

教師が非単調な場合のアンサンブル学習

三好 誠司[†] 原 一之^{††} 岡田 真人^{†††,††††,†††††}

[†] 神戸市立工業高等専門学校 〒 651-2194 神戸市西区学園東町 8-3

^{††} 東京都立工業高等専門学校 〒 140-0011 東京都品川区東大井 1-10-40

^{†††} 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

^{††††} 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

^{†††††} 科学技術振興機構 さきがけ

E-mail: [†]miyoshi@kobe-kosen.ac.jp, ^{††}hara@tokyo-tmct.ac.jp, ^{†††}okada@brain.riken.go.jp

あらまし アンサンブル学習の大きな特徴として、多数決などで生徒を組み合わせるにより、単一の生徒では表現できない入出力関係を実現できることがあげられる。その意味で、教師が生徒一個のモデル空間内にはないような場合のアンサンブル学習の解析は非常に興味深い。そこで本論文では、教師が非単調なパーセプトロンであり生徒が単純パーセプトロンである場合のアンサンブル学習を統計力学的なオンライン学習の枠組みで議論する。メトロポリス法により汎化誤差を計算した結果、ヘブ学習では学習の初期においてアンサンブルの効果があるものの、やがて生徒の多様性がなくなりアンサンブルの効果も消滅してしまうことがわかった。これに対し、パーセプトロン学習では生徒の多様性は消滅せず、そのために十分時間が経過した後もアンサンブルの効果が残ることがわかった。

キーワード アンサンブル学習, オンライン学習, 非単調教師, 汎化誤差

Analysis of Ensemble Learning for Non-Monotonic Teacher

Seiji MIYOSHI[†], Kazuyuki HARA^{††}, and Masato OKADA^{†††,††††,†††††}

[†] Kobe City College of Technology 8-3 Gakuenhigashimachi, Nishi-ku, Kobe-shi, 651-2194 Japan

^{††} Tokyo Metropolitan College of Technology 1-10-40 Higashi-oi, Shinagawa, Tokyo, 140-0011 Japan

^{†††} Division of Transdisciplinary Sciences, Graduate School of Frontier Sciences, The University of Tokyo 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba, 277-8561 Japan

^{††††} RIKEN Brain Science Institute 2-1 Hirosawa, Wako-shi, Saitama, 351-0198 Japan

^{†††††} JST PRESTO

E-mail: [†]miyoshi@kobe-kosen.ac.jp, ^{††}hara@tokyo-tmct.ac.jp, ^{†††}okada@brain.riken.go.jp

Abstract A major merit of ensemble learning is to realize the input-output relations by combining students that cannot be represented by one student. Therefore, ensemble learning in which a teacher isn't in the model space of one student is very attractive. In this paper ensemble learning, in which a teacher and students are a non-monotonic perceptron and simple perceptrons respectively, is discussed based on online learning theory and statistical mechanics. Hebbian learning doesn't keep the variety of students and the effect of ensemble disappears. On the contrary, perceptron learning keeps the variety of students and the effect of ensemble doesn't disappear.

Key words ensemble learning, online learning, non-monotonic teacher, generalization error

1. ま え が き

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おうとすることは一般にアンサンブル学習と呼ばれ、近年注目されている [1] ~ [3]。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んに行われている [4] ~ [8]。

著者らは [7], [8] において教師が単純パーセプトロンで生徒が K 個の単純パーセプトロンであるようなアンサンブル学習を、オンライン学習の枠組みで議論した。すなわち、まず K 個の生徒が多数決で統合出力を決定する場合の汎化誤差が教師と生徒の類似度と生徒間の類似度という二つの巨視的変数で計算できることを示した。次に、一般の学習則について、これらの巨視的変数のダイナミクスを記述する微分方程式を導出した。さ

らに、よく知られているヘブ学習、パーセプトロン学習、アダプトロン学習の三つの学習則 [9] ~ [11] について、この微分方程式を具体的に導出し、それらを解いた結果を用いて汎化誤差を数値的に求めた。その結果、こらら三つの学習則が「生徒の多様性維持」というアンサンブル学習との相性という点でそれぞれ異なった性質を有しており、アダプトロン学習が、アンサンブル学習との相性という点で最も優れているという興味深い事実が明らかになった。

一方、Inoue, Nishimori, Kabashima は教師が一個の非単調パーセプトロンであり生徒が一個の単純パーセプトロンである場合について解析した [12], [13]。彼らが扱ったモデルは、教師が生徒のモデル空間内にはない場合ということができる。

