{a_n\}_{n=1}^{\infty}$ が存在するとする：

(2.1) $\quad a_{n+1} \leqq (1-\gamma_{n+1})a_n + A_{n+1} \quad , \quad \forall n \geqq n_0$

(2.2) $\quad \sum_{n=1}^{\infty} \gamma_n = \infty ,$

(2.3) $\quad \lim_{n\to\infty} \gamma_n = 0 ,$

(2.4) $\quad \sum_{n=1}^{\infty} A_n < \infty ,$

ただし $\{\gamma_n\}_{n=1}^{\infty}$, $\{A_n\}_{n=1}^{\infty}$ は2つの非負の実数列で，n_0 は或る正整数とする．

このとき次の様な事柄が成立する．

$$\lim_{n\to\infty} a_n = 0 \quad .$$

補助定理3. 次の様な条件を満している非負の実数列 $\{a_n\}_{n=1}^{\infty}$ が存在するとする：

(2.5) $\quad a_{n+1} \leqq (1 - A/n^s)\, a_n + B/n^t \quad , \quad \forall n \geqq n_0 ,$

(2.6) $\quad t$ はある実数で $0<s<1$ である．

このとき次の様な事柄が成立する

$$\limsup_{n\to\infty} n^{t-s} a_n \leqq B/A \quad .$$

§3. teacher をもつ学習によるパターン分類

我々が"学習"と云う言葉について感じるいろいろの面を考えてみると，これまで論じられているモデル(I)を次の様に変型したモデルに想定することは自然である様に思われる．ここでは category の集合 $\Theta = \{\theta_1, \theta_2\}$ の場合を述べる．

time	1	2	‥‥	n	‥‥
category	θ^1	θ^2	‥‥	θ^n	‥‥
pattern space	X_1	X_2	‥‥	X_n	‥‥
pattern	x^1	x^2	‥‥	x^n	‥‥
transition prob. density fn. (未知)	$p(x^1, \theta^1)$	$p(x^2, \theta^2 \mid x^1, \theta^1)$	‥‥	$p(x^n, \theta^n \mid \xi^{n-1}; \alpha^{n-1})$ $\xi^{n-1} = (x^1, \cdots, x^{n-1})$ $\alpha^{n-1} = (\theta^1, \theta^2, \cdots, \theta^{n-1})$	‥‥

理想的な状態（各時点での transition prob density fn を知っている）では，Bayes の意味で optimal な判定をする。即ち各時点 n で x^n を観測したとき

$$D^*(x^n \mid \xi^{n-1}, \alpha^{n-1}) = \Pi_{x^n}(\alpha^{n-1}\theta_1) - \Pi_{x^n}(\alpha^{n-1}\theta_2) \quad \geqq 0 \quad \text{ならば}$$

x^n は category θ_1 をもつと判定する．

$$D^*(x^n \mid \xi^{n-1}, \alpha^{n-1}) = \Pi_{x^n}(\alpha^{n-1}\theta_1) - \Pi_{x^n}(\alpha^{n-1}\theta_2) \quad < 0 \quad \text{ならば}$$

x^n は category θ_2 であると判定する．ただし

$$\Pi_{x^n}(\alpha^{n-1}\theta_i) = \frac{\pi(\alpha^{n-1})\, p(x^n, \theta_i \mid \xi^{n-1}, \alpha^{n-1})}{\displaystyle\sum_{\tilde{\alpha}^{n-1} \in \Theta^{n-1}} \sum_{\tilde{\theta}^n \in \Theta} \pi(\tilde{\alpha}^{n-1})\, p(x^n, \tilde{\theta}^n \mid \xi^{n-1}, \alpha^{n-1})}$$

（$\pi(\alpha^{n-1})$ は $\Theta^{n-1} = \Theta \times \Theta \times \cdots \times \Theta \ni \alpha^{n-1}$ の上の確率）

上の様な判定方法は統計的決定理論より誤りの確率を最小にすることが知られている.

