社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

[チュートリアル講演] アンサンブル学習

岡田　真人[†,††]　　原　　一之[†††]　　三好　誠司[††††]

† 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
†† 科学技術振興事業団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
〒 351–0198 埼玉県和光市広沢 2–1
††† 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
†††† 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
E-mail: †okada@brain.riken.go.jp, ††hara@tokyo-tmct.ac.jp, †††miyoshi@kobe-kosen.ac.jp

あらまし　本講演では，バギングやパラレルブースティングなど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で議論する．教師と生徒は両方とも単純パーセプトロンである場合を議論する．アンサンブル学習機械の汎化誤差が生徒の結合荷重ベクトルの大きさ，教師と生徒の荷重ベクトルのオーバーラップ (方向余弦) および生徒の結合荷重ベクトル間のオーバーラップ (相関) にのみ依存することを示す．これらの巨視的な変数の学習のダイナミクスを記述する方程式を導出し，バギングとパラレルブースティングの性質を議論する．
キーワード　アンサンブル学習，バギング，パラレルブースティング，オンライン学習，統計力学，パーセプトロン

[Tutorial] Ensemble learning

Masato OKADA[†,††], Kazuyuki HARA[†††], and Seiji MIYOSHI[††††]

† RIKEN Brain Science Institute　2–1 Hirosawa, Wako-shi, Saitama, 351–0198 Japan
†† "Intelligent Cooperation and Control", PRESTO, JST
2–1 Hirosawa, Wako-shi, Saitama, 351–0198 Japan
††† Tokyo Metropolitan College of Technology　1–10–40 Higashi-oi, Shinagawa, Tokyo, 140–0011 Japan
†††† Kobe City College of Technology　8–3 Gakuenhigashimachi, Nishi-ku, Kobe-shi, 651–2194 Japan
E-mail: †okada@brain.riken.go.jp, ††hara@tokyo-tmct.ac.jp, †††miyoshi@kobe-kosen.ac.jp

Abstract　We discuss emsemble learning algorithms, i.e., bagging and parallel boosting based on the on-line learning from statisitcal mechnical point of view. We treat cases that both of teacher and students are simple perceptrons. We show the the generalization error of emsemble learnig machine depends only on the normls of the weight vectors of student perceptrons, overlaps (direction cosine) between the weight vector of the teacher perceptron and those of the student perceptrons, and correlations between the weight vectors of the student percetrons. We derive differential equations to describe the dynamics of these macrosocpic variables. Applying these equations to a linear perceptron case and a nonlinear perceptron case, we discuss the bagging and parallel boosting.
Key words　Ensemble learning, Bagging, Parallel boosting, On-line learning, Statistical mechanics, Perceptron

1. はじめに

近年，ブースティングの一種であるバギング [1] やアダブースト [2] などのように，多数の性能の劣る学習機械 (ウィークラーナー) を用いることによって，それらを個々に用いた場合に比べて性能を改善しようとする研究が多く行われ，単一の学習機械を用いた場合より優れた結果が得られている．これらはアンサンブル学習とも呼ばれ，注目されている．ここでは，アンサンブル学習の一種であるバギングとパラレルブースティング [3] の統計力学的なオンライン学習 [5] の枠組を解説する [4], [6]〜[8].

2. モデル

2.1 学習機械

図 1 に議論する教師と生徒を示す．教師は一つの単純パーセプトンであり，生徒は K 個の単純パーセプトンであるとする．これらの教師パーセプトロンと生徒パーセプトロンは図に示すように N 次元の入力 $\boldsymbol{x} = (x_1, \ldots x_N)$ を受けとるとする．こ

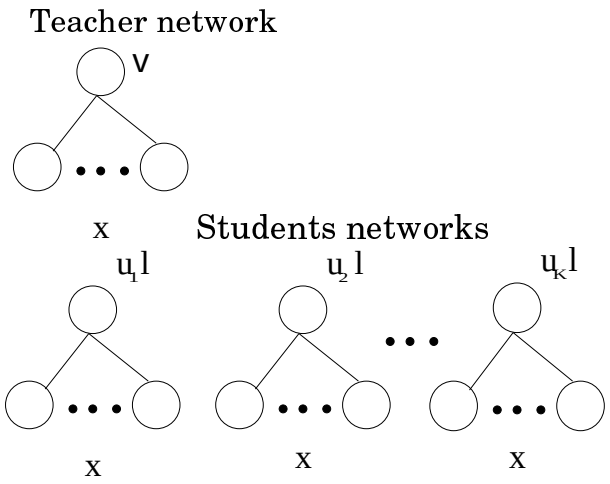


図 1　生徒と教師のネットワーク構造

こでは入力次元 $N \to \infty$ の熱力学的極限を考える．入力の各成分は平均 0 で分散が $1/N$ の確率分布から独立に生成されるものとする．この場合，入力ベクトルの大きさは 1 になる，

$$< x_i >= 0, \quad < (x_i)^2 >= \frac{1}{N}, \quad |\boldsymbol{x}| = 1, \tag{1}$$

