2004 年情報論的学習理論ワークショップ
2004 Workshop on Information-Based Induction Sciences (IBIS2004)
Tokyo, Japan, November 8 - 10, 2004.

教師がコミティマシンの場合のアンサンブル学習
Analysis of Ensemble Learning for Committee Machine Teacher

三好 誠司* 原 一之† 岡田 真人‡
Seiji Miyoshi Kazuyuki Hara Masato Okada

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 is not in the model space of one student is very attractive. In this paper ensemble learning, in which a teacher and students are a committee machine and simple perceptrons respectively, is discussed based on online learning theory and statistical mechanics. Hebbian learning gathers all students to the center of teacher units. Perceptron learning keeps a variety of students and the effect of ensemble does not disappear. AdaTron learning shows a kind of over-learning.

Keywords: ensemble learning, online learning, committee machine, 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, 2, 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, 5, 6, 7, 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，アダトロン学習の三つの学習則 [9, 10, 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ことにより，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，教師がコミティマシンであり，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．その結果，

*神戸市立工業高等専門学校 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
†東京都立工業高等専門学校 電子情報工学科，〒 140–0011 東京都品川区東大井 1–10–40,
Department of Electronics and Information Engineering, Tokyo Metropolitan College of Technology, 1–10–40 Higashi-oi, Shinagawa, Tokyo, 140–0011 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
科学技術振興機構 戦略的創造研究推進事業（さきがけ研究 21）「協調と制御」研究領域,
JST PRESTO

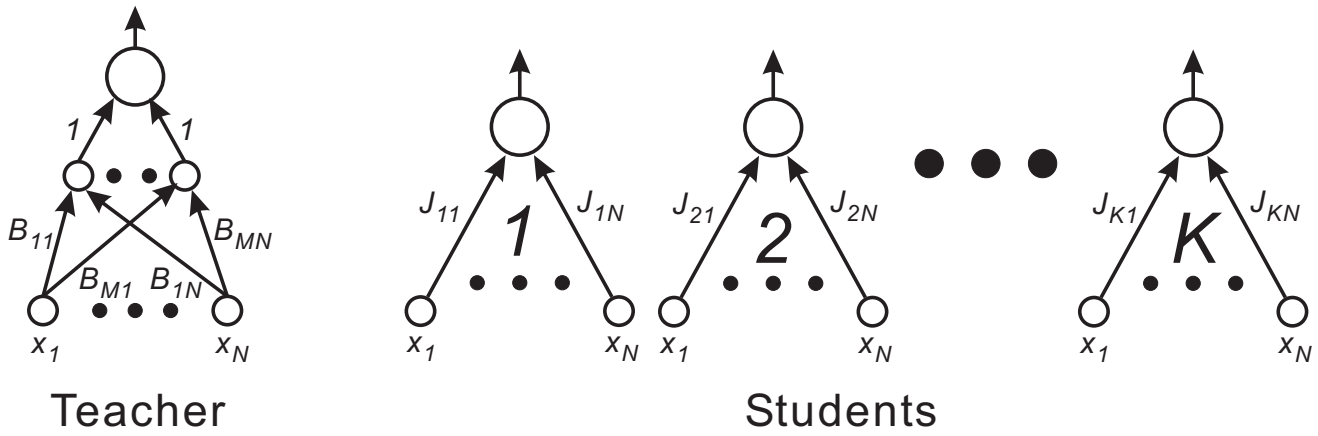

図 1: モデル

ヘブ学習では生徒が教師コミティマシン中間層ユニットの中央に集まること，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，アダトロン学習では一種の過学習が起こることなど，学習則毎の顕著な特徴が明らかになる．

2 モデル

図 1 に本論文で扱うモデルを示す．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 $\boldsymbol{J}_1, \boldsymbol{J}_2, \cdots, \boldsymbol{J}_K$ とする．$\boldsymbol{J}_k = (J_{k1}, \cdots, J_{kN}), k = 1, 2, \cdots, K$ と入力 $\boldsymbol{x} = (x_1, \cdots, x_N)$ は N 次元ベクトルであり，$\boldsymbol{x}$ 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，$\boldsymbol{J}_k$ の初期値 $\boldsymbol{J}_k^0$ 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，

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

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

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

$$u_k l_k = \boldsymbol{J}_k^T \boldsymbol{x}, \quad (3)$$

である．^{T} は転置を表す．l_k は生徒 $\boldsymbol{J}_k$ 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．

教師機械は中間層ユニット数 M のコミティマシンであるとする．各中間層ユニットは符号関数を出力関数とするパーセプトロンである．入力層から各中間層への結合荷重 $\boldsymbol{B}_m$ は N 次元ベクトルであり，その各要素 B_{mi} 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，

$$\langle B_{mi} \rangle = 0, \ \left\langle (B_{mi})^2 \right\rangle = 1. \quad (4)$$