然し我々の問題では各時点での transition prob. density f_n は未知であるが, 従属関係をもっている training sequence (x^1, θ^1), (x^2, θ^2), …が利用できる場合を考える. このとき各時点で与えられる orthonormal function の集合 $\{\varphi_i^{(n)}(x)\}_{i=1}^{N}$ と training sequence を用いて, 各時点 n で,

$$I_n = E[(D(x^n|\xi^{n-1}, d^{n-1}) - \sum_{i=1}^{N} c_i^{(n)}(\xi^{n-1}, d^{n-1})\varphi_i^{(n)}(x^n))^2 | \xi^{n-1}, d^{n-1}]$$

$$= \int_{X_n} (D(x^n|\xi^{n-1}, d^{n-1}) - \sum_{i=1}^{N} c_i^{(n)}(\xi^{n-1}, d^{n-1})\varphi_i^{(n)}(x^n))^2 p(x^n, \theta^n|\xi^{n-1}, d^{n-1}) dx^n$$

$$(D(x^n|\xi^{n-1}, d^{n-1}) = p(x^n, \theta_1|\xi^{n-1}, d^{n-1}) - p(x^n, \theta_2|\xi^{n-1}, d^{n-1}))$$

を最小にする $\{c_{i*}^{(n)}(\xi^{n-1}, d^{n-1})\}_{i=1}^{N}$ を求めて, $\sum_{i=1}^{N} c_{i*}^{(n)}(\xi^{n-1}, d^{n-1})\varphi_i^{(n)}(x^n)$ を判定関数として, パターンの category θ^n を判定することとしたい.

今 $\dfrac{\partial I_n}{\partial c_j^{(n)}} = 0 \quad (j=1, 2, \cdots N)$ より

$$c_{j*}^{(n)}(\xi^{n-1}, d^{n-1}) = E_{\theta_1}[\varphi_j^{(n)}(x^n)|\xi^{n-1}, d^{n-1}] - E_{\theta_2}[\varphi_j^{(n)}(x^n)|\xi^{n-1}, d^{n-1}]$$ が得られる

ただし $E_{\theta_i}[\varphi_j^{(n)}(x^n)|\xi^{n-1}, d^{n-1}] = \int_{X_n} \varphi_j^{(n)}(x^n) p(x^n, \theta_i|\xi^{n-1}, d^{n-1}) dx^n$.

上の議論から T. Kitagawa [6] によって導入され, V. Dupač [4] によって詳細に論じられた "modified stochastic approximation" の応用として次の様なアルゴリズムが考えられる.

<u>Algorithm</u>

最初に観測される x^1 と teacher に指示される θ^1 の対 (x^1,θ^1) より各 j について，次の様な型で $C_j^{(1)}(\xi^1,\alpha^1)$ を構成する：

$$C_j^{(1)}(\xi^1,\alpha^1) = C_j^{(0)} + \gamma_1\left[\rho^{(1)}(\theta^1)\varphi_j^{(1)}(x^1) - (1-\rho^{(1)}(\theta^1))\varphi_j^{(1)}(x^1)\right]$$

ただし　$C_j^{(0)} \equiv 0 \quad (j=1,2,\cdots,N)$

$$\rho^{(1)}(\theta^1) = \begin{cases} 1 & \theta^1=\theta_1 \\ 0 & \theta^1=\theta_2 \end{cases} .$$

次に観測される x^2 と teacher によって指示される θ^2 の対 (x^2,θ^2) より各 j について，次の様な型で $C_j^{(2)}(\xi^2,\alpha^2)$ を構成する：

$$C_j^{(2)}(\xi^2,\alpha^2) = C_j^{(1)}(\xi^1,\alpha^1) + \gamma_2\left[\rho^{(2)}(\theta^2)\varphi_j^{(2)}(x^2) - (1-\rho^{(2)}(\theta^2))\varphi_j^{(2)}(x^2)\right]$$

ただし

$$\rho^{(2)}(\theta^2) = \begin{cases} 1 & \theta^2=\theta_1 \\ 0 & \theta^2=\theta_2 \end{cases} .$$

一般には観測される x^n と teacher によって指示される θ^n の対 (x^n,θ^n) より各 j について次の様な型で $C_j^{(n)}(\xi^n,\alpha^n)$ を構成する：

$$C_j^{(n)}(\xi^n,\alpha^n) = C_j^{(n-1)}(\xi^{n-1},\alpha^{n-1}) + \gamma_n\left[\rho^{(n)}(\theta^n)\varphi_j^{(n)}(x^n) - (1-\rho^{(n)}(\theta^n))\varphi_j^{(n)}(x^n)\right]$$

ただし

$$\rho^{(n)}(\theta^n) = \begin{cases} 1 & \theta^n=\theta_1 \\ 0 & \theta^n=\theta_2 \end{cases}$$

ただし上のアルゴリズムで用いられている非負の実数列 $\{\gamma_n\}_{n=1}^{\infty}$ は次の様な性質を満している：

$$\sum_{n=1}^{\infty}\gamma_n=\infty \quad , \quad \sum_{n=1}^{\infty}\gamma_n^2<\infty \quad .$$

このとき $C_j^{(n)}(\zeta^n,\alpha^n)$ と $C_{j*}^{(n)}(\zeta^{n-1},\alpha^{n-1})$ について以下の様な定理が成立することが示される.

