

# 確率的フィルタリングを用いたアンサンブル学習の 統計力学的解析

## Statistical Mechanics of Ensemble Learning using Stochastic Filtering

三好 誠司\*  
Seiji MIYOSHI

岡田 真人†  
Masato OKADA

**Abstract:** Ensemble learning using stochastic filtering was proposed by Murata as an on-line version of Schapire's algorithm. The algorithm stochastically uses negative examples to maintain diversity of students. We analyze ensemble learning using the stochastic filtering with statistical mechanical method based on online learning theory. Analyzing a model composed of a nonmonotonic teacher and three students which are updated with Hebbian rule, we show that the stochastic filtering is effective to keep students' diversity rich and to improve the generalization performance. In addition, we show that the students' symmetric breaking occurs if a probability of negative example is larger than 0.5 and the dimension is finite.

*keywords:* ensemble learning, online learning, generalization error, stochastic filtering, boosting

### 1 まえがき

学習はバッチ学習とオンライン学習 [1] に大別できる。バッチ学習においては与えられたいくつかの例題を繰り返し使用する。この場合、生徒が適切な自由度を持っていればすべての例題に正しく答えられるようになるが、それまでに長い時間が必要である。また、多くの例題を蓄えておくメモリが必要である。これに対してオンライン学習では一度使った例題は捨ててしまう。この場合、過去に使った例題に対して生徒が必ず正しく答えられるとは限らないが、多くの例題を蓄えておくためのメモリが不要であり、また時間的に変化する教師にも追従できるなどの利点がある。

一方、精度の低いルールや学習機械（以後は生徒と呼

ぶ）を複数組み合わせることにより精度の高い予測や分類を行うことはアンサンブル学習と呼ばれ注目されている [2, 3, 4]。アンサンブル学習の汎化能力を統計力学的手法によって理論的に解析する研究もさかに行われている [5, 6, 7, 8]。アンサンブル学習においては生徒集団の多様性を維持することが重要である。Schapire は生徒の多様性を維持するためにそれぞれの生徒が用いる例題を確率的なフィルタリングにより選択するアルゴリズムを提案した [10]。Murata は Schapire のアルゴリズムをオンライン学習に適した形に修正したアルゴリズムを提案している [11]。このアルゴリズムは生徒の多様性を維持するために、わざと間違った例題（負例）も用いる点が特徴的である。

本論文では Murata のアルゴリズムの効果を理論的に明らかにするため、オンライン学習の枠組みで統計力学的手法を用いた解析を行う。非単調な出力特性を有する教師 [12, 13, 14] を単純パーセプトロンの生徒集団がヘブ則で学習するモデルについて解析を行った結果、確率的フィルタリングの効果により生徒集団の多様性が維持され、汎化能力が大きく改善されることが明らかになった。また、次元  $N$  が有限の場合には負例を用いる確率が 0.5 より大きいと生徒の対称性の破れが生じることがわかった。

\*神戸高専電子工学科, 651-2194 神戸市西区学園東町 8-3, tel. 078-795-3247, e-mail miyoshi@kobe-kosen.ac.jp,  
Department of Electronic Engineering, Kobe City College of Technology, 8-3 Gakuenhigashimachi, Nishi-ku, Kobe-shi, 651-2194 Japan

†東京大学大学院 新領域創成科学研究科 複雑理工学専攻, 〒 277-8561 千葉県柏市柏の葉 5-1-5,  
Division of Transdisciplinary Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba, 277-8561 Japan  
理化学研究所 脳科学総合研究センター, 〒 351-0198 埼玉県和光市広沢 2-1  
RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama, 351-0198 Japan