教師パーセプトロンの結合荷重を $\boldsymbol{B}$ とし，$\boldsymbol{B}$ の各成分 B_i を平均 0 分散 1 の確率分布から生成する．この場合，教師パーセプトロンの結合荷重をの大きさは N になる，

$$< B_i >= 0, \quad < (B_i)^2 >= 1, \quad |\boldsymbol{B}| = \sqrt{N}. \tag{2}$$

教師パーセプトロンの出力 z は出力関数 $F(\cdot)$ と入力和 v を用いて，

$$z = F(v), \quad v = \sum_{i=1}^{N} B_i x_i = \boldsymbol{B} \cdot \boldsymbol{x}, \tag{3}$$

で与えられるとする．ここでは，$F(\cdot)$ が線形の場合と符号関数で与えられる場合を議論する．式 (1) と (2) より教師の入力和 v は平均 0 分散 1 のガウス分布に従う．k 番目の生徒パーセプトロンの結合荷重を $\boldsymbol{J}^k$ とする．結合荷重の学習則は後で述べるが，$\boldsymbol{J}^k$ の各成分は入力次元 N に対して $O(1)$ であるとし，結合荷重の大きさを，

$$|\boldsymbol{J}|^{\boldsymbol{k}} = l_k \sqrt{N}, \tag{4}$$

とする．ここで l_k は $O(1)$ となる．k 番目の生徒パーセプトロンへの入力和を結合荷重の大きさ l_k を用いて $l_k u_k$ あらわし，出力 y_k を

$$y_k = F(l_k u_k), \quad l_k u_k = \sum_{i=1}^{N} J_i^k x_i = \boldsymbol{J}^k \cdot \boldsymbol{x}, \tag{5}$$

とする．生徒パーセプトロンの出力関数は教師パーセプトロンの出力関数と同じ $F(\cdot)$ を用いる．式 (1) と結合荷重 $\boldsymbol{J}^k$ の性質より，k 番目の生徒の入力和をあらわす u_k は平均 0 分散 1 のガウス分布に従う．

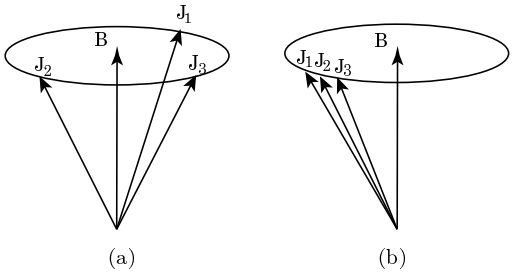


図 2　アンサンブル学習の効果 ($K = 3$)．(a) は生徒同士が似ていない (相関 q が小さい) 場合をあらわし，(b) は生徒同士が似ている (相関 q が大きい) 場合をあらわす．

2.2　学習アルゴリズム

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．ここでは，この手続きを学習と呼ぼう．線形パーセプトロンの学習則である勾配法や符号関数を出力関数する (非線形) パーセプトロンの代表的な学習則である，ヘブ学習，パーセプトロン学習，アダトロン学習などのほとどの学習則は，以下の式であらわすことができる，

$$\boldsymbol{J}^k(m+1) = \boldsymbol{J}^k(m) + f_k(m)\boldsymbol{x}(m) \tag{6}$$

$$f_k(m) = f(v(m), u_k(m), l_k(m)) \tag{7}$$

ここで，m は時刻ステップをあらわし，それぞれの変数についた (m) は，学習ステップ m でのそれぞれの変数がとる値を意味する．学習方法はオンライン学習を用いる．オンライン学習では入力 $\boldsymbol{x}$ が与えられると直ちに式 (7) を用いて結合荷重 $\boldsymbol{J}^k$ を更新する．そしてこのとき用いた入力 $\boldsymbol{x}$ は，以降の学習では使わない．このような定式化を行うと，入力 $\boldsymbol{x}$ と生徒の結合荷重 $\boldsymbol{J}^k$ は統計的に独立になるため，解析が容易になる．f の具体的な形は以下の式で与えられる，