<u>定理3.1</u> 次の様な条件を満している非負の実数列 $\{M_n\}_{n=1}^{\infty}$ と3つの非負の実数 K_1, K_2, K_3 が存在する;

(i) $P[(\theta_j^{(n)})^2\leqq\gamma_{n+1}M_n]=1 \quad \forall n, j$,

(ii) $\sum_{n=1}^{\infty}M_n<\infty$,

(iii) $Var[Y_j^{(n+1)} \mid \zeta^n,\alpha^n]\leqq K_1(U_j^{(n)})^2+K_2(\theta_j^{(n)})^2+K_3$,

ただし $\theta_j^{(n)}=C_{j*}^{(n)}(\zeta^{n-1},\alpha^{n-1})-C_{j*}^{(n+1)}(\zeta^n,\alpha^n)$, $U_j^{(n)}=C_j^{(n)}(\zeta^n,\alpha^n)-C_{j*}^{(n)}(\zeta^{n-1},\alpha^{n-1})$

$$Y_j^{(n+1)}=\rho^{(n+1)}(\theta^{n+1})\varphi_j^{(n+1)}(x^{n+1})-(1-\rho^{(n+1)}(\theta^{n+1}))\varphi_j^{(n+1)}(x^{n+1}) \quad .$$

このとき次の様な事柄が成立する

$$P[\lim_{n\to\infty}U_j^{(n)}=0]=1 \ \forall j, \quad \lim_{n\to\infty}E[(U_j^{(n)})^{2\beta}]=0 \quad , 0<\beta\leqq 1 .$$

<u>証明</u> 各 $j=1,2,\cdots,N$ に対して, $C_j^{(n+1)}(\zeta^{n+1},\alpha^{n+1})$ の構成方法より

$$C_j^{(n+1)}(\zeta^{n+1},\alpha^{n+1})-C_{j*}^{(n+1)}(\zeta^n,\alpha^n)=(1-\gamma_{n+1})[C_j^{(n)}(\zeta^n,\alpha^n)-C_{j*}^{(n)}(\zeta^{n-1},\alpha^{n-1})]+$$
$$(1-\gamma_{n+1})[C_{j*}^{(n)}(\zeta^{n-1},\alpha^{n-1})-C_{j*}^{(n+1)}(\zeta^n,\alpha^n)]+\gamma_{n+1}[\rho^{(n+1)}(\theta^{n+1})\varphi_j^{(n+1)}(x^{n+1})-(1-\rho^{(n+1)}(\theta^{n+1}))$$
$$\varphi_j^{(n+1)}(x^{n+1})-C_{j*}^{(n+1)}(\zeta^n,\alpha^n)] \quad .$$

上式に記号 $U_j^{(n)}, \theta_j^{(n)}$ を用いて平方することによって次式を得る

$$(U_j^{(n+1)})^2\leqq(1-\gamma_{n+1})^2(U_j^{(n)})^2+(1-\gamma_{n+1})^2(\theta_j^{(n)})^2+\gamma_{n+1}^2[Y_j^{(n+1)}-C_{j*}^{(n+1)}(\zeta^n,\alpha^n)]^2$$
$$+2(1-\gamma_{n+1})^2|U_j^{(n)}||\theta_j^{(n)}|+2(1-\gamma_{n+1})\gamma_{n+1}(Y_j^{(n+1)}-C_{j*}^{(n+1)}(\zeta^n,\alpha^n))(U_j^{(n)})$$
$$+2(1-\gamma_{n+1})\gamma_{n+1}(Y_j^{(n+1)}-C_{j*}^{(n+1)}(\zeta^n,\alpha^n))(\theta_j^{(n)}) \quad .$$

両辺の条件附き期待値をとり, 定理の条件を用いると

$$E[(u_j^{(n+1)})^2 \mid \xi^n, d^n] \leqq \{1-2\gamma_{n+1}+\gamma_{n+1}^2(1+K_1)\}(u_j^{(n)})^2+\{1+\gamma_{n+1}^2(1+K_2)\}(\theta_j^{(n)})^2$$
$$+2|u_j^{(n)}||\theta_j^{(n)}|$$