教師出力ユニットは中間層ユニット出力の単純多数決をとる．すなわち教師の出力は

$$d = \mathrm{sgn}\left(\sum_{m=1}^{M} \mathrm{sgn}(v_m) \right), \quad (5)$$

$$v_m = \boldsymbol{B}_m^T \boldsymbol{x}. \quad (6)$$

v_m を教師中間層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，教師中間層の結合荷重のことをそれぞれ単に生徒，教師中間層と呼ぶことにする．また，教師出力ユニットのことを単に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，

$$|\boldsymbol{x}| = 1, \ \ |\boldsymbol{B}_m| = \sqrt{N}, \ \ |\boldsymbol{J}_k^0| = \sqrt{N}, \quad (7)$$

となる．生徒の大きさ $|\boldsymbol{J}_k|$ は一般には時間の経過とともに変化するが，$\sqrt{N}$ に対する比を l_k とし，これを生徒 $\boldsymbol{J}_k$ の長さと呼ぶことにする．すなわち，

$$|\boldsymbol{J}_k| = l_k \sqrt{N} \quad (8)$$

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 $\boldsymbol{x}$ が同じ順序で与えられる．個々の生徒は入力 $\boldsymbol{x}$ 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，アダトロン学習がよく知られている [9, 10, 11, 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 $\boldsymbol{x}$ 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．

$$\boldsymbol{J}_k^{n+1} = \boldsymbol{J}_k^n + f(d^n, u_k^n)\boldsymbol{x}^n. \tag{9}$$

ここで，n 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 ϵ_g を理論的に求めることである．本論文では，K 個の単純パーセプトロンが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．すなわち，K 個の生徒のうち $+1$ を出力している生徒の方が -1 を出力している生徒より多い場合にはアンサンブルの出力は $+1$ とし，逆の場合には -1 とする．このとき，誤差 ϵ として，

$$\epsilon = \Theta\left(-d\ \sum_{k=1}^{K} \mathrm{sgn}\left(\boldsymbol{J}_k^T\boldsymbol{x}\right)\right) \tag{10}$$

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．

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

この場合，教師の出力と生徒アンサンブルの出力が同じであれば $\epsilon = 0$ となり，そうでなければ $\epsilon = 1$ となる．汎化誤差 ϵ_g は式 (10) を入力 $\boldsymbol{x}$ の確率分布 $p(\boldsymbol{x})$ で平均したものと定義する．すなわち，汎化誤差 ϵ_g は新たな入力 $\boldsymbol{x}$ に対するアンサンブ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，教師中間層の内部状態 v と生徒の規格化内部状態 u_k を用いて，$\epsilon = \epsilon(\{v_m\}, \{u_k\})$ と書くことができるので，汎化誤差 ϵ_g も v_m，u_k の確率分布 $p(\{v_m\}, \{u_k\})$ を用いて，

$$\epsilon_g = \int \prod_{m=1}^{M} dv_m \prod_{k=1}^{K} du_k p(\{v_m\}, \{u_k\})\epsilon(\{v_m\}, \{u_k\}) \tag{12}$$

と書ける．v_m と u_k は入力 $\boldsymbol{x}$ とそれとは無関係な結合荷重 $\boldsymbol{B}_m$，$\boldsymbol{J}_k$ で書けるので $p(\{v_m\}, \{u_k\})$ は平均 0 の多重ガウス分布である．ここで，v_m と u_k は平均 0 分散 1 のガウス分布にしたがうので，$p(\{v_m\}, \{u_k\})$ の共分散行列 $\boldsymbol{\Sigma}$ 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，教師中間層 $\boldsymbol{B}_m$ と生徒 $\boldsymbol{J}_k$ の方向余弦として R_{mk} を定義する．すなわち，

$$R_{mk} \equiv \frac{\boldsymbol{B}_m^T\boldsymbol{J}_k}{|\boldsymbol{B}_m||\boldsymbol{J}_k|} = \frac{1}{l_k N}\sum_{i=1}^{N} B_{mi}J_{ki}. \tag{13}$$

教師中間層 $\boldsymbol{B}_m$ と生徒 $\boldsymbol{J}_k$ に相関がなければ $R_{mk} = 0$ であり，両者の方向が同じであれば $R_{mk} = 1$ であるから，以後は R_{mk} のことを教師中間層と生徒の類似度と呼ぶことにする．R_{mk} は本研究で扱う二番目の巨視的変数である．