アンサンブル学習の大きな特徴として、多数決などで生徒を組み合わせるにより、単一の生徒では表現できない入出力関係を実現できることがあげられる [3]。その意味で、教師が生徒一個のモデル空間内にはないような場合のアンサンブル学習の解析 [14] ~ [16] は非常に興味深い。そこで本論文では、教師が一個の非単調パーセプトロンであり、生徒が単純パーセプトロンの集団であるようなアンサンブル学習についてオンライン学習の枠組みで議論する。ヘブ学習とパーセプトロン学習について解析を行った結果、ヘブ学習では学習の初期においてアンサンプルの効果があるものの、やがて生徒の多様性がなくなりアンサンプルの効果も消滅してしまうことがわかった。これに対し、パーセプトロン学習では生徒の多様性は消滅せず、そのために十分時間が経過した後でもアンサンプルの効果が残ることがわかった。

2. モデル

本論文で対象とする生徒は、符号関数を出力関数とするパーセプトロンである。\$K\$ 個の生徒からなるアンサンブルを考え、各生徒の結合荷重を \$J_1, J_2, \dots, J_K\$ とする。\$J_k = (J_{k1}, \dots, J_{kN}), k = 1, 2, \dots, K\$ と入力 \$x = (x_1, \dots, x_N)\$ は \$N\$ 次元ベクトルであり、\$x\$ の各要素 \$x_i\$ は平均 0、分散 \$1/N\$ のガウス分布に従う独立な確率変数であるとする。また、\$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する。すなわち、

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

ここで、\$\langle \cdot \rangle\$ は平均を表す。各生徒の出力は \$\text{sgn}(u_1 l_1), \text{sgn}(u_2 l_2), \dots, \text{sgn}(u_K l_K)\$ である。ここで、

$$\text{sgn}(ul) = \begin{cases} +1, & ul \geq 0, \\ -1, & ul < 0, \end{cases} \quad (2)$$

$$u_k l_k = J_k \cdot x, \quad (3)$$

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、\$u_k\$ を各生徒の規格化内部状態と呼ぶことにする。

教師は非単調パーセプトロンであるとする。その結合荷重 \$B\$ は \$N\$ 次元ベクトルであり、その各要素 \$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。す

なわち、

$$\langle B_i \rangle = 0, \langle (B_i)^2 \rangle = 1. \quad (4)$$

また、教師は以下のような非単調な出力特性を有するものとする。

$$d = \text{sgn}((v-a)v(v+a)), \quad (5)$$

$$v = B \cdot x. \quad (6)$$

\$v\$ を教師の内部状態と呼ぶことにする。なお、簡単のため以後は生徒の結合荷重、教師の結合荷重のことをそれぞれ単に生徒、教師と呼ぶことにする。

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

$$|x| = 1, |B_m| = \sqrt{N}, |J_k^0| = \sqrt{N}, \quad (7)$$

となる。生徒の大きさ \$|J_k|\$ 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_k\$ とし、これを生徒 \$J_k\$ の長さと呼ぶことにする。すなわち、\$|J_k| = l_k \sqrt{N}\$ である。\$l_k\$ は本論文で扱う巨視的変数のひとつである。

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、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、アダプトロン学習がよく知られている [9] ~ [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、学習は一般に以下のように表せる。

$$J_k^{m+1} = J_k^m + f(d^m, u_k^m) x^m. \quad (8)$$

ここで、\$m\$ は時刻ステップを表す。

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、\$K\$ 個の非線形単純パーセプトロンが単純多数決でアンサンブルとしての出力を決定するものとする。このとき、誤差 \$\epsilon\$ として、

$$\epsilon = \Theta \left(-d \sum_{k=1}^K \text{sgn}(J_k \cdot x) \right) \quad (9)$$

$$= \Theta \left(-\text{sgn}((v-a)v(v+a)) \sum_{k=1}^K \text{sgn}(J_k \cdot x) \right) \quad (10)$$

を用いることにする。ここで、\$\Theta(\cdot)\$ は以下のようなステップ関数である。

$$\Theta(z) = \begin{cases} +1, & z \geq 0, \\ 0, & z < 0. \end{cases} \quad (11)$$

汎化誤差 \$\epsilon_g\$ は式 (10) を入力 \$x\$ の確率分布 \$p(x)\$ で平均したものと定義する。すなわち、汎化誤差 \$\epsilon_g\$ は新たな入力 \$x\$ に対

するアンサンプルの出力が教師の出力と異なる確率と言うこともできる。誤差 ϵ は、教師の内部状態 v と生徒の規格化内部状態 u_k を用いて、 $\epsilon = \epsilon(v, \{u_k\})$ と書くことができるので、汎化誤差 ϵ_g も v, u_k の確率分布 $p(v, \{u_k\})$ を用いて、

$$\epsilon_g = \int dv \prod_{k=1}^K du_k p(v, \{u_k\}) \epsilon(v, \{u_k\}) \quad (12)$$

