

アンサンブル教師に対するオンライン学習の解析

Analysis of On-line Learning for Ensemble Teachers

三好 誠司*
Seiji MIYOSHI

岡田 真人†
Masato OKADA

Abstract: We have analyzed the generalization performance of a student in a model composed of linear perceptrons: a true teacher, ensemble teachers, and the student. Calculating the generalization error of the student analytically using statistical mechanics in the framework of on-line learning, it has been proven that when the learning rate $\eta < 1$, the larger the number K and the variety of the ensemble teachers are, the smaller the generalization error is. On the other hand, when the learning rate $\eta > 1$, the smaller the number and the variety of the ensemble teachers are, the smaller the generalization error is. When the variety of the ensemble teachers is rich enough, the direction cosine between the true teacher and the student becomes unity in the limit of $\eta \rightarrow 0$ and $K \rightarrow \infty$.

Keywords: ensemble teachers, on-line learning, generalization error, statistical mechanics, learning rate

1 まえがき

学習とは観測データを用いてその背後にあるデータの生成過程を推定することである。教師つき学習においては観測データは教師の入出力であり、これは例題とも呼ばれる。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。

これまでに我々はオンライン学習の枠組みで、特にアンサンブル学習 [2, 3, 4, 5, 6] の汎化能力について統計力学的手法を用いた解析を行ってきた [7, 8, 9, 10]。また最近同様の手法を用いて、教師が真の教師と一定の方向余弦を保ったまま真の教師のまわりを動き続ける場合の生徒の汎化能力についても解析を行った [11]。その結果、生徒が動く教師の入出力だけを例題として使用するにもかかわらず、動く教師の汎化誤差よりも生徒の汎化誤差の方が小さくなりうるということが明らかになった。現実の人間社会においても、生徒が入出力を観測できる教師は必ずしも正しい解答を示すとは限らず、また、教師自身も学習しており、変わり続ける存在である場合が多いことから、この解析は統計的学習理論と現実社会のアナロジーを考える上で興味深いものである。

さて、現実の人間社会においては、生徒が複数の教師の入出力を観測することができ [12]、また、その教師たちが少しずつ違っているという場合も多いことから、本論文では、このようなモデルの解析を行い、不完全な複数の教師たちの使い方について議論する。すなわち、まず真の教師というものを考え、 K 個の教師 (アンサンブル教師と呼ぶことにする) がそのまわりに存在するもの

*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
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

とする．生徒はアンサンブル教師の入出力を順番に例題として用いる．特に本論文では，真の教師，アンサンブル教師，生徒のいずれもが雑音の重畳された線形パーセプトロン [7] である場合を考え，オンライン学習の枠組みで統計力学的手法を用いることにより巨視的変数や汎化誤差を解析的に求める．その結果，生徒の学習係数 η が 1 より小さい場合には教師数 K が多いほど，また，教師の多様性が豊かであるほど生徒の汎化誤差が小さくなるが， η が 1 より大きい場合には逆であることが明らかになる．また，教師の多様性が十分に豊かであるときには， $\eta \rightarrow 0, K \rightarrow \infty$ の極限で真の教師と生徒の方向余弦が 1 になることが明らかになる．

2 モデル

本論文では 1 個の真の教師， K 個のアンサンブル教師，1 個の生徒を考える．それらはすべて線形パーセプトロンであるとし，それぞれの結合荷重を A, B_k, J とする．ただし $k = 1, \dots, K$ である．なお，本論文では簡単のため真の教師の結合荷重，アンサンブル教師の結合荷重，生徒の結合荷重のことをそれぞれ単に真の教師，アンサンブル教師，生徒と呼ぶことにする．真の教師 $A = (A_1, \dots, A_N)$ ，アンサンブル教師 $B_k = (B_{k1}, \dots, B_{kN})$ ，生徒 $J = (J_1, \dots, J_N)$ および入力 $x = (x_1, \dots, x_N)$ は N 次元ベクトルであり， A の各要素 A_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． B_k の N 個の要素 B_{ki} のうちいくつかは A の要素 A_i に -1 を乗じた値に等しく，それ以外は A_i に等しい．いま，どの要素 B_{ki} が $-A_i$ に等しいかが A_i の値に無関係であるとする． B_{ki} も A_i 同様に平均 0，分散 1 のガウス分布にしたがう．また， B_{ki} は A_i 同様に不変であるとする． B_k と A の方向余弦は R_{Bk} ， B_k と $B_{k'}$ の方向余弦は $q_{kk'}$ であるとする． J の初期値 J^0 の各要素 J_i^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， J と A の方向余弦は R_J ， J と B_k の方向余弦は R_{BkJ} 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，