また，生徒 $\boldsymbol{J}_k$ と生徒 $\boldsymbol{J}_{k'}$ の方向余弦として $q_{kk'}$ を定義する．すなわち，

$$q_{kk'} \equiv \frac{\boldsymbol{J}_k^T\boldsymbol{J}_{k'}}{|\boldsymbol{J}_k||\boldsymbol{J}_{k'}|} = \frac{1}{l_k l_{k'} N}\sum_{i=1}^{N} J_{ki}J_{k'i}, \tag{14}$$

ここで，$k \neq k'$ である．

生徒 $\boldsymbol{J}_k$ と生徒 $\boldsymbol{J}_{k'}$ 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．

教師中間層 $\boldsymbol{B}_m$ の内部状態 v_m と生徒 $\boldsymbol{J}_k$ の規格化内部状態 u_k の共分散は以下に示すように教師中間層 $\boldsymbol{B}_m$ と生徒 $\boldsymbol{J}_k$ の類似度 R_{mk} に等しい．

$$\begin{aligned} \langle v_m u_k \rangle &= \left\langle \frac{1}{l_k}\sum_{i=1}^{N} B_{mi}x_i \sum_{j=1}^{N} J_{kj}x_j \right\rangle && (15) \\ &= \frac{1}{l_k}\sum_{i=1}^{N} \langle B_{mi}J_{ki}\rangle \left\langle (x_i)^2 \right\rangle && (16) \\ &= R_{mk} && (17) \end{aligned}$$

また，生徒 $\boldsymbol{J}_k$ の規格化内部状態 u_k と生徒 $\boldsymbol{J}_{k'}$ の規格化内部状態 $u_{k'}$ の共分散は以下に示すように生徒間の類似度 $q_{kk'}$ に等しい．

$$\begin{aligned} \langle u_k u_{k'} \rangle &= \left\langle \frac{1}{l_k l_{k'}}\sum_{i=1}^{N} J_{ki}x_i \sum_{j=1}^{N} J_{k'j}x_j \right\rangle && (18) \\ &= \frac{1}{l_k l_{k'}}\sum_{i=1}^{N} \langle J_{ki}J_{k'i}\rangle \left\langle (x_i)^2 \right\rangle && (19) \\ &= q_{kk'} && (20) \end{aligned}$$

よって，式 (10)，(12) より汎化誤差 ϵ_g は R_{mk} と $q_{kk'}$ を用いて以下のように書ける．ここで $\boldsymbol{I}$ は $M \times M$ の単位行列である．

$$\begin{aligned} p(\{v_m\}, \{u_k\}) &= \frac{1}{(2\pi)^{\frac{M+K}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \\ &\times \exp\left(-\frac{(\{v_m\}, \{u_k\})\boldsymbol{\Sigma}^{-1}(\{v_m\}, \{u_k\})^T}{2}\right), \end{aligned} \tag{21}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{\Sigma}_B \\ \boldsymbol{\Sigma}_B^T & \boldsymbol{\Sigma}_D \end{pmatrix}, \tag{22}$$

$$\boldsymbol{\Sigma}_B = \begin{pmatrix} R_{1,1} & \dots & R_{1,K} \\ \vdots & \ddots & \vdots \\ R_{M,1} & \dots & R_{M,K} \end{pmatrix}, \tag{23}$$

$$\boldsymbol{\Sigma}_D = \begin{pmatrix} 1 & q_{1,2} & \cdots & q_{1,K} \\ q_{2,1} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & q_{K-1,K} \\ q_{K,1} & \dots & q_{K,K-1} & 1 \end{pmatrix}. \tag{24}$$

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)，(21) より，汎化誤差 ϵ_g は R_{mk} と $q_{kk'}$ がすべてわかれば計算できる．本論文では，入力，教師中間層，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．自己平均性に基づく一般の学習則の l_k，R_{mk}，$q_{kk'}$ に関する微分方程式は先行研究において以下のように導出されている [7, 8, 9]．ここで t は時刻ステップ n を次元 N で正規化した時刻 $t = n/N$ である．

$$\frac{dl_k}{dt} = \langle f_k u_k \rangle + \frac{\langle f_k^2 \rangle}{2l_k}, \tag{25}$$

$$\frac{dR_{mk}}{dt} = \frac{\langle f_k v_m \rangle - \langle f_k u_k \rangle R_{mk}}{l_k} - \frac{R_{mk}}{2l_k^2} \langle f_k^2 \rangle, \tag{26}$$

$$\frac{dq_{kk'}}{dt} = \frac{\langle f_{k'} u_k \rangle - q_{kk'} \langle f_{k'} u_{k'} \rangle}{l_{k'}} + \frac{\langle f_k u_{k'} \rangle - q_{kk'} \langle f_k u_k \rangle}{l_k} + \frac{\langle f_k f_{k'} \rangle}{l_k l_{k'}} - \frac{q_{kk'}}{2} \left(\frac{\langle f_k^2 \rangle}{l_k^2} + \frac{\langle f_{k'}^2 \rangle}{l_{k'}^2} \right). \tag{27}$$