と書ける。 v と u_k は入力 x とそれとは無関係な結合荷重 B, J_k で書けるので $p(v, \{u_k\})$ は平均 0 の多重ガウス分布である。ここで、 v と u_k は平均 0 分散 1 のガウス分布にしたがうので、 $p(v, \{u_k\})$ の共分散行列 Σ の対角要素は 1 である。次に、この行列の非対角要素を求めるために、結合荷重間の方向余弦を議論する。まず、教師 B と生徒 J_k の方向余弦として R_k を定義する。すなわち、

$$R_k \equiv \frac{B \cdot J_k}{|B||J_k|} = \frac{1}{l_k N} \sum_{i=1}^N B_i J_{ki} \quad (13)$$

教師 B と生徒 J_k に相関がなければ $R_k = 0$ であり、両者の方向が同じであれば $R_k = 1$ であるから、以後は R_k のことを教師と生徒の類似度と呼ぶことにする。 R_k は本研究で扱う二番目の巨視的変数である。

また、生徒 J_k と生徒 $J_{k'}$ の方向余弦として $q_{kk'}$ を定義する。すなわち、

$$q_{kk'} \equiv \frac{J_k \cdot J_{k'}}{|J_k||J_{k'}|} = \frac{1}{l_k l_{k'} N} \sum_{i=1}^N J_{ki} J_{k'i} \quad (14)$$

ここで、 $k \neq k'$ である。

生徒 J_k と生徒 $J_{k'}$ に相関がなければ $q_{kk'} = 0$ であり、両者の方向が同じであれば $q_{kk'} = 1$ であるから、以後は $q_{kk'}$ のことを生徒間の類似度と呼ぶことにする。 $q_{kk'}$ は本研究で扱う三番目の巨視的変数である。

教師 B の内部状態 v と生徒 J_k の規格化内部状態 u_k の共分散は教師 B と生徒 J_k の類似度 R_k に等しい。また、生徒 J_k の規格化内部状態 u_k と生徒 $J_{k'}$ の規格化内部状態 $u_{k'}$ の共分散は生徒間の類似度 $q_{kk'}$ に等しい。よって、式 (12) の $p(v, \{u_k\})$ は R_k と $q_{kk'}$ を用いて以下のように書ける。

$$p(v, \{u_k\}) = \frac{1}{(2\pi)^{\frac{1+K}{2}} |\Sigma|^{\frac{1}{2}}} \times \exp\left(-\frac{(v, \{u_k\}) \Sigma^{-1} (v, \{u_k\})^T}{2}\right), \quad (15)$$

$$\Sigma = \begin{pmatrix} 1 & R_1 & R_2 & \dots & R_K \\ R_1 & 1 & q_{1,2} & \dots & q_{1,K} \\ R_2 & q_{2,1} & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & q_{K-1,K} \\ R_K & q_{K,1} & \dots & q_{K,K-1} & 1 \end{pmatrix} \quad (16)$$

3.2 巨視的変数の微分方程式

統計的学習理論の目的のひとつは汎化誤差 ϵ_g を理論的に求めることである。また、式 (12), (15), (16) より、汎化誤差 ϵ_g は R_k と $q_{kk'}$ がすべてわかれば計算できる。本論文では、入力、

教師、生徒の大きさを式 (7) のように設定しているので、 N が大きい極限では入力の影響は入力に関する平均 (サンプル平均) で置き換えることができる。この考え方を統計物理の分野では自己平均性と呼ぶ。一般の学習則の $l_k, R_k, q_{kk'}$ のダイナミクスは自己平均性に基づいて導出される以下のような決定論的連立微分方程式で記述することができる [7]~[9]。ここで t は時刻ステップ m を次元 N で正規化した時刻 $t = m/N$ である。

$$\frac{dl_k}{dt} = \langle f_k u_k \rangle + \frac{\langle f_k^2 \rangle}{2l_k}, \quad (17)$$

$$\frac{dR_k}{dt} = \frac{\langle f_k v \rangle - \langle f_k u_k \rangle R_k}{l_k} - \frac{R_k}{2l_k^2} \langle f_k^2 \rangle, \quad (18)$$

$$\begin{aligned} \frac{dq_{kk'}}{dt} &= \frac{\langle f_{k'} u_k \rangle - q_{kk'} \langle f_{k'} u_{k'} \rangle}{l_{k'}} \\ &+ \frac{\langle f_k u_{k'} \rangle - q_{kk'} \langle f_k u_k \rangle}{l_k} \\ &+ \frac{\langle f_k f_{k'} \rangle}{l_k l_{k'}} - \frac{q_{kk'}}{2} \left(\frac{\langle f_k^2 \rangle}{l_k^2} + \frac{\langle f_{k'}^2 \rangle}{l_{k'}^2} \right). \end{aligned} \quad (19)$$