## 2 モデル

本論文では1個の教師と3個の生徒を考える．教師は結合荷重  $B$  でしきい値  $a$  の非単調なパーセプトロン，生徒は結合荷重  $J$  の単純パーセプトロンであるとする．すなわち，教師の出力  $d$  は，

$$d = \text{sgn}((v - a)v(v + a)), \quad (1)$$

$$v = B \cdot x, \quad (2)$$

であり，各生徒の出力は  $\text{sgn}(u_1 l_1)$ ,  $\text{sgn}(u_2 l_2)$ ,  $\text{sgn}(u_3 l_3)$  である．ここで，

$$\text{sgn}(ul) = \begin{cases} +1, & ul \geq 0, \\ -1, & ul < 0, \end{cases} \quad (3)$$

$$u_k l_k = J_k \cdot x, \quad (4)$$

である． $l_k$  は生徒  $J_k$  の長さであり，これは本論文で扱う巨視的変数のひとつであるが詳しくは後で述べる．また， $v$  を教師の内部状態， $u_k$  を各生徒の規格化内部状態と呼ぶことにする．なお，本論文では簡単のため教師の結合荷重，生徒の結合荷重のことをそれぞれ単に教師，生徒と呼ぶことにする．教師  $B = (B_1, \dots, B_N)$ ，生徒  $J_k = (J_{k1}, \dots, J_{kN})$ ,  $k = 1, 2, 3$  および入力  $x = (x_1, \dots, x_N)$  は  $N$  次元ベクトルであり， $B$  の各要素  $B_i$  は平均0，分散1のガウス分布にしたがい独立に生成され，不変であるとする． $J_k$  の初期値  $J_k^0$  の各要素  $J_{ki}^0$  は平均0，分散1のガウス分布にしたがい独立に生成されるものとし， $B$  と  $J_k$  の方向余弦は  $R_k$ ， $J_k$  と  $J_{k'}$  の方向余弦は  $q_{kk'}$  であるとする．また， $x$  の各要素  $x_i$  は平均0，分散  $1/N$  のガウス分布にしたがい独立に生成されるものとする．以上より，

$$\langle B_i \rangle = 0, \quad \langle (B_i)^2 \rangle = 1, \quad (5)$$

$$\langle J_{ki} \rangle = 0, \quad \langle (J_{ki})^2 \rangle = 1, \quad (6)$$

$$\langle x_i \rangle = 0, \quad \langle (x_i)^2 \rangle = \frac{1}{N}, \quad (7)$$

$$R_k = \frac{B \cdot J_k}{\|B\| \|J_k\|}, \quad q_{kk'} = \frac{J_k \cdot J_{k'}}{\|J_k\| \|J_{k'}\|}. \quad (8)$$

ここで， $\langle \cdot \rangle$  は平均を表す．

本論文では， $N \rightarrow \infty$  の熱力学的極限を考えることにする．このとき，

$$\|B\| = \sqrt{N}, \quad \|J_k\| = \sqrt{N}, \quad \|x\| = 1, \quad (9)$$

となる．生徒のノルム  $\|J\|$  は一般には時間の経過とともに変化するが，初期値  $\sqrt{N}$  に対する比を  $l^m$  とし，生徒の長さと呼ぶことにする．すなわち， $\|J^m\| = l^m \sqrt{N}$  である．ここで， $m$  は時刻ステップである．

教師と個々の生徒には共通の入力  $x$  が同じ順序で与えられる．個々の生徒は入力  $x$  に対する教師の出力と自分の出力を比べ，教師と同じ出力を出す確率が上がるように，必要に応じて自分の結合荷重を修正していく．この手続きを学習と呼ぶ．修正の方法は学習則と呼ばれ，ヘブ学習，パーセプトロン学習，アダルトン学習がよく知られている [9, 12, 13]．自分自身に関する情報以外に生徒が修正のために使える情報は，入力  $x$  とそれに対する教師の出力  $d$  だけであるから，学習は一般に  $J_k^{m+1} = J_k^m + f(d^m, u_k^m) x^m$  と書ける．

本論文の記法を用いた場合，Murata の確率的フィルタリングのアルゴリズム [11] は以下の通りである．

**step0:** three different classifiers  $J_1, J_2, J_3$

**step1:** pick an example  $(x, d)$

**step2:** choose  $i \in \{1, 2, 3\}$  randomly

**step3:** calculate  $H(x) = \sum_{j \neq i} \text{sgn}(J_j \cdot x)$

**step4:** if  $H(x) = 0$ , train  $J_i$  with  $(x, d)$   
otherwise, flip a fair/unfair coin

**head:** train  $J_i$  with  $(x, d)$

**tail:** train  $J_i$  with  $(x, -d)$

**step5:** goto step1 until some condition is fulfilled

step4 の tail の場合が，生徒の多様性を維持するために，わざと間違った例題（負例）を与えることを表している．このアルゴリズムによる更新は，生徒  $J_1$  を例にとると以下のように書ける．