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

$$\langle f_k u_k \rangle = \int \prod_{i=1}^{M} dv_i du_k p_1 f(d, u_k) u_k, \tag{28}$$

$$\langle f_k v_m \rangle = \int \prod_{i=1}^{M} dv_i du_k p_1 f(d, u_k) v_m, \tag{29}$$

$$\langle f_k^2 \rangle = \int \prod_{i=1}^{M} dv_i du_k p_1 \left(f(d, u_k) \right)^2, \tag{30}$$

$$\langle f_k u_{k'} \rangle = \int \prod_{i=1}^{M} dv_i du_k du_{k'} p_2 f(d, u_k) u_{k'}, \tag{31}$$

$$\langle f_{k'} u_k \rangle = \int \prod_{i=1}^{M} dv_i du_k du_{k'} p_2 f(d, u_{k'}) u_k, \tag{32}$$

$$\langle f_k f_{k'} \rangle = \int \prod_{i=1}^{M} dv_i du_k du_{k'} p_2 f(d, u_k) f(d, u_{k'}) \tag{33}$$

である．ここで，$d = \mathrm{sgn}\left(\sum_{j=1}^{M} \mathrm{sgn}(v_j) \right)$ である．また $p_1 = p_1(\{v_m\}, u_k)$，$p_2 = p_2(\{v_m\}, u_k, u_{k'})$ は式 (21)–(24) 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 $\boldsymbol{J}_k$ の初期値 $\boldsymbol{J}_k^0$，教師中間層 $\boldsymbol{B}_m$ 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，$N \to \infty$ 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，

$$R_{mk}^0 = 0, \quad q_{kk'}^0 = 0 \tag{34}$$

である．式 (34) と生徒の対称性より，式 (27) において，

$$\langle f_k u_{k'} \rangle = \langle f_{k'} u_k \rangle, \quad \langle f_k f_{k'} \rangle = \langle f_{k'} f_k \rangle \tag{35}$$

が成り立つ．また，式 (34) と生徒の対称性より，式 (25)–(27) の巨視的変数 $l_k, R_{mk}, q_{kk'}$ から添え字 m, 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，アダトロン学習はそれぞれ以下の式で更新を行う学習則である．

$$f(d, u) = d, \tag{36}$$

$$f(d, u) = \Theta(-ud)\, d, \tag{37}$$

$$f(d, u) = -u\Theta(-ud). \tag{38}$$

それぞれの学習則について式 (25)–(27) を数値的に解いて R,q のダイナミクスを求めた．その際，式 (25)–(27) 中の各サンプル平均は $p_1(\{v_m\}, u_k)$，$p_2(\{v_m\}, u_k, u_{k'})$ を求める際の $\boldsymbol{\Sigma}^{-1}$ の逆行列計算を不要にするために積分変数を直交化したうえで式 (28)–(33) の数値積分をメトロポリス法により実行することにより求めた．その際，モンテカルロステップ数は 10^6 とした．

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 (21) の逆行列計算を不要にするために積分変数を直交化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．結果を図 2，4，6 に示す．また，計算機シミュレーションの結果を図 3，5，7 に示す．計算機シミュレーションにおいては $N = 1000$ とし，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汎化誤差を計算した．