が成立し，更に不等式

$$2|u_j^{(n)}||\theta_j^{(n)}| \leqq \gamma_{n+1}(u_j^{(n)})^2+(\theta_j^{(n)})^2/\gamma_{n+1}$$ を用いることによって

(3.1) $$E[(u_j^{(n+1)})^2 \mid \xi^n, d^n] \leqq \{1+\gamma_{n+1}^2(1+K_1)\}(u_j^{(n)})^2-\gamma_{n+1}(u_j^{(n)})^2+\{1+\gamma_{n+1}^2(1+K_2)\}\cdot$$
$$\cdot(\theta_j^{(n)})^2+(\theta_j^{(n)})^2/\gamma_{n+1}+\gamma_{n+1}^2K_3$$

を得る。これ故に補助定理1を用いて次の結果を得る:

$$P[\lim_{n\to\infty} u_j^{(n)}=0]=1,\quad \lim_{n\to\infty}E[(u_j^{(n)})^{2\beta}]=0,\ 0<{}^\forall\beta<1.$$

次に(3.1)の両辺の期待値をとることより次の式が得られる:

$$E[(u_j^{(n+1)})^2] \leqq [1-\gamma_{n+1}\{1-\gamma_{n+1}(1+K_1)\}]E[(u_j^{(n)})^2]+\{1+\gamma_{n+1}^2(1+K_2)\}\cdot$$
$$\cdot E[(\theta_j^{(n)})^2]+E[(\theta_j^{(n)})^2]/\gamma_{n+1}+\gamma_{n+1}^2K_3 .$$

これ故に補助定理2を用いて次の結果を得る:

$$\lim_{n\to\infty}E[(u_j^{(n)})^2]=0 .$$

更に平均収束の order については次の様な定理が成立する.

<u>定理3.2</u>　定理3.1の条件(iii)の他に次の条件が成立する:

(iv) $\gamma_n=a/n^\alpha$, $a>0$, $1/2<\alpha<1$,

(v) $E[(\theta_j^{(n)})^2]=O(n^{-2w})$, $w>\alpha$.

このとき次の様な事柄が成立する.

$$E[(U_j^{(n)})^2] = \begin{cases} O(n^{-2(w-\alpha)}) & , w < 3/2\,\alpha \\ O(n^{-\alpha}) & , w \geqq 3/2\,\alpha \end{cases}$$

証明　定理3.1の証明より，上の定理の条件を用いることによって次式が成立する：　$\forall n \geqq N$

$$E[(U_j^{(n+1)})^2] \leqq (1 - C_1/n^{\alpha})E[(U_j^{(n)})^2] + C_2/n^{2\alpha} + C_3/n^{2w-\alpha} .$$

上式より　$w < (3/2)\alpha$ ならば

$$E[(U_j^{(n+1)})^2] \leqq (1 - C_1/n^{\alpha})E[(U_j^{(n)})^2] + C_4/n^{2w-\alpha} ,$$

$w \geqq (3/2)\alpha$ ならば

$$E[(U_j^{(n+1)})^2] \leqq (1 - C_1/n^{\alpha})E[(U_j^{(n)})^2] + C_5/n^{2\alpha}$$

ただし C_1, C_2, C_3, C_4, C_5 は或る定数である．それ故に補助定理3を用いて定理の結果が得られる．

例3.1

(i)　パターン空間 $X_n = X$ $\forall n$,　$\Theta = \{\theta_1, \theta_2\}$,

(ii)　$p(x^{n+1}, \theta_i \mid \xi^n, \alpha^n) = q^{(n+1)}(\theta_i \mid \theta^n) f_i(x^{n+1})$　$\forall n$,

(iii)　$\displaystyle\sum_{\theta_i \in \Theta} q^{(n+1)}(\theta_i \mid \theta^n) = 1$,

(iv)　$\displaystyle\int_X f_i(x)\,dx = 1$,

(v)　次の様な条件をみたす $1 \geqq q_1 \geqq 0$ なる q_1 が存在する：

$$P[\,|q^{(n+1)}(\theta_1 \mid \theta^n) - q_1| \leqq C/(n+1)^{1+\alpha}\,] = 1 \quad \forall n$$

ただし C, α は或る正数である．

上の様な特別なモデルに対して，更に

(vi) $\gamma_n = 1/n$

(vii) $E_{\theta_i}[\varphi_j^2(x)] = \int_X \varphi_j^2(x) f_i(x)\, dx < \infty \quad \forall i, j$