$$f(v,u,l) = (v - lu), \qquad \text{勾配法} \tag{8}$$

$$f(v,u,l) = \mathrm{sgn}(v), \qquad \text{ヘブ学習} \tag{9}$$

$$f(v,u,l) = \Theta(-uv)\mathrm{sgn}(v), \qquad \text{パーセプトロン学習} \tag{10}$$

$$f(v,u,l) = -u\Theta(-uv), \qquad \text{アダトロン学習} \tag{11}$$

式 (7) の学習則を用いて，各生徒の結合荷重 $\boldsymbol{J}^k$ を学習した時に，$\boldsymbol{J}^k$ がどのように変化するかを考えよう．例えば，生徒の結合荷重の学習が無相関な状態から出発したとしよう，ある程度学習が進むと，K 個の生徒の結合荷重 $\boldsymbol{J}^k$ は図 2(a) のように教師の結合荷重 $\boldsymbol{B}$ の近くに当距離に分布しているはずである．この場合，生徒の結合荷重の平均，

$$\bar{\boldsymbol{J}} = \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{J}^k, \tag{12}$$

は先生の結合荷重 $\boldsymbol{B}$ を良く近似している．つまり，何らかの形で生徒の出力を平均すれば，個々の生徒よりも先生を良く近似できるはずである．これがアンサンブル学習の直観的な理解である．次の節では，この定性的な理解をもとに，アンサンブル学習を定式化する．

2.3 アンサンブル学習

アンサンブル学習機械の出力 y を，

$$y = F\left(\sum c_k y_k\right) = F\left(\sum c_k F(\boldsymbol{J}^k \cdot \boldsymbol{x})\right), \tag{13}$$

とする．ただし c_k は k 番目の生徒パーセプトロンの荷重平均の重みで，

$$\sum_{k=1}^{K} c_k = 1, \tag{14}$$

を満たすものとする．通常の平均を用いる場合，すなわち $c_k = 1/K$ の場合をバギングと呼び，荷重平均を用いることをパラレルブースティングと呼ぶ．

ここでもっとも簡単である出力関数が線形の場合のバギングを用いて，アンサンブル学習の効果を定性的に説明する．式 (13) に $F(x) = x$ と $C_k = 1/K$ を代入すると，アンサンブル学習機械の出力 y は式 (12) の平均結合荷重 $\bar{\boldsymbol{J}}$ を用いて書くことができる，

$$y = \bar{\boldsymbol{J}} \cdot \boldsymbol{x}. \tag{15}$$

図 2 に示すように，この平均結合荷重 $\bar{\boldsymbol{J}}$ は教師の結合荷重 $\boldsymbol{B}$ に近いことが予想されるので，アンサンブル学習機械は個別の生徒より，教師を良く近似できることがわかる．

3. 理　　論

3.1 汎化誤差 – 何を求めれば良いか –

統計的学習理論の目的の一つは，汎化誤差を理論的に求めることである．ここでは，線形パーセプトロンに関しては，誤差 ϵ として自乗誤差を用い，

$$\epsilon = \frac{1}{2}(z - y)^2, \tag{16}$$

非線形パーセプトロンに関しては，

$$\epsilon = \Theta(-zy), \tag{17}$$

を用いる．ここで，式 (17) の誤差は，先生の出力とアンサンブル学習機械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 (16) または (17) の誤差 ϵ を入力 $\boldsymbol{x}$ の確率分布 $p(\boldsymbol{x})$ で平均したもので定義する．誤差 ϵ は，教師と生徒の入力和 v と $l_k u_k$ および平均の荷重 c_k を用いて，$\epsilon = \epsilon(v, \{u_k\}, \{l_k\}, \{c_k\})$ と書くことができるので，汎化誤差も入力和の確率分布 $p(v, \{u_k\})$ を用いて，

$$\begin{aligned}\epsilon_g &\equiv \int d\boldsymbol{x} p(\boldsymbol{x}) \epsilon \\ &= \int dv \prod_k du_k p(v, \{u_k\}) \epsilon(v, \{u_k\}, \{l_k\}, \{c_k\}).\end{aligned} \tag{18}$$

となる．入力和は入力 $\boldsymbol{x}$ と入力 $\boldsymbol{x}$ とは無相関な変数 $\boldsymbol{B}$ と $\boldsymbol{J}^k$ で書けるので，$p(v, \{u_k\})$ は平均 0 の多重ガウス分布に従う．前節で述べたように，v と u_k は平均 0 分散 1 のガウス分布に従うので，$p(v, \{u_k\})$ の共分散行列の対角要素は 1 である．共分散行列の非対角要素を求める前に，その前に準備として結合荷重間の方向余弦を議論する．先生と生徒の結合荷重のオーバラップを R_k とし，

$$R_k \equiv \frac{1}{|\boldsymbol{B}||\boldsymbol{J}^k|} \sum_{i=1}^{N} B_i J_i^k = \frac{1}{l_k N} \sum_{i=1}^{N} B_i J_i^k, \tag{19}$$

k 番目と k' 番めの生徒間の結合荷重のオーバラップを qkk' とする，

$$q_{kk'} \equiv \frac{1}{|\boldsymbol{J}^k||\boldsymbol{J}^{k'}|} \sum_{i=1}^{N} J_i^k J_i^{k'} = \frac{1}{l_k l_{k'} N} \sum_{i=1}^{N} J_i^k J_i^{k'}. \tag{20}$$