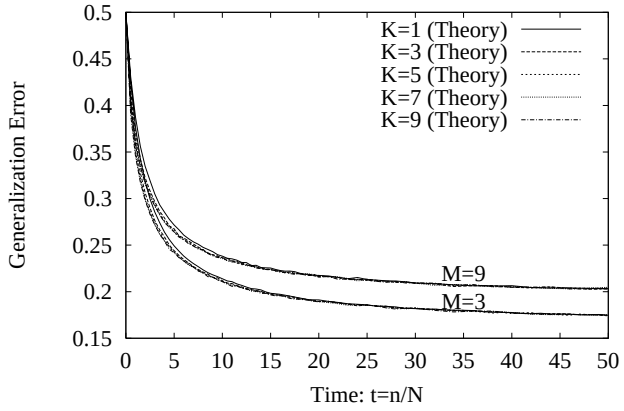

図 2: ヘブ学習の ϵ_g（理論）

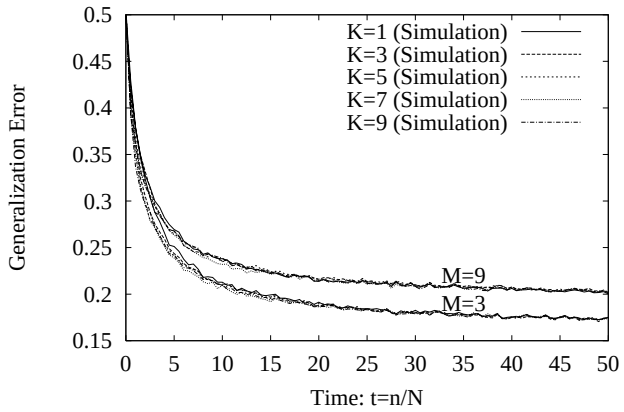

図 3: ヘブ学習の ϵ_g（計算機シミュレーション）

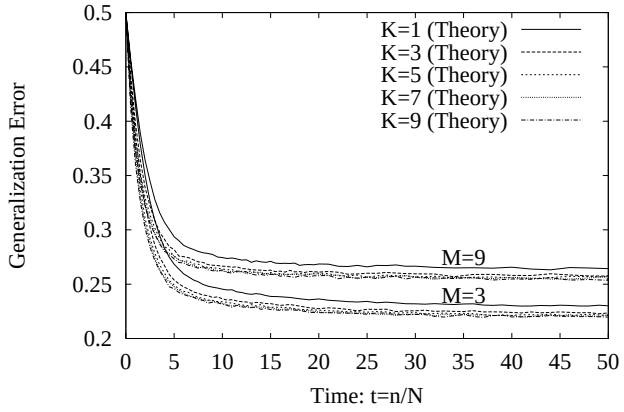

図 4: パーセプトロン学習の ϵ_g（理論）

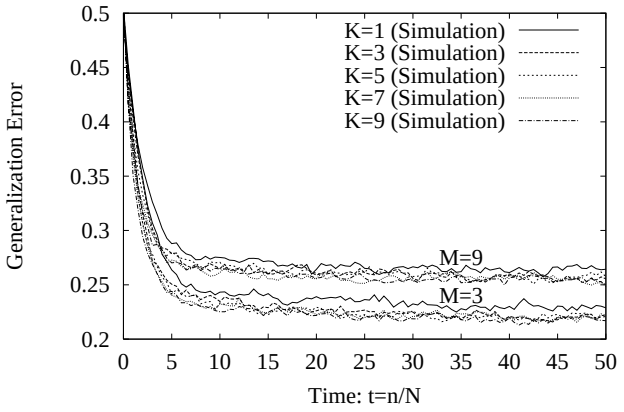

図 5: パーセプトロン学習の ϵ_g（計算機シミュレーション）

5 議論

図 2–7 より，いずれの学習則においてもダイナミクスのある期間においては K が大きいほど ϵ_g が小さいことがわかる．すなわち，教師機械がコミティマシンの場合でもアンサンブル学習の効果がある．

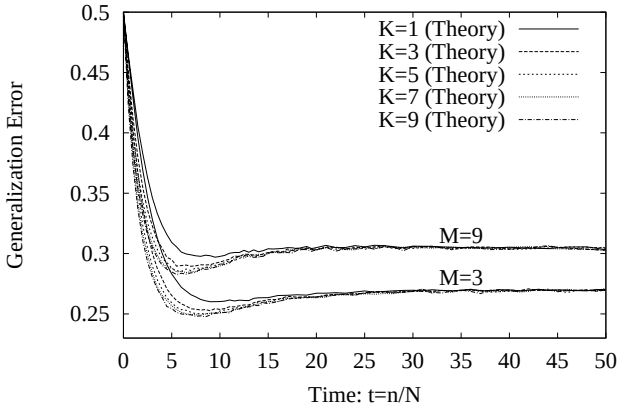

図 6: アダトロン学習の ϵ_g（理論）

図 8–10 は R と q のダイナミクスを示す．図 8 よりヘブ学習では q が急速に立ち上がり，$t \approx 20$ でほぼ 1 になっている．すなわち，ヘブ学習では $t \approx 20$ で生徒の多様性は消滅し，すべての生徒が同一になる．生徒が同一になってしまえば多数決をとる意味はないので，このことに対応して，図 2 においても $t \approx 20$ で生徒の数 K による汎化誤差の違いは消滅している．すなわち，ヘブ学習においてはアンサンブルの効果は学習の初期においてのみ存在し，その後は消滅する．また，図 8 よりヘブ学習における R の定常値は $M = 3$ のとき 0.577，$M = 9$ のとき 0.333 であるが，これは直交する M 個の教師中間層ユニットの中央（平均）にすべての生徒が漸近することを示している．図 11 に $M = 3$ の場合の定常状態