ここで、 $\langle \cdot \rangle$ はサンプル平均を表す。すなわち、

$$\langle f_k u_k \rangle = \int dv du_k p_1 f(d, u_k) u_k, \quad (20)$$

$$\langle f_k v \rangle = \int dv du_k p_1 f(d, u_k) v, \quad (21)$$

$$\langle f_k^2 \rangle = \int dv du_k p_1 (f(d, u_k))^2, \quad (22)$$

$$\langle f_k u_{k'} \rangle = \int dv du_k du_{k'} p_2 f(d, u_k) u_{k'}, \quad (23)$$

$$\langle f_{k'} u_k \rangle = \int dv du_k du_{k'} p_2 f(d, u_{k'}) u_k, \quad (24)$$

$$\langle f_k f_{k'} \rangle = \int dv du_k du_{k'} p_2 f(d, u_k) f(d, u_{k'}) \quad (25)$$

である。ここで、 $d = \text{sgn}((v-a)v(v+a))$ であり、 $p_1 = p_1(v, u_k)$ 、 $p_2 = p_2(v, u_k, u_{k'})$ は式 (15), (16) においてそれぞれ $K=1, K=2$ とした多重ガウス関数である。

4. 結 果

本論文では生徒 J_k の初期値 J_k^0 、教師 B_m 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、

$$R_k^0 = 0, \quad q_{kk'}^0 = 0 \quad (26)$$

である。式 (26) と生徒の対称性より、式 (19) において、

$$\langle f_k u_{k'} \rangle = \langle f_{k'} u_k \rangle \quad (27)$$

が成り立つ。また、式 (26) と生徒の対称性より、式 (17)–(19) 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。

ヘブ学習、パーセプトロン学習はそれぞれ $f(d, u) = d$ 、 $f(d, u) = \Theta(-ud)$ で更新を行う学習則である。ヘブ学習とパーセプトロン学習のそれぞれについて式 (17)–(19) を

数値的に解くために必要なサンプル平均は式 (20)–(25) を実行することにより導出できる．これらのうち、 $\langle f_k u_k \rangle$ 、 $\langle f_k v \rangle$ 、 $\langle f_k^2 \rangle$ については Inoue らが [13] で導出している．それらに加えて今回我々は $\langle f_k u_{k'} \rangle$ 、 $\langle f_k f_{k'} \rangle$ を新たに導出した．結果を以下にまとめて示す．

ヘブ学習の場合

$$\langle f_k u_k \rangle = \frac{2R}{\sqrt{2\pi}} \left(2 \exp\left(-\frac{a^2}{2}\right) - 1 \right) \quad (28)$$

$$\langle f_k v \rangle = \frac{2}{\sqrt{2\pi}} \left(2 \exp\left(-\frac{a^2}{2}\right) - 1 \right) \quad (29)$$

$$\langle f_k^2 \rangle = 1 \quad (30)$$

$$\langle f_k u_{k'} \rangle = \langle f_{k'} u_{k'} \rangle = \langle f_k u_k \rangle \quad (31)$$

$$\langle f_k f_{k'} \rangle = \langle f_k^2 \rangle = 1 \quad (32)$$

パーセプトロン学習の場合

$$\langle f_k u_k \rangle = 2 \int_0^\infty Du u \left[H\left(\frac{a - Ru}{\sqrt{1 - R^2}}\right) - H\left(\frac{-Ru}{\sqrt{1 - R^2}}\right) - H\left(\frac{a + Ru}{\sqrt{1 - R^2}}\right) \right] \quad (33)$$

$$\langle f_k v \rangle = \frac{1}{\sqrt{2\pi}} \left(2 \exp\left(-\frac{a^2}{2}\right) - 1 - R \right) \quad (34)$$

$$\langle f_k^2 \rangle = 2 \left[\int_a^\infty Dv H\left(\frac{Rv}{\sqrt{1 - R^2}}\right) + \int_0^a Dv H\left(-\frac{Rv}{\sqrt{1 - R^2}}\right) \right] \quad (35)$$

$$\langle f_k u_{k'} \rangle = \frac{1}{1 - R^2} \left((q - R^2) \langle f_k u_k \rangle + R(1 - q) \langle f_k v \rangle \right) \quad (36)$$

$$\langle f_k f_{k'} \rangle = 2 \left(\int_{-\infty}^{-a} Dv \int_{-\frac{Rv}{\sqrt{1 - R^2}}}^\infty H(z) + \int_0^a Dv \int_{-\frac{Rv}{\sqrt{1 - R^2}}}^\infty H(z) \right) \quad (37)$$

$$z \equiv \frac{(q - R^2)u + R\sqrt{1 - R^2}v}{\sqrt{(1 - q)(1 + q - 2R^2)}} \quad (38)$$