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

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

$$\langle J_i^0 \rangle = 0, \quad \langle (J_i^0)^2 \rangle = 1, \quad (3)$$

$$\langle x_i \rangle = 0, \quad \langle (x_i)^2 \rangle = \frac{1}{N}, \quad (4)$$

$$R_{Bk} = \frac{A \cdot B_k}{\|A\| \|B_k\|}, \quad q_{kk'} = \frac{B_k \cdot B_{k'}}{\|B_k\| \|B_{k'}\|}, \quad (5)$$

$$R_J = \frac{A \cdot J}{\|A\| \|J\|}, \quad R_{BkJ} = \frac{B_k \cdot J}{\|B_k\| \|J\|}. \quad (6)$$

ここで， $\langle \cdot \rangle$ は平均を表す．真の教師 A ，アンサンブル教師 B_k ，生徒 J および方向余弦 $q_{kk'}$ ， R_{Bk} ， R_J ， R_{BkJ} の関係を図 1 に示す．

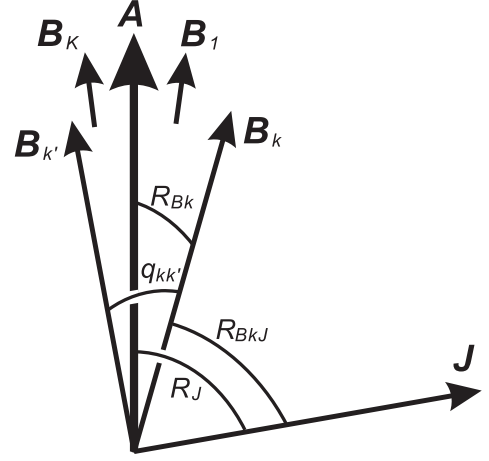


図 1: 真の教師 A ，アンサンブル教師 B_k ，生徒 J ． $q_{kk'}$ ， R_J ， R_{Bk} ， R_{BkJ} は方向余弦である．

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，

$$\|A\| = \sqrt{N}, \quad \|B_k\| = \sqrt{N}, \quad \|J^0\| = \sqrt{N}, \quad \|x\| = 1. \quad (7)$$

となる．生徒の大きさ $\|J\|$ 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． m は時間ステップである．

真の教師の出力は $y^m + n_A^m$ ，アンサンブル教師の出力は $v_k^m + n_{Bk}^m$ ，生徒の出力は $u^m l^m + n_J^m$ である．ここで，

$$y^m = A \cdot x^m, \quad (8)$$

$$v_k^m = B_k \cdot x^m, \quad (9)$$

$$u^m l^m = J^m \cdot x^m, \quad (10)$$

$$n_A^m \sim \mathcal{N}(0, \sigma_A^2), \quad (11)$$

$$n_{Bk}^m \sim \mathcal{N}(0, \sigma_{Bk}^2), \quad (12)$$

$$n_J^m \sim \mathcal{N}(0, \sigma_J^2). \quad (13)$$

である． $\mathcal{N}(0, \sigma^2)$ は平均 0，分散 σ^2 のガウス分布を表す．つまり，真の教師の出力，アンサンブル教師の出力，生徒の出力にはそれぞれ分散 $\sigma_A^2, \sigma_{Bk}^2, \sigma_J^2$ の互いに独立なガウス雑音が重畳されている．またこのとき， y, v_k, u は平均 0，分散 1 のガウス分布にしたがう確率変数となる．

いま，真の教師 A とアンサンプル教師 B_k の誤差 ϵ_{Bk} を両者の出力の二乗誤差で定義する．すなわち，

$$\epsilon_{Bk}^m \equiv \frac{1}{2} (y^m + n_A^m - v_k^m - n_{Bk}^m)^2. \quad (14)$$

同様に，アンサンプル教師 B_k と生徒 J の誤差 ϵ_{BkJ} を両者の出力の二乗誤差で定義する．すなわち，

$$\epsilon_{BkJ}^m \equiv \frac{1}{2} (v_k^m + n_{Bk}^m - u^m l^m - n_J^m)^2. \quad (15)$$

生徒 J は入力 x とそれに対する K 個のアンサンプル教師 B_k の出力を順番に用いて勾配法により学習を行うものとする．すなわち，

$$\mathbf{J}^{m+1} = \mathbf{J}^m - \eta \frac{\partial \epsilon_{BkJ}^m}{\partial \mathbf{J}^m} \quad (16)$$

$$= \mathbf{J}^m + \eta (v_k^m + n_{Bk}^m - u^m l^m - n_J^m) \mathbf{x}^m, \quad (17)$$

$$k = \text{mod}(m, K) + 1. \quad (18)$$

ここで， η は生徒の学習係数であり定数とする．また， $\text{mod}(m, K)$ は m を K で割った余りを表す．

学習則を一般化すると以下のように表せる．

$$\mathbf{J}^{m+1} = \mathbf{J}^m + f_k \mathbf{x}^m \quad (19)$$

$$= \mathbf{J}^m + f(v_k^m + n_{Bk}^m, u^m l^m + n_J^m) \mathbf{x}^m, \quad (20)$$

$$k = \text{mod}(m, K) + 1. \quad (21)$$

ここで， f は更新量を表す関数であり学習則によって決定される．

また，真の教師 A と生徒 J の誤差 ϵ_J も両者の出力の二乗誤差で定義しておく．すなわち，

$$\epsilon_J^m \equiv \frac{1}{2} (y^m + n_A^m - u^m l^m - n_J^m)^2. \quad (22)$$