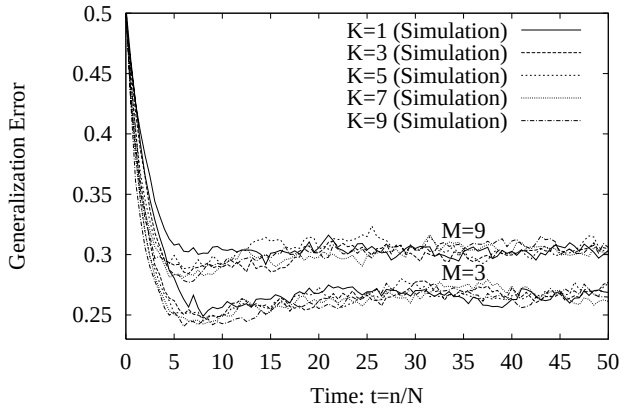

図 7: アダトロン学習の ϵ_g（計算機シミュレーション）

における教師中間層ユニット結合荷重 $\boldsymbol{B}_1$, $\boldsymbol{B}_2$, $\boldsymbol{B}_3$ と生徒 $\boldsymbol{J}_k$ の関係を示す.

パーセプトロン学習の場合にはヘブ学習とは異なるふるまいが見られる．すなわち，図 9 よりパーセプトロン学習の場合も $t \approx 20$ で R と q は定常状態に達するが，そのときの q の値は 1 よりも小さい．つまり，パーセプトロン学習においては生徒の多様性が消滅せずに残る．このことに対応して，図 4 において汎化誤差の定常値（残留汎化誤差）が生徒数 K に依存しており，その値は K が大きいほど小さい．つまり，パーセプトロン学習ではアンサンブルの効果が消滅せずに残る．なお，図 9 は $t = 50$ までの計算であるが，式 (25)–(27) の（左辺）$= 0$ とおいた定常解析を $M = 3$ のパーセプトロン学習について数値的に行ったところ，$l = 1.15, R = 0.50, q = 0.94$ が平衡点であることが確認できた.

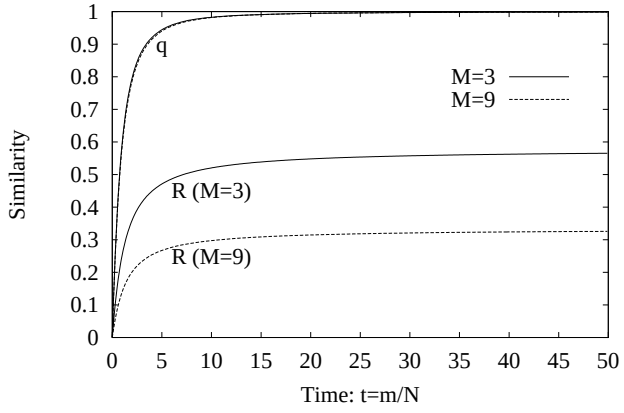

図 8: ヘブ学習の R と q（理論）

アダトロン学習の場合にはまた異なった特徴が見られる．すなわち，図 6 よりアダトロン学習では $t = 5 \sim 10$ で汎化誤差がいったん最小値をとり，その後少し増大して定常値に漸近することがわかる．このことは，アダト