$$J_1^{m+1} = J_1^m + \Theta(-u_2^m u_3^m) f_1 x^m + \Theta(u_2^m u_3^m) (s^m f_1^m + (1 - s^m) g_1^m) x^m, \quad (10)$$

$$P(s^m = 1) = 1 - P(s^m = 0) = b. \quad (11)$$

ここで， $f$  は学習則により決定される関数であり， $g$  は負例を用いる場合の関数である．また  $b$  は step4 で head となる確率である．ここで， $\Theta(\cdot)$  は以下のようなステップ関数である．

$$\Theta(z) = \begin{cases} +1, & z \geq 0, \\ 0, & z < 0. \end{cases} \quad (12)$$

なお，本論文においては統計力学的な取り扱いを容易にするため，step2 は以下とする．

step2:  $i = m \pmod{3} + 1$

すなわち本論文においては3個の生徒が順番に更新を行うものとする。

### 3 理論

#### 3.1 汎化誤差

統計的学習理論の目的のひとつは汎化誤差  $\epsilon_g$  を理論的に求めることである。本論文では、3個の生徒は多数決で集団としての出力を決定するものとする。よって、誤差  $\epsilon$  として、

$$\epsilon = \Theta \left( -d \sum_{k=1}^3 \text{sgn}(\mathbf{J}_k \cdot \mathbf{x}) \right) \quad (13)$$

$$= \Theta \left( -\text{sgn}((v-a)v(v+a)) \sum_{k=1}^3 \text{sgn}(u_k l_k) \right), \quad (14)$$

を用いることにする。

汎化誤差  $\epsilon_g$  は誤差  $\epsilon$  を入力  $\mathbf{x}$  の確率分布  $p(\mathbf{x})$  で平均したものと定義する。ここで誤差  $\epsilon$  は生徒の規格化内部状態  $\{u_k\}$  と教師の内部状態  $v$  を用いて  $\epsilon(\{u_k\}, v)$  と書くことができるので、 $\epsilon_g$  は確率分布  $p(\{u_k\}, v)$  を用いて、

$$\epsilon_g = \int d\mathbf{x} p(\mathbf{x}) \epsilon \quad (15)$$

$$= \int \left( \prod_{k=1}^3 du_k \right) dv p(\{u_k\}, v) \epsilon(\{u_k\}, v), \quad (16)$$

と書ける。 $\{u_k\}, v$  は  $\{\mathbf{J}_k\}, \mathbf{B}$  とそれらとは独立な入力  $\mathbf{x}$  で書けるので  $p(\{u_k\}, v)$  は平均0の多重ガウス分布であり、その共分散行列  $\Sigma$  は  $R_k$  と  $q_{kk'}$  を用いて書くことができる。すなわち、

$$p(\{u_k\}, v) = \frac{1}{(2\pi)^2 |\Sigma|^{\frac{1}{2}}} \exp \left( -\frac{(\{u_k\}, v) \Sigma^{-1} (\{u_k\}, v)^T}{2} \right), \quad (17)$$

$$\Sigma = \begin{pmatrix} 1 & q_{12} & q_{13} & R_1 \\ q_{21} & 1 & q_{23} & R_2 \\ q_{31} & q_{32} & 1 & R_3 \\ R_1 & R_2 & R_3 & 1 \end{pmatrix}. \quad (18)$$

#### 3.2 巨視的変数の微分方程式

式(16)–(18)より、汎化誤差  $\epsilon_g$  は巨視的変数  $R_k, q_{kk'}$  がすべてわかれば計算できる。そこで、これらの巨視的変数のダイナミクスを記述する連立微分方程式を熱力学的極限における自己平均性に基づき式(11)から決定論

的な形で導出した。結果を次頁の式(19)–(21)に示す。ここで  $r_k \equiv l_k R_k$  と  $Q_{kk'} \equiv l_k l_{k'} q_{kk'}$  は解析を容易にするために導入した補助的な巨視的変数である。 $t$  は時刻ステップ  $m$  を次元  $N$  で正規化した時刻  $t = m/N$ 、 $\langle \cdot \rangle$  はサンプル平均を表す。