ここでは R と区別するために，q を相関と呼ぶ，先生の入力和 v と生徒の入力和 u_k の分散は下記に示すように，先生と生徒のオーバラップになる，

$$\begin{aligned}< v u_k > &= \frac{1}{l_k} \sum_i \sum_j B_i x_i J_j^k x_j \\ &= \frac{1}{l_k} \sum_j < B_i J_i^k >< (x_i)^2 > \\ &= \frac{1}{N l_k} \sum_j < B_i J_i^k >= R_k\end{aligned} \tag{21}$$

まず，理論解析のための仮定について説明する．同様に u_k と $u_{k'}$ の相関を求めると，

$$\begin{aligned}< u_k u_{k'} > &= \frac{1}{l_k l_{k'}} \sum_i \sum_j J_i^k x_i J_j^{k'} x_j \\ &= \frac{1}{l_k l_{k'}} \sum_j < J_i^k J_i^{k'} >< (x_i)^2 > \\ &= \frac{1}{N l_k} \sum_j < J_i^k J_i^{k'} >= q_{kk'},\end{aligned} \tag{22}$$

となる．よって式 (18) の汎化誤差 ϵ は，生徒の荷重ベクトルの大きさ l_k と教師と生徒のオーバーラップ R_k と生徒間の相関 $q_{kk'}$ および平均の荷重 c_k を用いて，$\epsilon_g = \epsilon_g(\{l_k\}, \{R_k\}, \{q_{kk'}\}, \{c_k\})$ と書くことができる．アンサンブル学習の平均の荷重 c_k は設計者が事前に決めるので，この系の汎化誤差は l_k，R_k，$q_{kk'}$ の三種類の巨視的変数 (オーダーパラメータ) が学習の進行に従ってどのように変化するかを決めればよい．

3.2 オーダーパラメータのダイナミクス

ここではオーダーパラメータ l_k，R_k，$q_{kk'}$ が学習によりどのように変化するかを議論する．一つの生徒に関するオーダーパラメータである l_k と R_k は

$$\frac{dl_k}{dt} = \langle f_k u_k \rangle + \frac{\langle f_k^2 \rangle}{2l}, \tag{23}$$

$$\frac{dR_k}{dt} = \frac{\langle f_k v_k \rangle - \langle f_k u_k \rangle R_k}{l_k} - \frac{R_k}{2l_k^2} \langle f_k^2 \rangle. \tag{24}$$

となる．これらの式の導出に関して次の文献を参考にせよ [5], [9], [10]．$q_{kk'}$ の導出を説明する．$\boldsymbol{J}_k$ と $\boldsymbol{J}_{k'}$ に関する式 (7) の両辺をかけあわせると，

$$\begin{aligned}&N l_k(m+1) l_{k'}(m+1) q_{kk'}(m+1) \\ &= N l_k(m) l_{k'}(m) q_{kk'}(m) + f_{k'}(m) l_k(m) u_k(m) \\ &+ f_k(m) l_{k'}(m) u_{k'}(m) + f_k(m) f_{k'}(m),\end{aligned} \tag{25}$$

を得る．ここで式 (25) の N 依存性に注目されたい．N は十分大きいため一個の学習サンプルでは $l_k(m+1)$ や $q_{kk'}(m+1)$ は，$l_k(m)$ や $q_{kk'}(m)$ に比べてほとんど変化しない．$l_k(m)$ や $q_{kk'}(m)$ が $O(1)$ 変化するためには，式 (7) であらわされるオンライン学習が $O(N)$ 回ほどこされる必要がある．そこでオンライン学習のステップ m を $m = Nt$ とおき，学習過程を連続変数である時刻 t であらわす．時刻が t から $t + dt$ になった場合を考えよう．ここで dt は微小量であるが N に対しては $O(1)$ である．微小時間 dt の間に l_k は dl_k だけ変化し，$q_{kk'}$ は $dq_{kk'}$ だけ変化したとすると，式 (25) は，

$$\begin{aligned} &N(l_k + dl_k)(l_{k'} + dl_{k'})(q_{kk'} + dq_{kk'}) \\ &= Nl_k l_{k'} q_{kk'} + l_k N dt\langle f_{k'} u_k\rangle + l_{k'} N dt\langle f_k u_{k'}\rangle \\ &+ N dt\langle f_k f_{k'}\rangle, \end{aligned} \tag{26}$$