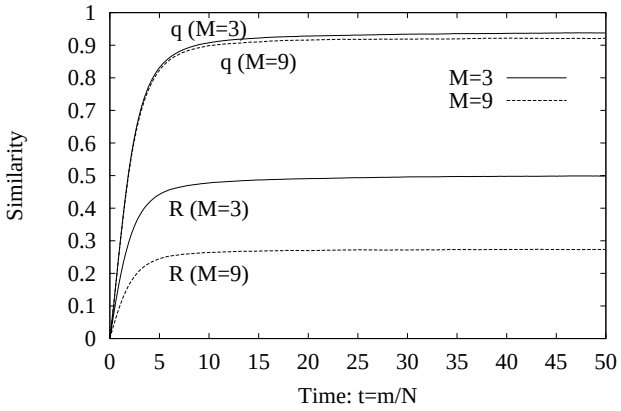

図 9: パーセプトロン学習の R と q（理論）

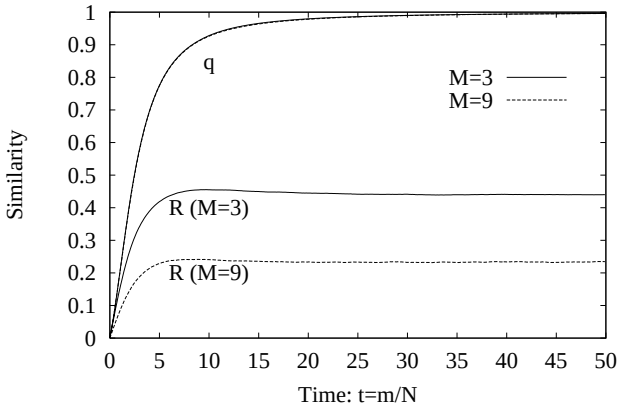

図 10: アダトロン学習の R と q（理論）

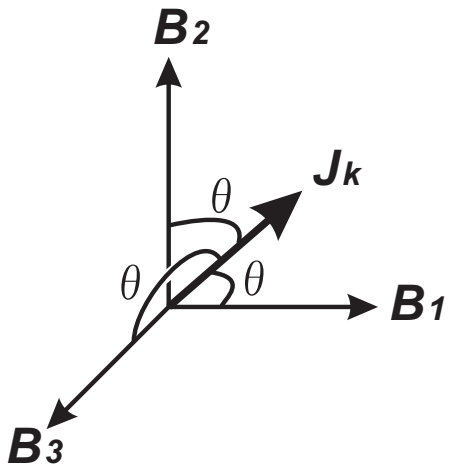

図 11: ヘブ学習の定常状態における教師と生徒．$M = 3$. $\cos\theta = 0.577$.

ロン学習の場合には学習パターン数に最適値が存在し，それ以上の数のパターンを用いると生徒集団の性能が下がってしまうことを意味している．つまり，アダトロン学習では一種の過学習が起こる．これはヘブ学習やパーセプトロン学習では見られない現象であり興味深い．なお，汎化誤差が最小値をとる時刻は K が大きいほど早い．本論文で扱っているモデルでは教師が生徒単体のモデル空間内にないが，汎化誤差が最小になるときに生徒アンサンブルの学習モデルが教師モデルにもっとも近くなっていると言える．また，図 10 をよく見るとアダトロン学習では R に関しても $M = 3$ のとき $t = 9.9$ で，$M = 9$ のとき $t = 8.6$ でいったん最大値をとり，その後少し減少して定常値に漸近することがわかる．このことに対応して図 6 においては生徒数 K が 1 の場合でも ϵ_g がいったん最小値をとる現象が見られる．すなわち，アダトロン学習の過学習は生徒が一個の場合でも起こっている．さらに，学習の初期においては生徒の多様性が維持されているためアンサンブルによって汎化誤差が大きく改善されるので生徒数が多いほど過学習のようすが顕著になる．

図 12 は各学習則の残留汎化誤差を示す．残留汎化誤差はヘブ学習で最も小さく，アダトロン学習で最も大きいことがわかる．

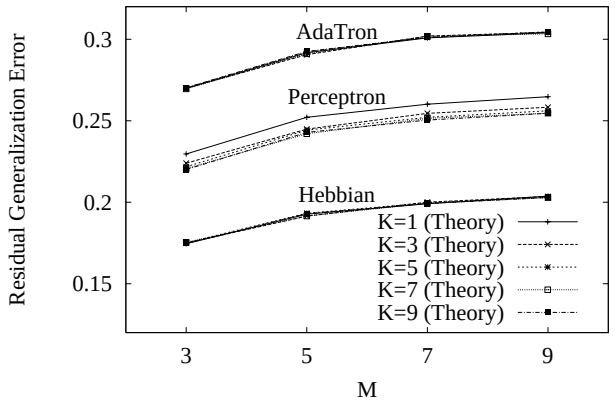

図 12: 残留汎化誤差（理論）

R と q が定常状態に達した後の生徒のふるまいを調べるために $t = 100$ における生徒とその後のその生徒の類似度を $N = 1000$ の計算機シミュレーションで調べた．結果を図 13–15 に示す．

図 13 より，ヘブ学習の場合には $t = 100$ の生徒とその後のその生徒の類似度が 1 であることがわかる．つまり，その生徒は停止している．このことと，図 8 において q の定常値が 1 であること，つまりすべての生徒が同一になること，および，R の定常値が直交する M 個の教師中間層の中央（平均）に対応する値であることを考