### 4 結果と考察

以下では3個の生徒の対称性を仮定するとともに、それらがすべてヘブ則を用いる場合を考える。この場合、 $f_k = -g_k = d$  である。いま、式(1)の非単調教師を考えているから、

$$f_k = \text{sgn}((v-a)v(v+a)), \quad (22)$$

$$g_k = -\text{sgn}((v-a)v(v+a)), \quad (23)$$

となる。また、本論文では生徒  $\mathbf{J}_k$  の初期値  $\mathbf{J}_k^0$ 、教師  $\mathbf{B}$  の各要素は平均0、分散1のガウス分布にしたがい独立に生成され、また、 $N \rightarrow \infty$  の熱力学的極限を考えているので、初期状態においてこれらはすべて直交しており、

$$R_k^0 = 0, \quad q_{kk'}^0 = 0, \quad (24)$$

である。式(24)と生徒の対称性より式(19)–(21)の中の15個のサンプル平均  $A_1 \sim A_{15}$  には以下の関係がある。

$$A_1 = \langle \Theta(-u_2 u_3) f_1 u_1 \rangle \quad (25)$$

$$A_2 = \langle \Theta(u_2 u_3) f_1 u_1 \rangle \quad (26)$$

$$A_3 = \langle \Theta(u_2 u_3) g_1 u_1 \rangle = -A_2 \quad (27)$$

$$A_4 = \langle \Theta(-u_2 u_3) f_1^2 \rangle \quad (28)$$

$$A_5 = \langle \Theta(u_2 u_3) f_1^2 \rangle = 1 - A_4 \quad (29)$$

$$A_6 = \langle \Theta(u_2 u_3) g_1^2 \rangle = 1 - A_4 \quad (30)$$

$$A_7 = \langle \Theta(-u_2 u_3) f_1 v \rangle \quad (31)$$

$$A_8 = \langle \Theta(u_2 u_3) f_1 v \rangle \quad (32)$$

$$A_9 = \langle \Theta(u_2 u_3) g_1 v \rangle = -A_8 \quad (33)$$

$$A_{10} = \langle \Theta(-u_1 u_3) f_2 u_1 \rangle \quad (34)$$

$$A_{11} = \langle \Theta(u_1 u_3) f_2 u_1 \rangle \quad (35)$$

$$A_{12} = \langle \Theta(u_1 u_3) g_2 u_1 \rangle = -A_{11} \quad (36)$$

$$A_{13} = \langle \Theta(-u_2 u_3) f_1 u_2 \rangle = A_{10} \quad (37)$$

$$A_{14} = \langle \Theta(u_2 u_3) f_1 u_2 \rangle = A_{11} \quad (38)$$

$$A_{15} = \langle \Theta(u_2 u_3) g_1 u_2 \rangle = -A_{11} \quad (39)$$

さらに生徒の対称性から式(19)–(21)の巨視的変数  $l_k, r_k, Q_{kk'}$  から添え字  $k, k'$  を落としてそれぞれを  $l, r, Q$  と書くことにする。このとき式(19)–(21)は以下のよう

$$\frac{dl_1}{dt} = \frac{1}{3} \left( \langle \Theta(-u_2 u_3) f_1 u_1 \rangle + b \langle \Theta(u_2 u_3) f_1 u_1 \rangle + (1-b) \langle \Theta(u_2 u_3) g_1 u_1 \rangle \right) + \frac{1}{6l_1} \left( \langle \Theta(-u_2 u_3) f_1^2 \rangle + b \langle \Theta(u_2 u_3) f_1^2 \rangle + (1-b) \langle \Theta(u_2 u_3) g_1^2 \rangle \right), \quad (19)$$

$$\frac{dr_1}{dt} = \frac{1}{3} \left( \langle \Theta(-u_2 u_3) f_1 v \rangle + b \langle \Theta(u_2 u_3) f_1 v \rangle + (1-b) \langle \Theta(u_2 u_3) g_1 v \rangle \right), \quad (20)$$

$$\frac{dQ_{12}}{dt} = \frac{l_1}{3} \left( \langle \Theta(-u_1 u_3) f_2 u_1 \rangle + b \langle \Theta(u_1 u_3) f_2 u_1 \rangle + (1-b) \langle \Theta(u_1 u_3) g_2 u_1 \rangle \right) + \frac{l_2}{3} \left( \langle \Theta(-u_2 u_3) f_1 u_2 \rangle + b \langle \Theta(u_2 u_3) f_1 u_2 \rangle + (1-b) \langle \Theta(u_2 u_3) g_1 u_2 \rangle \right). \quad (21)$$