$(\varphi_j^{(1)}(x) = \varphi_j^{(2)}(x) = \cdots = \varphi_j^{(n)}(x) = \cdots)$

なる orthonormal fn の集合$\{\varphi_i(x)\}_{i=1}^N$が存在する。

上の様な状態では定理3.1の結果が成立する。

<u>例3.2</u>　例3.1における条件(i)(ii)(iii)(iv)(vii)の他に

(i) $\gamma_n = a/n^\alpha$, $a > 0$, $(1/2) < \alpha < 1$,

(ii) 次の様な条件をみたす$1 \geq q_1 \geq 0$なるq_1が存在する:

$$E[(q_j^{(n+1)}(\theta_1 | \theta^n) - q_1)^2] = O((n+1)^{-2w}), \quad w > \alpha.$$

このような状態では定理3.2の結果が成立する。

§4. Self-learning によるパターン分類

ここでは§3と同じ type のモデルについて考える。ただしこの場合には category の集合$\Theta = \{\theta_1, \theta_2, \cdots, \theta_s\}$についてもきちんとした知識はなく，training sequence (θ_iについての情報を含む)も利用できない。この場合には，各時点で prob. density fn. $p(x^n | \xi^{n-1}, d^{n-1}) = \sum_{\theta_i \in \Theta} p(x^n, \theta_i | \xi^{n-1}, d^{n-1})$のモードに一つの category を対応させるようにする。

従ってここでは各時点$n+1$で与えられた orthonormal function の集合$\{\varphi_i^{(n+1)}(x^{n+1})\}_{i=1}^N$の一次式で未知な関数$p(x^{n+1} | \xi^n, d^n)$を近似することを考える。この方法は各時点$n+1$で

$$J_{n+1} = \int_{X_{n+1}} [p(x^{n+1} | \xi^n, d^n) - \sum_{i=1}^N c_i^{(n+1)}(\xi^n, d^n) \varphi_i^{(n+1)}(x^{n+1})]^2 dx^{n+1}$$

を最小にする $\{C_{j*}^{(n+1)}(\xi^n, d^n)\}_{j=1}^{N}$ を求める問題となる。

今 $\dfrac{\partial J_{n+1}}{\partial C_j^{(n+1)}} = 0$ として

$$C_{j*}^{(n+1)}(\xi^n, d^n) = E[\varphi_j^{(n+1)}(x^{n+1}) | \xi^n, d^n] = \sum_{i=1}^{N} \int_{X_{n+1}} \varphi_j^{(n+1)}(x^{n+1})\, p(x^{n+1}, \theta_i | \xi^n, d^n)\, dx^{n+1}$$

を得る。

上の議論から "*modified stochastic approximation*" の応用として次の様なアルゴリズムが考えられる。

Algorithm

最初に 観測される x^1 より次の様な型で $C_j^{(1)}(\xi^1, d^1)$ を構成する：

$$C_j^{(1)}(\xi^1, d^1) = C_j^{(0)} + \gamma_1(\varphi_j^{(1)}(x^1) - C_j^{(0)}) \quad , j=1,2,\cdots,N$$

ただし $C_j^{(0)} = 0 \quad (j=1,2,\cdots,N)$.

次に 観測される x^2 より各 j について 次の様な型で $C_j^{(2)}(\xi^2, d^2)$ を構成する；

$$C_j^{(2)}(\xi^2, d^2) = C_j^{(1)}(\xi^1, d^1) + \gamma_2(\varphi_j^{(2)}(x^2) - C_j^{(1)}(\xi^1, d^1))$$

一般には 観測される x^n より 各 j について 次の様な型で $C_j^{(n)}(\xi^n, d^n)$ を構成する：

$$C_j^{(n)}(\xi^n, d^n) = C_j^{(n-1)}(\xi^{n-1}, d^{n-1}) + \gamma_n(\varphi_j^{(n)}(x^n) - C_j^{(n-1)}(\xi^{n-1}, d^{n-1})) \quad .$$

ただし上のアルゴリズムで用いられている非負の実数列 $\{\gamma_n\}_{n=1}^{\infty}$ は次の様な性質を満している：

$$\sum_{n=1}^{\infty} \gamma_n = \infty \quad , \quad \sum_{n=1}^{\infty} \gamma_n^2 < \infty \; .$$