ここで、

$$H(u) \equiv \int_u^\infty Dx, \quad Dx \equiv \frac{dx}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (39)$$

式 (28)–(39) を用いて連立微分方程式 (17)–(19) を数値的に解き R, q のダイナミクスを求めた．さらに、得られた R, q を使って式 (12) の数値積分を実行することにより汎化誤差 ϵ_g のダイナミクスを求めた．数値積分は式 (15) の逆行列計算を不要にするために積分変数を直交化 [7] したうえでメトロポリス法を用いて実行した．その際モンテカルロステップ数は 10^7 とした．ヘブ学習、パーセプトロン学習それぞれについて生徒数 $K = 3$ の汎化誤差 ϵ_g 、 $K = 1$ の ϵ_g 、 l 、 R 、 q を図 1–10 に示す．図中では $N = 10^3$ で実行した計算機シミュレーションの結果を重ねてプロットしている．計算機シミュレーションの ϵ_g は

各時点で 10^5 個のランダム入力によりテストを行うことにより求めている．

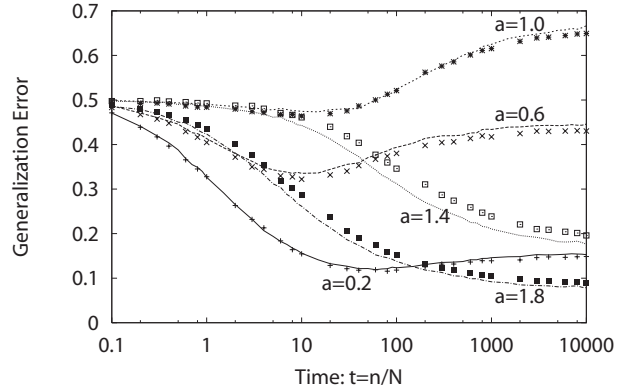


図 1 ヘブ学習の ϵ_g ($K = 3$)

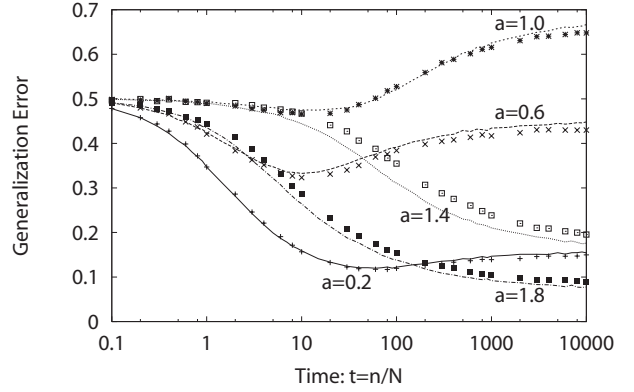


図 2 ヘブ学習の ϵ_g ($K = 1$)

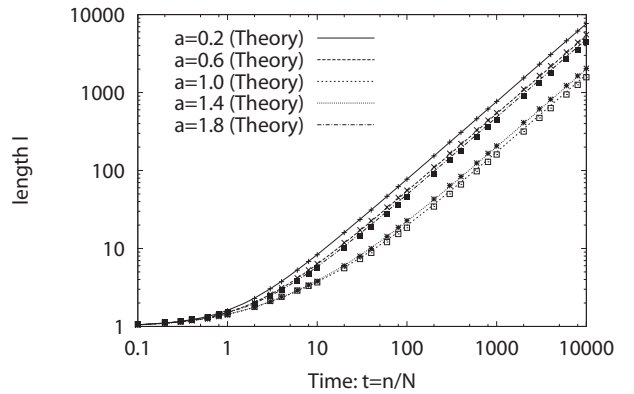


図 3 ヘブ学習の l

5. 議 論

5.1 ヘブ学習

図 1, 2 より、教師が非単調な場合にヘブ学習の汎化誤差 ϵ_g が定常値に達するためには 10^4 程度の時間がかかることがわかる．これは、後述するパーセプトロン学習 (図 6, 7) や教師がコミティマシンの場合 [14]–[16] のヘブ学習やパーセプトロン学習と比較してかなり遅い．また、図 1 と図 2 で、特にしきい値 $a = 0.2$ の場合をよく見ると、学習の初期段階で $K = 3$ の