になり，計算すべきサンプル平均は  $A_1, A_7, A_{10}$  の 3 個であることがわかる．

$$\frac{dl}{dt} = \frac{1}{3} A_1 + \frac{1}{6l} \quad (40)$$

$$\frac{dr}{dt} = \frac{1}{3} A_7 \quad (41)$$

$$\frac{dQ}{dt} = \frac{2l}{3} A_{10} \quad (42)$$

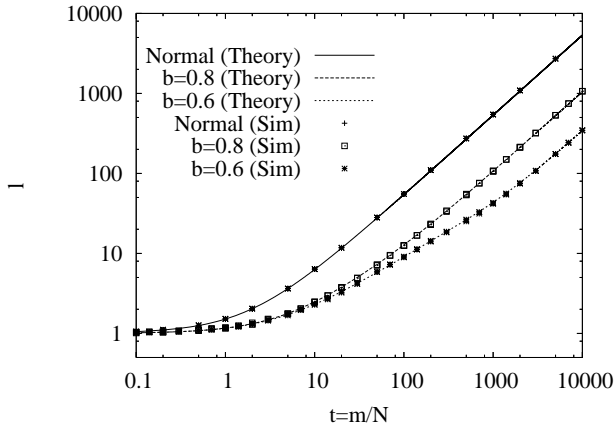


図 1:  $l$  のダイナミクス ( $b = 0.8, 0.6$ )

非単調教師のしきい値が  $a = 0.6$  の場合について理論的に求めた  $l, R, q, \epsilon_g$  のダイナミクスを計算機シミュレーションの結果と重ねて図 1-8 にそれぞれ示す．いずれの図においても，曲線が理論計算の結果を表し，+, × 等の印が計算機シミュレーションの結果を表す．また，比較の対象とするため通常のアンサンブル学習の結果を“Normal”として重ねている．通常のアンサンブル学習においては 3 個の生徒が同じ例題を使って，他の生徒とは情報交換することなしに独立に更新してゆく．step4 で head となる確率は  $b = 0.8, 0.6, 0.4, 0.2$  の 4 通りについて計算した．理論計算においては，式 (40)–(42) に現れる 3 個のサンプル平均  $A_1, A_7, A_{10}$  をモンテカルロステップ数  $MC = 10^6$  のメトロポリス法で求めながら，ルンゲクッタ法により連立微分方程式を数値的に解いた．得られた  $R, q$  を用いて式 (16) の積分を再度メトロポリ

ス法 ( $MC = 10^9$ ) で実行することにより  $\epsilon_g$  を求めた．計算機シミュレーションは  $b = 0.8$  と  $b = 0.6$  の場合は  $N = 1000$ ， $b = 0.4$  と  $b = 0.2$  の場合は  $N = 5000$  で実行し， $\epsilon_g$  は各時点で  $10^5$  個のランダム入力でテストを行うことにより求めた．

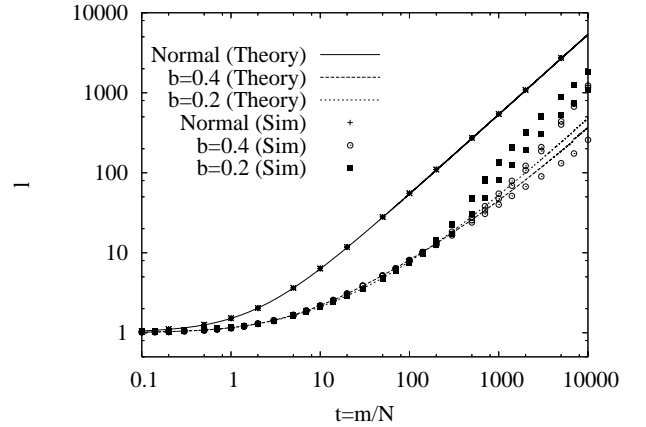


図 2:  $l$  のダイナミクス ( $b = 0.4, 0.2$ )