となり，$q_{kk'}$ の微分方程式は，

$$\frac{dq_{kk'}}{dt} = \frac{\langle f_{k'} u_k\rangle}{l_{k'}} + \frac{\langle f_k u_{k'}\rangle}{l_k} + \frac{\langle f_k f_{k'}\rangle}{l_k l_{k'}} - \frac{q_{kk'}}{l_k}\frac{dl_k}{dt} - \frac{q_{kk'}}{l_{k'}}\frac{dl_{k'}}{dt}, \tag{27}$$

となる．

4. 線形パーセプトロン

4.1 汎化誤差とオーダーパラメータダイナミクス

出力関数が $F(x) = x$ である線形パーセプトロンを議論しよう．この場合は，関数の線形性のため式 (18) の汎化誤差 ϵ_g を解析的に求めることができる，

$$\begin{aligned} &\epsilon_g(\{l_k\}, \{R_k\}, \{q_{kk'}\}, \{c_k\}) \\ &= \frac{1}{2}(1 - 2\sum_{k=1}^{K} c_k R_k l_k + \sum_{k=1}^{K}\sum_{k'=1}^{K} c_k c_{k'} q_{kk'} l_k l_{k'}). \end{aligned} \tag{28}$$

オーダーパラメータの微分方程式は，

$$\frac{dl_k}{dt} = \frac{1 - l_k^2}{2l_k}, \tag{29}$$

$$\frac{dR_k}{dt} = \frac{1}{l_k} - \frac{R_k}{2}\left(1 + \frac{1}{l_k^2}\right), \tag{30}$$

$$\frac{dq_{kk'}}{dt} = \frac{1}{l_k l'_k} - \frac{q_{kk'}}{2l_k^2} - \frac{q_{kk'}}{2l_{k'}^2}, \tag{31}$$

となる．

4.2 生徒の結合荷重が統計的に一様である場合

もっとも単純である，生徒の結合荷重が統計的に一様である場合を議論する．この場合一様性の仮定より，$R_k = R$, $l_k = l$, $q_{kk'} = q$ とする．また荷重平均の重みも $c_k = 1/K$ とする．後の議論からもわかるように，生徒の結合荷重が統計的に一様である場合，荷重平均の重みが一様であることは汎化誤差を小さくする点最適である．この時，式 (18) および式 (28) の汎化誤差は $R,\ l,\ q,\ K$ の関数となる，

$$\begin{aligned} &\epsilon_g(l, R, q, K) \\ &= \frac{1}{2}\left(\frac{l^2(1-q)}{K} + (q - R^2)l^2 + (Rl - 1)^2\right). \end{aligned} \tag{32}$$

式 (32) の右辺の第一項は生徒の個数 K に依存し，$K \to \infty$ の極限では 0 に収束するので，K が十分大きい時は第一項を無視できる．式 (32) の右辺の第三項は，個々の生徒の特性をあらわす l と R だけで書かれているので，第三項はアンサンブル学習の効果をあらわさない．第二項は教師と生徒のオーバーラップ R と生徒同士の相関 q からなる．q が R^2 になるべく近い方が汎化誤差が小さくなることがわかる．つまりアンサンブル学習を成功させるためには，教師とのオーバーラップを大きくするとともに，生徒同士の相関を小さくする必要があることがわかる．

より具体的に，ここでは教師の結合荷重 $\boldsymbol{B}$ と同様に，生徒パーセプトロンの結合荷重 $\boldsymbol{J}^k$ の各成分 J_i^k を平均 0 分散 1 の確率分布から生成した場合を議論する．この場合，各生徒パーセプトロンの結合荷重をの大きさは N になり，l_k の初期値は k に依存せず $l_k(0) = 1$ となる．この初期条件を式 (29) に代入して解を求めると，

$$\frac{dl}{dt} = 0, \quad l(t) = 1, \tag{33}$$

となる．線形安定性解析より $l(t) = 1$ は安定である．教師の結合荷重 $\boldsymbol{B}$ も平均 0 分散 1 の確率分布から独立に生成したので，各生徒のパーセプトロンの結合荷重 $\boldsymbol{J}^k$ と教師の結合荷重 $\boldsymbol{B}$ のオーバーラップ R_k の初期値も k に依存せず $R_k(0) = 0$ となる．$l(t) = 1$ を式 (30) に代入すると，