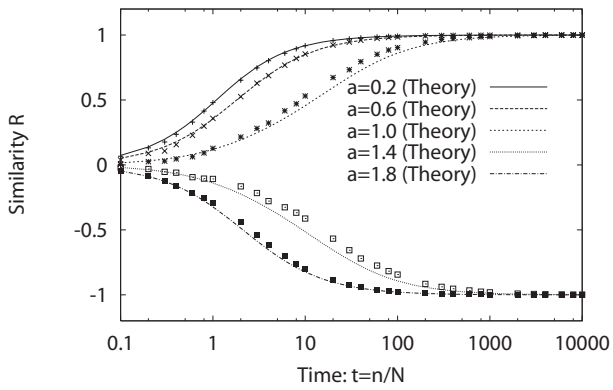


図 4 ヘブ学習の R

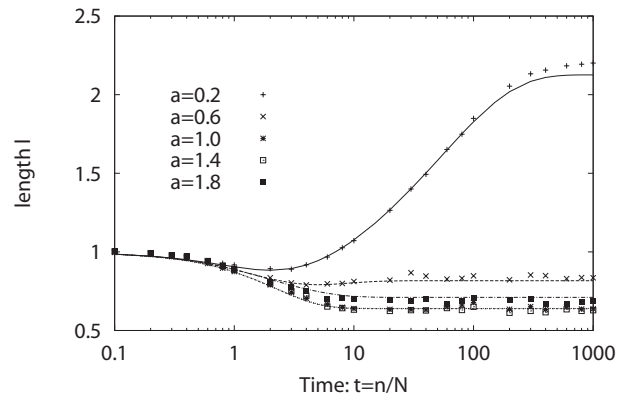


図 8 パーセプトロン学習の l

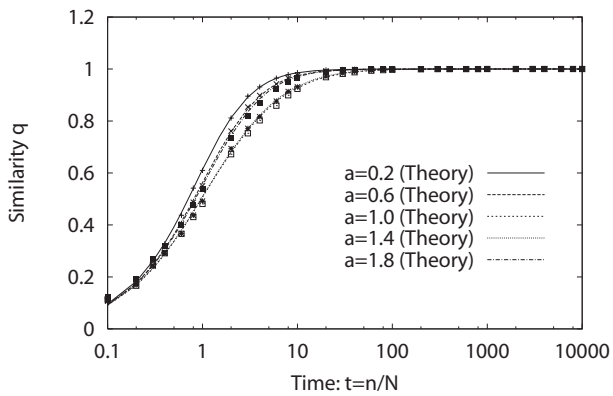


図 5 ヘブ学習の q

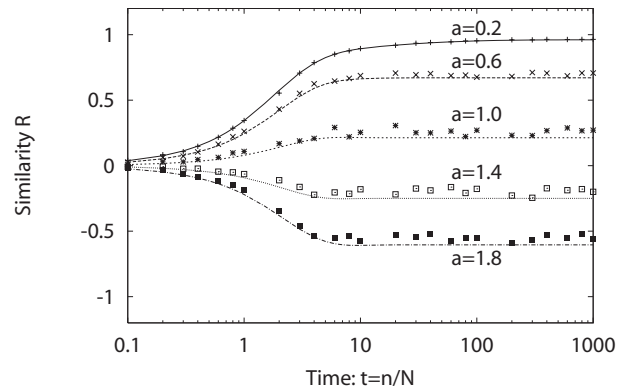


図 9 パーセプトロン学習の R

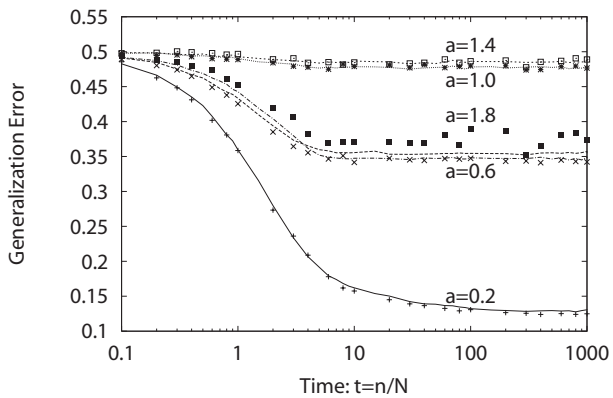


図 6 パーセプトロン学習の ϵ_g ($K = 3$)