図 1-8 から， $b = 0.8, 0.6$  の場合には理論と計算機シミュレーションはよく合っているのに対し， $b = 0.4, 0.2$  の場合には，学習の途中までしか合っていないことがわかる．すなわち，図 4，図 6 の計算機シミュレーションにおいて， $b = 0.4$  の場合には  $t = 50$  付近で， $b = 0.2$  の場合には  $t = 10$  付近で，それぞれ生徒の対称性が破れ，それ以後は理論と計算機シミュレーションは合っていない．

$b = 0.8, 0.6$  の計算機シミュレーションでは生徒の対称性が維持されるのに対し， $b = 0.4, 0.2$  では生徒の対称性が破れてしまう理由について考察する．図 4 をよく見ると，生徒の対称性が破れた後，十分な時間が経過すると 3 個の生徒のうち 2 個は  $B$  との方向余弦が 1 に近づき，残りの 1 個は  $-1$  に近づいている．いま， $R_1 \simeq 1, R_2 \simeq 1, R_3 \simeq -1$  の場合を考えると，確率的フィルタリングのアルゴリズムの step4 において  $J_1$  や  $J_2$  の更新を行う場合には  $H(x) = 0$  となり通常のヘブ則で

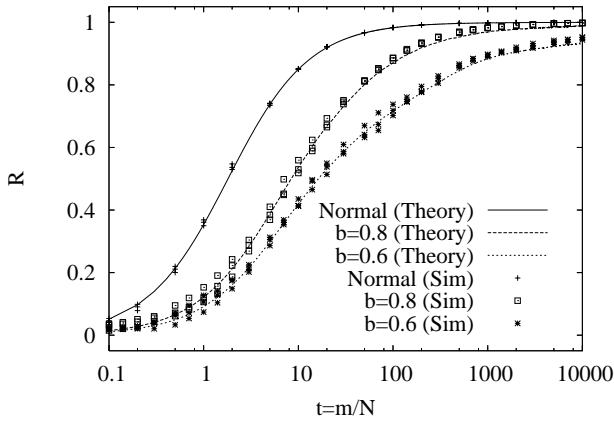


図 3:  $R$  のダイナミクス ( $b = 0.8, 0.6$ )

更新が行われるので  $R_1$  や  $R_2$  はさらに 1 に近づいていく [12, 13]. 一方,  $J_3$  の更新を行う場合には  $H(x) = \pm 2$  となり coin が振られることになる. このとき,  $b < 0.5$  であると, 負例を用いる確率が正例を用いる確率より大きいことになり,  $R_3$  はさらに  $-1$  に近づいていく. このように,  $b < 0.5$  の場合には  $R_1 = R_2 = -R_3 = 1$  は安定平衡点となる. よって,  $b < 0.5$  の場合には次元  $N$  が有限であることによる生徒の対称性の微小な破れが徐々に拡大し, この安定平衡点にまで至っていると理解できる.

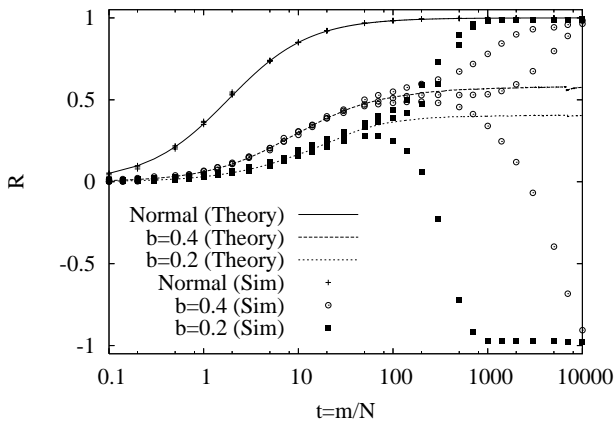


図 4:  $R$  のダイナミクス ( $b = 0.4, 0.2$ )

本論文で扱っているモデルのように教師が非単調な場合, 通常のアサンブル学習では,  $\epsilon_g$  は非単調なダイナミクスを呈する [14]. 図 7, 8 から, 学習の初期を除いては, 確率的フィルタリングにより汎化能力が大きく改善されていることがわかる.