$$\frac{dR_k}{dt} = 1 - R_k, \quad R_k(t) = 1 - \exp(-t), \tag{34}$$

となり，R_k はは添え字 k に因らなくなり以下では R と書く．またこのとき，$\boldsymbol{J}^k$ と $\boldsymbol{J}_{k'}$ は，初期状態では独立になるため $q_{kk'}(0) = 0$ となり，

$$\frac{dq_{kk'}}{dt} = 1 - q_{kk'}, \quad q_{kk'} = 1 - \exp(-t) \tag{35}$$

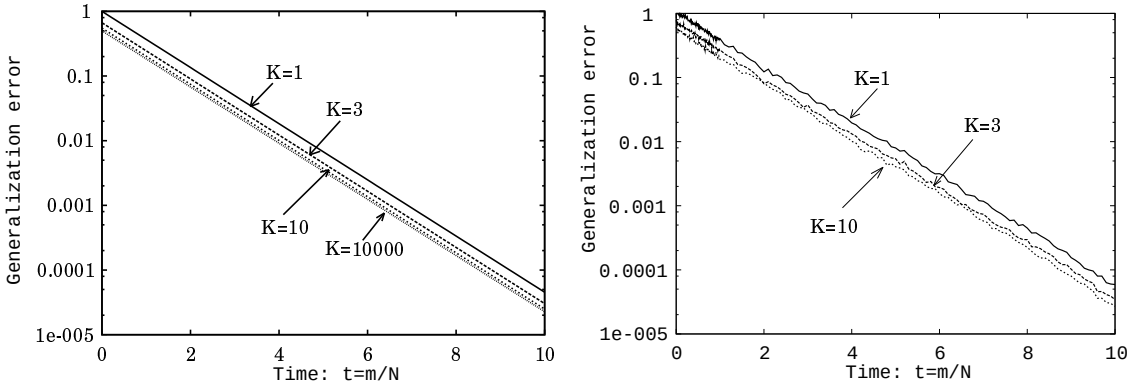
同様の理由により $q_{kk'}$ も添え字に因らなくなり以下では q と書く．ここで，R と q が同じ時間発展方程式に従うことから，$q = R$ であることがわかる．式 (33) から式 (35) および $c_k = 1/K$ を式 (32) に代入し整理すると，

$$\epsilon_g = \frac{1}{2}\left\{\frac{1-R}{K} + (1-R)\right\} \tag{36}$$

$$= \frac{1}{2}\left\{\frac{\mathrm{e}^{-t}}{K} + \mathrm{e}^{-t}\right\} \tag{37}$$

となる．これらの式で，右辺括弧内の第 1 項は生徒のパーセプトロンの数 K に依存した量であり，パーセプトロンの数が無限大の極限では 0 に収束する．一方，第 2 項は生徒のパーセプトロンの数 K に依存しない量であり，残留誤差である．式 (36) で $K = 1$ と置くと，汎化誤差は $\epsilon_g = 1 - R$ となり，単一のパーセプトロンの汎化誤差に一致する．また，K を無限大とした場合，アンサンブル学習の汎化誤差は $(1 - R)/2$ に漸近する．つまり，K 無限大の極限では，アンサンブル学習は単一の学習機械の汎化誤差の $1/2$ に収束することがわかった．

式 (37) を用いて求めた汎化誤差の K 依存性を図 3(a) に示す．図中で上から $K = 1, 3, 10, 10000$ の結果を示す．横軸は時刻であり，単位時間は入力パターンを N 回提示し学習したことに対応する．図のように K が大きくなるにしたがって，汎



(a) 理論解析　　(b) 数値シミュレーション

図 3　汎化誤差に対するアンサンブル学習に用いる単純パーセプトロン数 K の依存性

Fig. 3　Dependency of the generalization error of the ensemble learning related to number of the simple perceptrons K.