このとき $\varphi_j^{(n)}(\xi^n, d^n)$ と $\varphi_{j*}^{(n)}(\xi^{n-1}, d^{n-1})$ について以下の様な定理が成立する。

定理4.1　次の様な条件を満している非負の実数列 $\{M_n\}_{n=1}^{\infty}$ と 3つの非負の実数 K_1, K_2, K_3 が存在する：

(i)　$P[\ (\theta_j^{(n)})^2 \leqq \gamma_{n+1} M_n\ ] = 1 \qquad \forall n, j$

(ii)　$\sum_{n=1}^{\infty} M_n < \infty$

(iii)　$\mathrm{Var}[\ \varphi_j^{(n+1)}(x^{n+1}) \mid \xi^n, d^n\ ] \leqq K_1 (u_j^{(n)})^2 + K_2 (\theta_j^{(n)})^2 + K_3$

このとき次の様な事柄が成立する

$$P[\ \lim_{n\to\infty} u_j^{(n)} = 0\ ] = 1, \ \forall j \quad , \quad \lim_{n\to\infty} E[(u_j^{(n)})^{2\beta}] = 0 \quad 0 < \overset{\forall}{\beta} \leqq 1, \ \forall j$$

更に平均収束の order については次の様な定理が成立する.

定理4.2　定理4.1 の条件(iii)の他に

(i)　$\gamma_n = a/n^{\alpha} \quad , \ a > 0 \ , \quad (1/2) < \alpha < 1$

(ii)　$E[\ (\theta_j^{(n)})^2\ ] = O(n^{-2\omega}) \qquad , \ \omega > \alpha$.

このとき次の様な事柄が成立する：

$$E[\ (u_j^{(n)})^2\ ] = \begin{cases} O(n^{-2(\omega-\alpha)}) & , \ \omega < (3/2)\alpha \\ O(n^{-\alpha}) & , \ \omega \geqq (3/2)\alpha \end{cases} .$$

例4.1

(i) パターン空間 $X_n = X \quad \forall n, \quad \Theta = \{\theta_1, \theta_2, \cdots, \theta_s\}$,

(ii)　$p(x^{n+1}, \theta_i \mid \xi^n, d^n) = g^{(n+1)}(\theta_i \mid \theta^n) f_i(x^{n+1}) \qquad \forall i$,

(iii) $\sum_{\theta_i \in \Theta} q^{(n+1)}(\theta_i | \theta^n) = 1$,

(iv) $\int_X f_{\theta_i}(x)\,dx = 1 \quad \forall i$,

(v) 次の様な条件をみたす実数の集合 $\{q_i\}_{i=1}^{s}$ $(1 \geqq q_i \geqq 0, \sum_{i=1}^{s} q_i = 1)$ が存在する

$$P[\,|q^{(n+1)}(\theta_i | \theta^n) - q_i| \leqq C/(n+1)^{1+\alpha}\,] = 1 \qquad \forall i, n$$

ただし C, α は或る正数である。

上の様な特別なモデルに対して，更に

(vi) $\gamma_n = 1/n$

(vii) $E_{\theta_i}[\varphi_j^2(x)] = \int_X \varphi_j^2(x) f_i(x)\,dx < \infty \qquad \forall i, j$

$$(\varphi_j^{(1)}(x) = \varphi_j^{(2)}(x) = \cdots = \varphi_j^{(n)}(x) = \cdots)$$

なる orthornomal fnの集合 $\{\varphi_i(x)\}_{i=1}^{N}$ が存在する.

上の様な状態では定理4.1の結果が成立する。

<u>例4.2</u> 例4.1の中の条件 (i) (ii) (iii) (iv) (vii) の他に

(i) $\gamma_n = a/n^{\alpha}$, $a > 0$, $(1/2) < \alpha < 1$

(ii) 次の様な条件を満す実数の集合 $\{q_i\}_{i=1}^{s}$ $(1 \geqq q_i \geqq 0, \sum_{i=1}^{s} q_i = 1)$ が存在する:

$$E[\,(q^{(n+1)}(\theta_i | \theta^n) - q_i)^2\,] = O((n+1)^{-2\omega}), \ \omega > \alpha, \ \forall i.$$