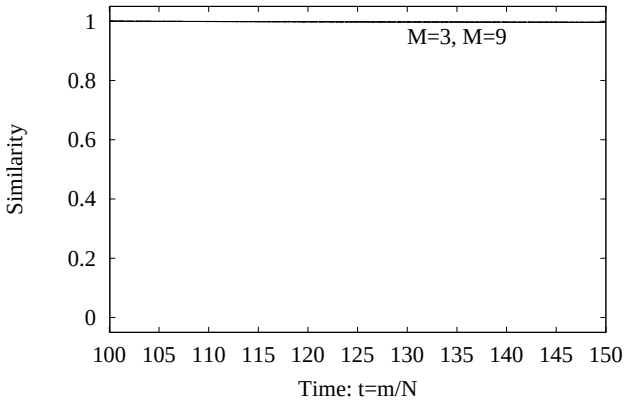

図 13: $t = 100$ における生徒とその後の生徒の類似度（ヘブ学習，計算機シミュレーション）

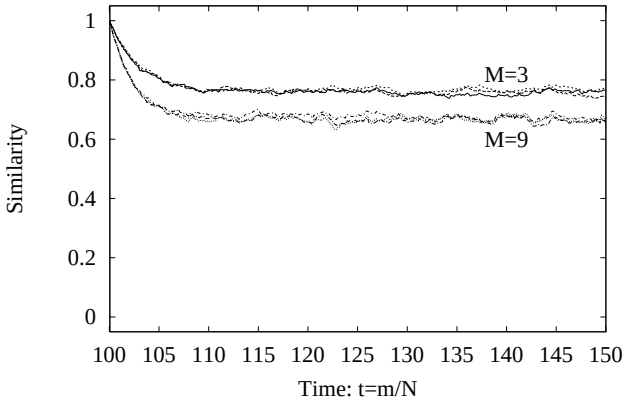

図 14: $t = 100$ における生徒とその後の生徒の類似度（パーセプトロン学習，計算機シミュレーション）

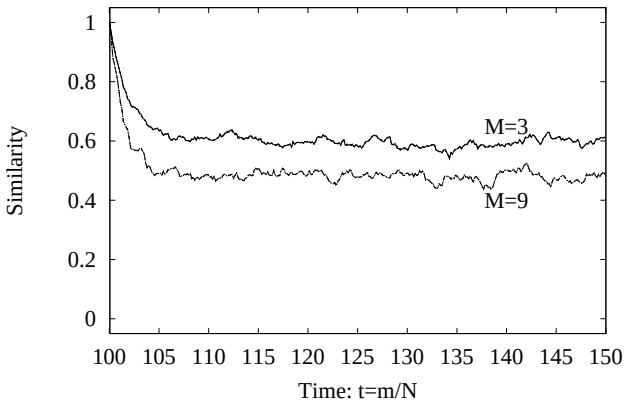

図 15: $t = 100$ における生徒とその後の生徒の類似度（アダトロン学習，計算機シミュレーション）

え合わせると，ヘブ学習ではすべての生徒は教師中間層の中央の一点に漸近し，そこで一体となって停止することがわかる．

これに対して，パーセプトロン学習の場合には図 14 より $t = 100$ の生徒とその後のその生徒の類似度が 1 より小さいことがわかる．つまり，R と q が定常値に達した後も生徒は動き続ける．このことと，図 9 において q の定常値が 1 より小さいこと，つまり生徒の多様性が残ることを考え合わせると，パーセプトロン学習の場合には K 個の生徒は一体にはならずに離れたままで動き続けることがわかる．なお，図 14 において $M = 3$ よりも $M = 9$ の類似度の方が小さいことから教師コミティマシンの中間層ユニット数 M が大きいほど生徒が動く範囲は広いと言える．

図 15 よりアダトロン学習の場合もパーセプトロン学習同様に $t = 100$ の生徒とその後のその生徒の類似度が 1 より小さいことがわかる．つまり，R と q が定常値に達した後も，生徒は動き続ける．このことと，図 10 において q の定常値が 1 であること，つまり生徒の多様性が消滅し生徒は一体となることを考え合わせると，アダトロン学習の場合には K 個の生徒は一体になって動き続けることがわかる．なお，図 15 よりアダトロン学習においてもパーセプトロン学習同様に教師コミティマシンの中間層ユニット数 M が大きいほど生徒が動く範囲は広い．また，図 15 を図 14 と比較するとアダトロン学習における $t = 100$ の生徒とその後のその生徒の類似度はパーセプトロン学習の場合のそれよりも小さいことがわかる．つまり，アダトロン学習の場合に生徒が動く範囲はパーセプトロン学習の場合よりも広い．

6 むすび

教師がコミティマシンであり，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，ヘブ学習ではすべての生徒は教師中間層の中央に漸近すること，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，アダトロン学習では一種の過学習が起こることなど，学習則毎の顕著な特徴が明らかになった．

謝辞

本論文の一部は科学研究費補助金（課題番号 13780313, 14084212, 15500151, 16500093）によるものであり，ここに感謝いたします．

参考文献

[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 一之，岡田 真人，“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，” IBIS 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”, 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).