化誤差が $K=1$ の $1/2$ の値に $1/K$ で収束することがわかる．図 3(b) は数値シミュレーションの結果である．$N=1000$ とした．図中で一番上の曲線から $K=1,3,10$ の結果を示す．理論解析と数値シミュレーションの結果は良く一致した．このことから理論解析の妥当性が示せた．理論と数値シミュレーションが一致したことから，以下の解析では理論解析のみを用いることとする．

4.3　生徒の結合荷重が統計的に一様でない場合

つぎに，生徒の結合荷重が統計的に一様でない場合について考える．簡単のために，学習の初期状態において $l(0)=1$ であると仮定する．R_k と $q_{kk'}$ の初期値は任意の値 $R_k(0)$ と $q_{kk'}(0)$ をとるものとする．この場合，式 (29) から (31) の微分方程式は解析的に解くことができて，

$$l_k(t) = 1 \tag{38}$$

$$R_k(t) = 1-(1-R_k(0))\exp(-t) \tag{39}$$

$$q_{kk'}(t) = 1-(1-q_{kk'}(0))\exp(-t) \tag{40}$$

となる．これらの式を汎化誤差の式 (28) に代入すると，

$$\begin{aligned}
&\epsilon_g(\{l_k\},\{R_k\},\{q_{kk'}\},\{c_k\}) \\
&= \exp(-t)\left\{1-R_K(0)-\sum_{k=1}^{K-1}c_k(t)(R_k(0)-R_K(0))\right. \\
&+ \sum_{k=1}^{K-1}c_k^2(t)(1-q_{kK}(0)) - \sum_{k=1}^{K-1}c_k(t)(1-q_{kK}(0)) \\
&\left. + \sum_{k=1}^{K-1}\sum_{k=2}^{K-1}c_k(t)(1+q_{kk'}(0)-q_{kK}(0)-q_{k'K}(0))\right\}
\end{aligned} \tag{41}$$

ただし，$c_K = 1-\sum_{k=1}^{K-1}c_k$ を用いた．

ここまで平均の荷重 c_k は設計者が与えるものとしたが，与えられたオーダーパラメータに対して，汎化誤差が最小になるように c_k を決めることができる．このように何らかの基準を用いて，一様な平均ではなく，荷重 c_k を用いて平均を行なうアンサンブル学習法をパラレルブーティングと呼ぶ [3]．例えば，生徒のうち結合荷重の初期値が $\boldsymbol{J}^1(0)=\boldsymbol{J}^2(0)$ であり，$\boldsymbol{J}^3(0)$ が他の 2 個と独立であった場合を考えよう．この場合，単純な平均をとると同じ結合荷重を持つ生徒パーセプトロンを重みをつけることになるので，$c_1=c_2=0.25$，$c_3=0.5$ とする方が汎化誤差が小さくなることは容易に想像つくだろう．

今回紹介した枠組の中で，汎化誤差を最小にする重み c_k をもとめる一つの方法は，

$$\frac{\partial \epsilon_g}{\partial c_k}=0, \quad k=1,2,\cdots,K-1, \tag{42}$$

を満たす c_k を求めることである．一般にはオーダーパラメータ l_k，R_k，$q_{kk'}$ は時間変化するので，式 (42) を満たす荷重もも時間の関数 $c_k(t)$ となることが予想される．しかしながら，汎化誤差を最小にする重み c_k は式 (41) を C_k で偏微分して 0 と等しいと置くので，c_k は初期値 $R_k(0)$，$q_{kk'}(0)$ にのみにしか依存することがわかる．

図 4 に $K=3$ の場合のバギングとパラレルブースティングの汎化誤差の差を示す．図中のラベル eg^B-eg^{PB} はバギングとパラレルブースティングの汎化誤差の差を表す．図より，学習の初期ではパラレルブースティングにより汎化誤差が大きく改善され，学習が進むにしたがって，その効果が小さくなることが分かった．

5.　非線形パーセプトロン

非線形パーセプトロン $F(x)=\mathrm{sgn}(x)$ について簡単に述べる．詳細はこの後の発表を参考にされたい [8]．前節の線形パーセプトロンの議論から，アンサンブル学習においては q と R の関係が本質的であることが予想された．そこで，式 (9) から (11) の三つの学習則について，q と R の関係を明確にするため，R と q を軸に取ったベクトル軌跡を図 5 に示す．この図を見ると，三つの学習則のうち，R と比較して q がもっとも小さい学習則はアダトロン学習であることがわかる．言い換えるならば，q の立ち上がりが最も遅く，生徒の多様性が長時間維持される学習則はアダトロン学習である．よって学習の初期で，アンサンブル学習を行うメリットがもっとも大きい学習則はアダトロン学習であることが期待され，予想通りの結果が得られ