このときの理論計算における  $R$  と  $q$  の関係を図 9 に示す. いずれの曲線も  $R = q = 0$  の端点  $t = 0$  に対応し, もう一方の端点  $t = 10000$  に対応する. 確率的フィルタリングにより,  $R$  に対して  $q$  が非常に小さく抑

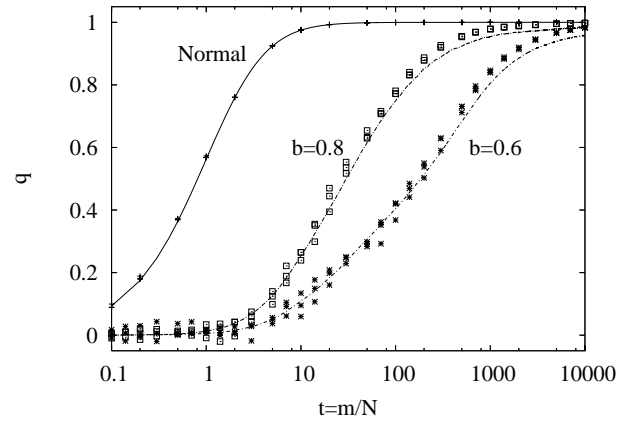


図 5:  $q$  のダイナミクス ( $b = 0.8, 0.6$ )

えられ, 生徒集団の多様性が維持されることがわかる.  $b$  を小さくしていくと  $q$  は小さくなり, 特に  $b = 0.2$  の場合には  $q$  が負になっている点も興味深い.

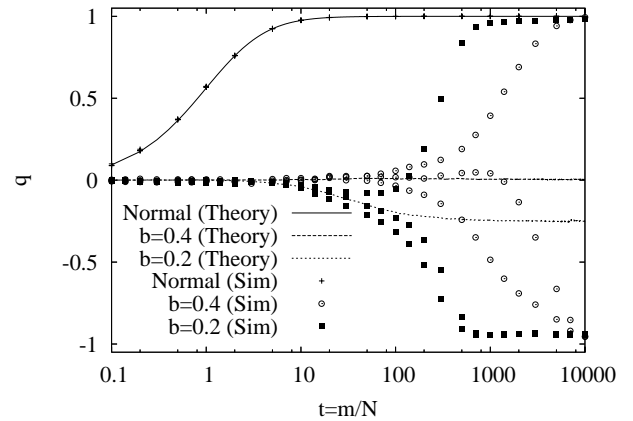


図 6:  $q$  のダイナミクス ( $b = 0.4, 0.2$ )

## 5 むすび

アサンブル学習における Murata の確率的フィルタリングの効果をオンライン学習の枠組みで統計力学的手法を用いて解析した. 教師が非単調な出力特性を有し, 3 個の生徒がヘブ則を用いるモデルについて調べた結果, 確率的フィルタリングにより生徒集団の多様性が維持され, 汎化能力が大きく改善されることが明らかになった. また, 次元  $N$  が有限の場合には負例を用いる確率が 0.5 より大きいと生徒の対称性の破れが生じることがわかった.

## 謝辞

本研究の一部は科学研究費補助金 ( 課題番号 15500151, 18500183 ) によるものであり, ここに感謝いたします.

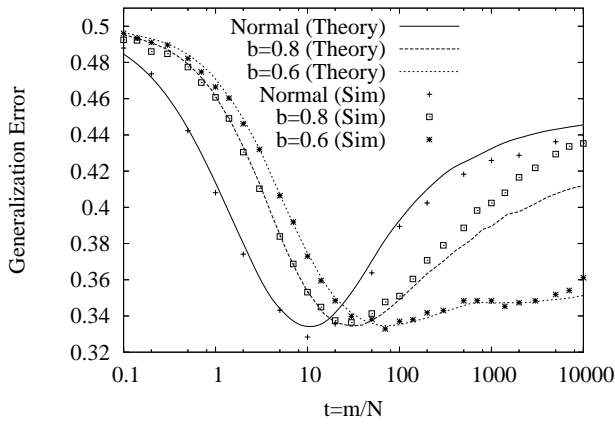


図 7:  $\epsilon_g$  のダイナミクス ( $b = 0.8, 0.6$ )