このとき定理4.2の結果が成立する。

§.5 未知関数の構成

ある object の特性が関数であらわされる場合に，その input と output を観測することより，その関数を構成することを第

一章で述べた。これを更に次の様に一般化する：

(1) 各時点で input 空間 X の中に観測される x^n は $x^1, x^2, \cdots x^{n-1}$ の結果に従属する．

(2) 各時点での未知関数 $f^{(n)}(x)$ が時間と共に少しずつ変化しながら安定する．

従ってモデルは次の様な型になる．

time	1	2	……	n	…
input space	X	X	……	X	…
input	x^1	x^2	……	x^n	…
output	y^1 $(=f^{(1)}(x^1))$	y^2 $(=f^{(2)}(x^2))$	……	y^n $(=f^{(n)}(x^n))$	…
transition prob. density fn (未知)	$p(x^1)$	$p(x^2 \mid x^1)$	……	$p(x^n \mid \xi^{n-1})$ $\xi^{n-1}=(x^1, x^2, \cdots, x^n)$	…

我々の問題では各時点での prob. density fn. は未知であるが，観測可能な input と output の列 $(x^1, y^1)\ (x^2, y^2) \cdots$ を用いて，各時点 $n+1$ で与えられた linearly independent continuous function の集合 $\{\varphi_i^{(n+1)}(x)\}_{i=1}^{N}$ の一次式で近似することを考える。

この方法では，各時点 $n+1$ で

$$I_{n+1} = E\left[\left(y^{n+1} - \sum_{i=1}^{N} c_i^{(n+1)}(\xi^n)\,\varphi_i^{(n+1)}(x^{n+1})\right)^2 \,\middle|\, \xi^n\right]$$

$$= E\left[\left(y^{n+1} - \mathbb{C}^{(n+1)}(\xi^n)\,\varphi^{(n+1)}(x^{n+1})\right)^2 \,\middle|\, \xi^n\right]$$

を最小にする vector $\mathbb{C}_*^{(n+1)}(\xi^n)$ を求める問題になる，ただし

$$\mathbb{C}^{(n+1)}(\xi^n)=(c_1^{(n+1)}(\xi^n),\ c_2^{(n+1)}(\xi^n),\ \cdots\cdots,\ c_N^{(n+1)}(\xi^n))$$

$$\varphi^{(n+1)}(x^{n+1})=(\varphi_1^{(n+1)}(x^{n+1}),\ \varphi_2^{(n+1)}(x^{n+1}),\ \cdots,\ \varphi_N^{(n+1)}(x^{n+1}))'\ 。$$

今 $\nabla_{\mathbb{C}^{(n+1)}} I_{n+1}=0$ より

$E[(y^{n+1}-\mathbb{C}_*^{(n+1)}(\xi^n)\varphi^{(n+1)}(x^{n+1}))\varphi^{(n+1)'}(x^{n+1})\mid\xi^n]=0$ が得られる.

上の議論から"*modified stochastic approximation*"の応用として次の様なアルゴリズムが考えられる。

<u>Algorithm</u>

最初に観測されたinput x^1とoutput y^1の対(x^1,y^1)を用いて次の様な型でvector $\mathbb{C}^{(1)}(\xi^1)$を構成する：

$$\mathbb{C}^{(1)}(\xi^1)=\mathbb{C}^{(0)}+\gamma_1(y^1-\mathbb{C}^{(0)}\varphi^{(1)}(x^1))\varphi^{(1)'}(x^1)$$

ただし　$\mathbb{C}^{(0)}\equiv 0$.

次に観測されたinput x^2とoutput y^2の対(x^2,y^2)を用いて次の様な型でvector $\mathbb{C}^{(2)}(\xi^2)$を構成する：

$$\mathbb{C}^{(2)}(\xi^2)=\mathbb{C}^{(1)}(\xi^1)+\gamma_2(y^2-\mathbb{C}^{(1)}(\xi^1)\varphi^{(2)}(x^2))\varphi^{(2)'}(x^2)\ .$$

一般には観測されたinput x^nとoutput y^nの対(x^n,y^n)を用いて次の様な型でvector $\mathbb{C}^{(n)}(\xi^n)$を構成する：

$$\mathbb{C}^{(n)}(\xi^n)=\mathbb{C}^{(n-1)}(\xi^{n-1})+\gamma_n(y^n-\mathbb{C}^{(n-1)}(\xi^{n-1})\varphi^{(n)}(x^n))\varphi^{(n)'}(x^n)$$

ただし上のアルゴリズムで用いられている非負の実数列$\{\gamma_n\}_{n=1}^{\infty}$は次の様な性質を満している：

$$\sum_{n=1}^{\infty}\gamma_n=\infty\quad,\quad\sum_{n=1}^{\infty}\gamma_n^2<\infty\ 。$$

このとき$\mathbb{C}^{(n)}(\xi^n)$と$\mathbb{C}_*^{(n)}(\xi^{n-1})$に関して次の様な定理が成立する.