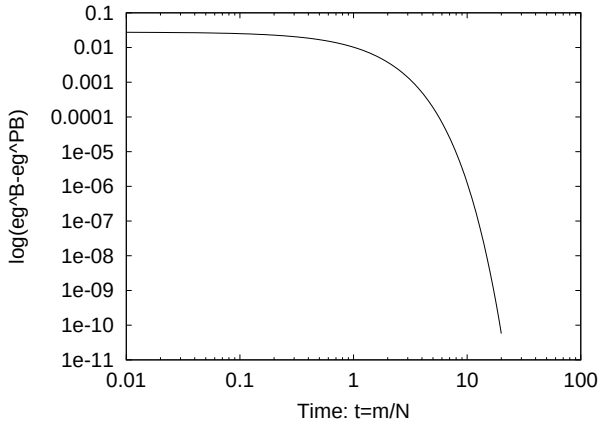


図 4　パラレルブースティングとバギングの比較

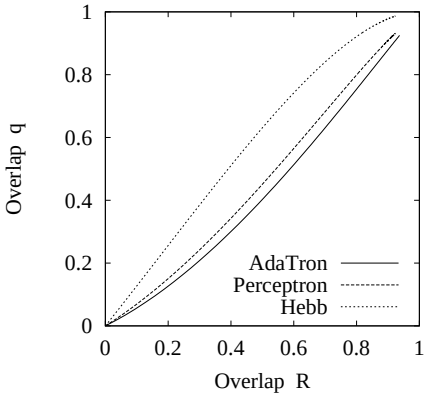


図 5　R と q の関係（理論）

た [8].

6. ま と め

アンサンブル学習の一種であるバギングやパラレルブースティングを統計力学的なオンライン学習の理論で議論した．アンサンブル学習機械の出力は各パーセプトロンの重み付き平均で与えた．まず，このアンサンブル学習機械の汎化誤差を導出し，汎化誤差がオーダーパラメーターである結合荷重の大きさ l_k，教師と生徒のオーバーラップ R_k および生徒同士の相関 $q_{kk'}$ だけで書けることを示した．さらに，これらのオーダーパラメーターが従う従う連立微分方程式を導出した．

この理論をまず線形パーセプトロンに適用した．教師と生徒の各パーセプトロンを平均 0，分散 1 のガウス乱数で初期化した場合，オーダーパラメータである結合荷重の長さ l, 教師と生徒の結合荷重のオーバーラップ R, 生徒の結合係数の相関 q の初期値は一様になる．この場合の，重みを均一にし，平均を出力とするバギングが有効である．我々はオーダーパラメータのダイナミクス，およびオーダーパラメータで記述される汎化誤差のダイナミクスを求めた．その結果，生徒のパーセプトロンの数 K によって減少し，アンサンブル学習が有効に働く項と，K に依存しない項があることが明らかにした．そして $K \to \infty$ の極限では，単一の線形パーセプトロンの汎化誤差の $1/2$ に収束することが分かった．また，有限の K に対しては，$K \to \infty$ の値に対し，$1/K$ で収束することがわかった．一方，初期の生徒の結合荷重の相関が一様でない場合に付いても解析を行なった．この場合には重みを最適化するパラレルブースティングが有効である．パラレルブースティングでは汎化誤差を最小にするように平均を求める重みを決定する．最適な重みはオーバーラップ，結合荷重の相関の初期値にしか依存しないことがわかった．最後に，理論を非線形パーセプロンに適用し，学習アルゴリズムによってアンサンブル学習の効果が異なることを示した．

謝　　辞

この研究の一部は科学研究費補助金（課題番号 13780313, 14084212, 14580438, 15500151）の援助を受けた．

文　　献

[1] Breiman L., Machine Learning, **24**, 123 (1996).
[2] Freund Y., Shapire R.E., Journal of Comp. and Sys. Sci., 55 119 (1997).
[3] 山名美智子，中原裕之，Massimiliano PONTIL，甘利俊一，”パラレルブースティングによるカーネルマシンを用いたアンサンブル学習”, 信学技法 NC2002-52, 47 (2002).
[4] Urbanczik R., ”Online learning with ensembles,” Phys. Rev. E, **62**, 1448 (2000).
[5] 西森秀稔, スピングラス理論と情報統計力学，岩波書店 (1999).
[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”，情報論的学習理論ワークショップ　予稿集 pp. 113-118, 9 月　富士吉田 (2002).
[7] 原一之，岡田真人，”パラレルブースティングのオンラインラーニングの理論”，信学技法，NC2003-14, pp. 13–18 (2003).
[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”，信学技法 NC2003-??, ?? (2003).
[9] 原 一之，岡田真人，”マージンを用いた単純パーセプトロンの学習法のオンランラーニングの理論”，電子情報通信学会和文論文誌 (D-II), **J85-D-II**, No.10, 1563-1570, (2002).
[10] 岡田真人，原 一之，”学習の問題を統計力学で取り扱う: 線形パーセプトロンのアンサンブル学習を一例として”，Computer Today 3 月号 情報論的学習理論- 機械学習のさまざまな形 -, pp. 23-28，(2003).