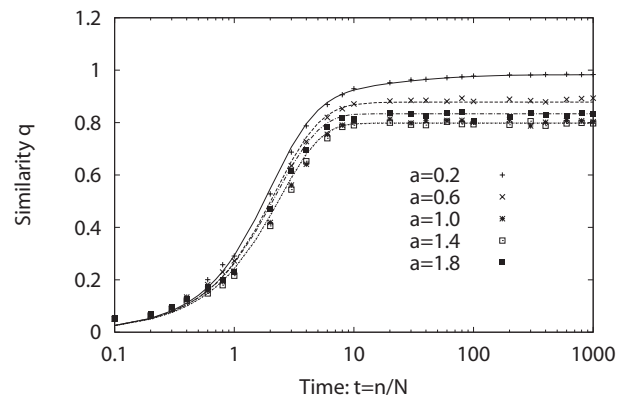


図 10 パーセプトロン学習の q

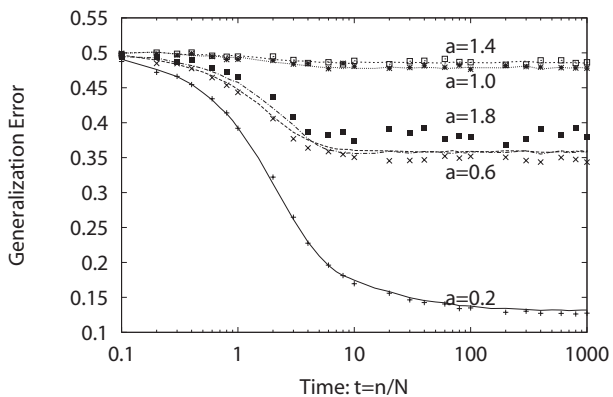


図 7 パーセプトロン学習の ϵ_g ($K = 1$)

汎化誤差が $K = 1$ の場合よりもわずかに小さいことから、学習の初期においてはわずかではあるがアンサンプルの効果があることがわかる。しかし、 $t > 10$ ではアンサンプルの効果は消滅している。図 5 よりしきい値 a の値によらず q は 1 に漸近することからヘブ学習の場合には生徒の多様性は消滅してしまうことがわかる。このためにアンサンプル学習の効果も消滅してしまう。さらに、図 1, 2 において、Inoue らが [13] で指摘しているように一種の過学習が見られる。すなわち、汎化誤差 ϵ_g がいったん最小値になってから若干増大して定常値に達する現象があることがわかる。特に $a = 1.0$ では汎化誤差 ϵ_g の定常値が初期値 0.5 よりも大きくなっている [12], [13]。つまり、ヘブ学習の場合、しきい値 a の値によっては、ランダムに決めた初期状態よりも生徒の汎化能力が悪くなってしまふ場合がある。

また, Inoue ら [12], [13] が解析している通り図 4 からヘブ学習の R は $+1$ または -1 に漸近することがわかる. なお, $+1$ と -1 のいずれに向かうかの分岐は $a_c = \sqrt{2 \ln 2} = 1.18$ で起こる [13].

5.2 パーセプトロン学習

図 6, 7 より, 教師が非単調な場合のパーセプトロン学習の汎化誤差 ϵ_g は $t = 1000$ でほぼ定常に達することがわかる. そこで, パーセプトロン学習の場合のアンサンブル学習の効果を明らかにするため, しいき値 a と $t = 1000$ における汎化誤差 ϵ_g の関係を調べた. 結果を図 11 に示す. また, アンサンブル学習の汎化誤差 ϵ_g を $K = 1$ の ϵ_g 理論値で割ることにより, アンサンブル学習の効果をよりわかりやすくした結果を図 12 に示す. これらの図から, パーセプトロン学習の場合はヘブ学習と異なり, アンサンブルの効果が消滅せずに残ることがわかる. 図 10 より, パーセプトロン学習の場合は q が漸近する値が 1 よりも小さいことがわかる. すなわち, パーセプトロン学習の場合はヘブ学習と異なり, 生徒の多様性が残る. このためにアンサンブル学習の効果が残る. 教師が単調なパーセプトロンの場合には生徒数 $K \rightarrow \infty$ で汎化誤差 ϵ_g は $K = 1$ のときの 0.72 倍に改善されることが知られている [7] が, 図 12 において a が大きい場合のアンサンブルの効果は生徒数 K が大きいときにこの値に近づくようすが見られる. しいき値 a が 0 または ∞ の場合には教師は単調なパーセプトロンと等価になることを考えると, この結果は自然である. また, 図 9 からパーセプトロン学習の R が漸近する値は $-1 \sim +1$ であり, その値はしいき値 a により決まる. Inoue ら [12], [13] が解析している通り, 式 (17), (18) の左辺を 0 と置くことによりこの値は $1 - 2 \exp(-a^2/2)$ と解析的に求められる.