定理5.1 全ての x^{n+1}, ξ^n に対して次の様な条件が成立する。

(1) $p(x^{n+1}|\xi^n) > 0$

(2) $E[|f^{(n+1)}(x^{n+1})\varphi_i^{(n+1)}(x^{n+1})|\,|\xi^n] \leqq M_1 \qquad \forall i, j, n$

$E[|\varphi_i^{(n+1)}(x^{n+1})\varphi_j^{(n+1)}(x^{n+1})|\,|\xi^n] \leqq M_2$

ただし M_1, M_2 はそれぞれ或る実数である。

(3) $A^{(n+1)}(\xi^n) = E[\varphi^{(n+1)}(x^{n+1})\varphi^{(n+1)\prime}(x^{n+1})|\xi^n]$ の最小の固有値 $k_0(\xi^n)$ に対して，次の様な k_0 が存在する

$$0 < k_0 \leqq k_0(\xi^n) \quad .$$

(4) $E[\|Y^{(n+1)} - E[Y^{(n+1)}|\xi^n]\|^2|\xi^n] \leqq K_1\|U^{(n)}\|^2 + K_2\|\Theta^{(n)}\|^2 + K_3$,

ただし $U^{(n)} = \mathbb{C}^{(n)}(\xi^n) - \mathbb{C}_*^{(n)}(\xi^{n-1})$, $\Theta^{(n)} = \mathbb{C}_*^{(n)}(\xi^{n-1}) - \mathbb{C}_*^{(n+1)}(\xi^n)$

$$Y^{(n+1)} = (y^{(n+1)} - \mathbb{C}^{(n)}(\xi^n)\varphi^{(n+1)}(x^{n+1}))\varphi^{(n+1)}(x^{n+1})'$$

(5) $P[\|\Theta^{(n)}\|^2 \leqq \gamma_{n+1} M_n] = 1$, $\sum_{n=1}^{\infty} M_n < \infty$.

このとき次の様な事柄が成立する：

$$P[\lim_{n\to\infty} U^{(n)} = 0] = 1, \quad \lim_{n\to\infty} E[\|U^{(n)}\|^{2\beta}] = 0, \quad 0 < {}^{\forall}\beta < 1.$$

更に平均収束の order については次の様な定理が成立する。

定理5.2 定理5.1の条件 (1)(2)(3)(4) の他に

(1)′ $\gamma_n = a/n^{\alpha}$, $a > 0$, $1/2 < \alpha < 1$,

(2)′ $E[\|\Theta^{(n)}\|^2] = O(n^{-2\omega})$, $\omega > \alpha$.

このとき次の様な事柄が成立する：

$$E[\|U^{(n)}\|^2] = \begin{cases} O(n^{-2(\omega-\alpha)}) & , \ \omega < (3/2)\alpha \\ O(n^{-\alpha}) & , \ \omega \geqq (3/2)\alpha . \end{cases}$$

更に input に対応して観測される output が input に関係している noise をもつ場合にも，noise の平均が 0 で分散が有界ならば同様のアルゴリズムで同じ問題がとかれる。

参 考 文 献

[1] Браверман, Э.М. и Розоноэр, Л.И.: Сходимость случайных процессов в теории обучения машин I, Автоматика и Телемеханика, № 1, (1969), 57–77.

[2] Браверман, Э.М. и Розоноэр, Л.И.: Сходимость случайных процессов в теории обучения машин. II, Автоматика и Телемеханика, № 3, (1969), 87–103.

[3] Chung, K.L.: On a stochastic approximation method, Ann. Math. Stat., vol. 25, (1954), 463–483.

[4] Dupač, V.: A dynamic stochastic approximation method, Ann. Math. Stat., vol. 36, (1965), 1695–1702.

[5] Kitagawa, T.: Successive process of statistical controls I, Mem. Fac. Sci. Kyshu Univ., Ser. A., vol. 7, (1952), 13–28.

[6] Kitagawa, T.: Successive process of statistical controls II, Mem. Fac. Sci. Kyush Univ., Ser. A., vol. 13, (1959), 1—16.

[7] Schumpert, J. M. and Yau, S. S.: Design of pattern classifiers with the updating property using stochastic approximation techniques, IEEE Trans. Computers, vol. C—17, No. 9, (1968), 861—872.