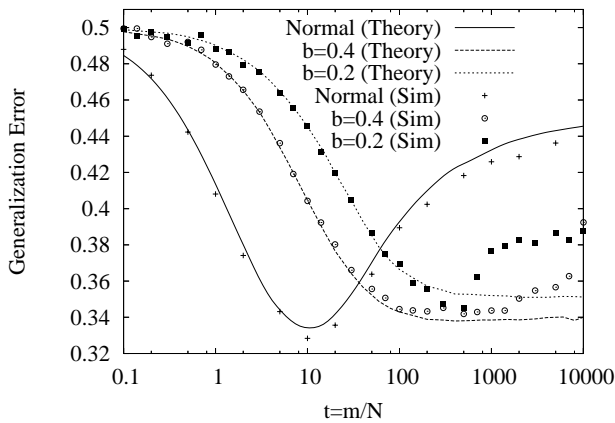


図 8:  $\epsilon_g$  のダイナミクス ( $b = 0.4, 0.2$ )

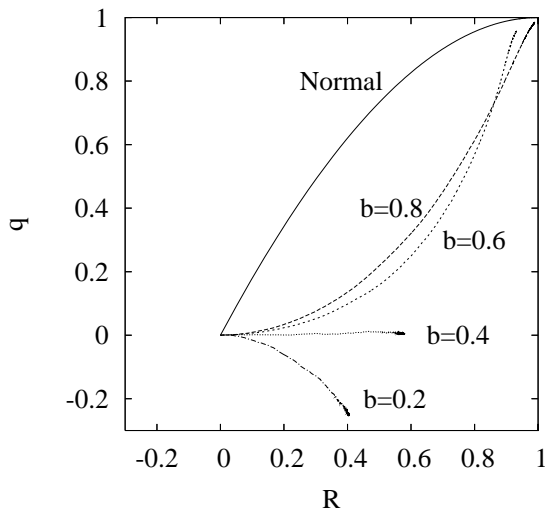


図 9:  $R$  と  $q$  の関係 (理論計算のみ)

## 参考文献

- [1] D. Saad (ed.), On-line Learning in Neural Networks, Cambridge University Press, (1998)
- [2] Y. Freund and R. E. Shapire, (安倍直樹訳), “ブースティング入門,” 人工知能学会誌, **14**(5), 771–780 (1999).
- [3] <http://www.boosting.org/>
- [4] 麻生 英樹, 津田 宏治, 村田 昇, “パターン認識と学習の統計学,” 岩波書店, 東京, 2003.
- [5] A. Krogh and P. Sollich, “Statistical mechanics of ensemble learning,” Phys. Rev. E, **55**(1), 811–825 (1997).
- [6] R. Urbanczik, “Online learning with ensembles,” Phys. Rev. E, **62**(1), 1448–1451 (2000).
- [7] K. Hara and M. Okada, “Ensemble learning of linear perceptron; Online learning theory”, J. Phys. Soc. Jpn. **74**, 2966, (2005).
- [8] S. Miyoshi, K. Hara and M. Okada, “Analysis of ensemble learning using simple perceptrons based on online learning theory.”, Phys. Rev. E, **71**, 036116, (2005).
- [9] 西森 秀俊, “スピングラス理論と情報統計力学,” 岩波書店, 東京, 1999.
- [10] R. E. Schapire, “The strength of weak learnability.”, Machine Learning, **5**, 197–227, (1990).
- [11] N. Murata, “Stochastic filtering for on-line boosting.”, NHC & SMAPIP-2005, Proc. 2005 Workshop on Randomness and Computation, 41–42, (2005).
- [12] J. I. Inoue and H. Nishimori, “On-line AdaTron learning of a unlearnable rules.”, Phys. Rev. E, **55**(4), 4544–4551, (1997).
- [13] J. I. Inoue, H. Nishimori and Y. Kabashima, “On-line learning of non-monotonic rules by simple perceptron,” J. Phys. A: Math. Gen. **30**, 3795–3816, (1997).
- [14] 三好 誠司, 原 一之, 岡田 真人, “教師が非単調な場合のアンサンブル学習”, 信学技報, NC2004-214, 123–128, (2005).