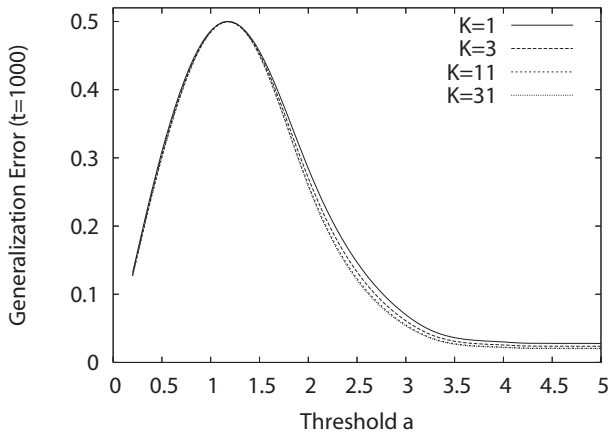


図 11 しいき値 a と ϵ_g ($t = 1000$, パーセプトロン学習)

6. むすび

教師が非単調なパーセプトロンであり, 生徒が単純パーセプトロンの集団であるようなアンサンブル学習についてオンライン学習の枠組みで議論した. その結果, ヘブ学習では生徒の多様性が消滅してしまうのでアンサンブルの効果が残らないこと, パーセプトロン学習では生徒の多様性が消滅せず, そのためにアンサンブルの効果が残ることがわかった.

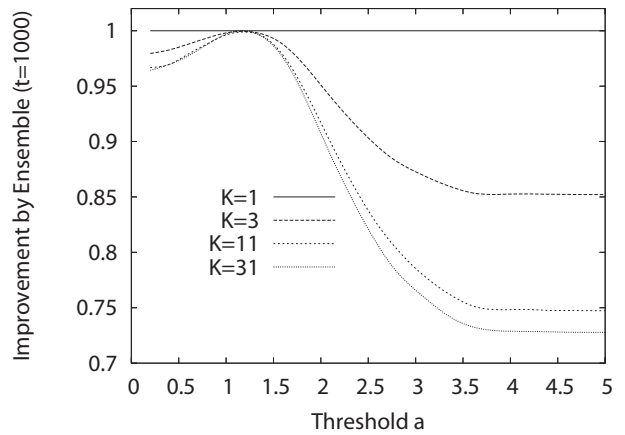


図 12 しいき値 a とアンサンブルの効果 ($t = 1000$, パーセプトロン学習) ϵ_g を $K = 1$ の ϵ_g 理論値で割った値をプロットしている.

謝 辞

本論文の一部は科学研究費補助金 (課題番号 14084212, 14580438, 15500151, 16500146) によるものであり, ここに感謝いたします.

文 献

- [1] Freund, Y. and Shapire, R.E., (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).
- [2] <http://www.boosting.org/>
- [3] 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.
- [4] Krogh, A. and Sollich, P., “Statistical mechanics of ensemble learning,” Phys. Rev. E, 55(1), 811-825 (1997).
- [5] Urbanczik, R., “Online learning with ensembles,” Phys. Rev. E, 62(1), 1448-1451 (2000).
- [6] 原, 岡田, “線形ウィークラーナーによるアンサンブル学習の汎化誤差の解析,” 情報論的学習理論ワークショップ予稿集, 113-118 (2002).
- [7] Miyoshi, S., Hara, K. and Okada, M., “Analysis of ensemble learning using simple perceptrons based on online learning theory,” Phys. Rev. E (in press), cond-mat/0403632
- [8] 三好誠司, 原一之, 岡田真人, “オンライン学習理論に基づく単純パーセプトロンのアンサンブル学習の解析,” 信学論 DII, J87-D-II(7), pp.1391-1401 (2004).
- [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] Anlauf, J.K. and Biehl, M., “The AdaTron: an adaptive perceptron algorithm,” Europhys. Lett., 10(7), 687-692 (1989).
- [11] Biehl, M. and Riegler, P., “On-line learning with a perceptron,” Europhys. Lett., 28(7), 525-530 (1994).
- [12] Inoue, J. and Nishimori, H., “On-line AdaTron learning of a unlearnable rules,” Phys. Rev. E, 55(4), 4544-4551 (1997).
- [13] Inoue, J., Nishimori, H. and Kabashima, Y., “A simple perceptron that learns non-monotonic rules,” cond-mat/9708096 (1997).
- [14] 三好誠司, 原一之, 岡田真人, “教師がコミティマシンの場合のアンサンブル学習,” 日本神経回路学会第 14 回全国大会 (JNNS2004) 講演論文集, pp.36-37, 2004.9
- [15] 三好誠司, 原一之, 岡田真人, “教師がコミティマシンの場合のアンサンブル学習,” 信学技報, NC2004-79, pp.63-68, 2004.10
- [16] 三好誠司, 原一之, 岡田真人, “教師がコミティマシンの場合のアンサンブル学習,” 情報論的学習理論ワークショップ (IBIS2004) 予稿集, pp.178-185, 2004.11