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を理論的に求めることである．汎化誤差は未知の入力に関する真の教師との誤差の平均であるから，個々のアンサンプル教師 B_k の汎化誤差 ϵ_{Bkg} ，生徒 J の汎化誤差 ϵ_{Jg} はそれぞれ以下のように計算される．なお，以後は時間ステップを表す添字 m は特に必要な場合を除いて省略する．

$$\epsilon_{Bkg} = \int d\mathbf{x} dn_A dn_{Bk} P(\mathbf{x}, n_A, n_{Bk}) \epsilon_{Bk} \quad (23)$$

$$= \int dy dv_k dn_A dn_{Bk} P(y, v_k, n_A, n_{Bk}) \times \frac{1}{2} (y + n_A - v_k - n_{Bk})^2 \quad (24)$$

$$= \frac{1}{2} (-2R_{Bk} + 2 + \sigma_A^2 + \sigma_{Bk}^2), \quad (25)$$

$$\epsilon_{Jg} = \int d\mathbf{x} dn_A dn_J P(\mathbf{x}, n_A, n_J) \epsilon_J \quad (26)$$

$$= \int dy du dn_A dn_J P(y, u, n_A, n_J) \times \frac{1}{2} (y + n_A - ul - n_J)^2 \quad (27)$$

$$= \frac{1}{2} (-2R_{Jl} + (l)^2 + 1 + \sigma_A^2 + \sigma_J^2). \quad (28)$$

ここで積分の実行には， y, v_k, u が平均 0，分散 1 のガウス分布に従うこと， y と v_k の共分散が R_{Bk} ， v_k と u の共分散が R_{BkJ} ， y と u の共分散が R_J であること，および， n_A, n_{Bk}, n_J はいずれも他の確率変数とは独立であることを利用した．

3.2 巨視的変数の微分方程式とその解

解析を容易にするため，以下の補助的な巨視的変数を導入する．

$$r_J \equiv R_{Jl}, \quad (29)$$

$$r_{BkJ} \equiv R_{BkJl}. \quad (30)$$

今回，巨視的変数のダイナミクスを記述する連立微分方程式 [13, 14] を熱力学的極限における自己平均性に基づき以下のような決定論的な形で導出した [8]．ここで次元 N はアンサンプル教師の数 K よりも十分大きいものとし，また，時間ステップ m を次元 N で正規化した値を時間 t としている．

$$\frac{dr_{BkJ}}{dt} = \frac{1}{K} \sum_{k'=1}^K \langle f_{k'} v_k \rangle, \quad (31)$$

$$\frac{dr_J}{dt} = \frac{1}{K} \sum_{k=1}^K \langle f_k y \rangle, \quad (32)$$

$$\frac{dl}{dt} = \frac{1}{K} \sum_{k=1}^K \left(\langle f_k u \rangle + \frac{1}{2l} \langle f_k^2 \rangle \right). \quad (33)$$

本論文では線形なパーセプトロンを考えているので，これらの連立微分方程式に現れるサンプル平均は以下のように容易に計算することができる．

$$\langle f_k u \rangle = \eta \left(\frac{r_{BkJ}}{l} - l \right), \quad (34)$$

$$\langle f_k^2 \rangle = \eta^2 \left((l)^2 - 2r_{BkJ} + 1 + \sigma_{Bk}^2 + \sigma_J^2 \right), \quad (35)$$

$$\langle f_k y \rangle = \eta (R_{Bk} - r_J), \quad (36)$$

$$\frac{1}{K} \sum_{k'=1}^K \langle f_{k'} v_k \rangle = \eta \left(-r_{BkJ} + \frac{1}{K} \sum_{k'=1}^K q_{kk'} \right) \quad (37)$$

本論文では真の教師 A ，生徒 J の初期値 \mathbf{J}^0 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直交しており,

$$R_J^0 = 0 \quad (38)$$

である. また,

$$l^0 = 1 \quad (39)$$

である. 式 (34)–(39) を用いて連立微分方程式 (31)–(33) は以下のように解析的に解ける.

$$r_{BkJ} = \frac{1}{K} \sum_{k'=1}^K q_{kk'} (1 - e^{-\eta t}), \quad (40)$$

$$r_J = \frac{1}{K} \sum_{k=1}^K R_{Bk} (1 - e^{-\eta t}), \quad (41)$$

$$\begin{aligned} (l)^2 &= \frac{1}{2-\eta} \left[2(1-\eta)\bar{q} + \eta(1 + \sigma_B^2 + \sigma_J^2) \right] \\ &+ \left[1 + \frac{1}{2-\eta} \left(\eta(1 + \sigma_B^2 + \sigma_J^2) - 2\bar{q} \right) \right] \\ &\times e^{\eta(\eta-2)t} - 2\bar{q}e^{-\eta t}, \end{aligned} \quad (42)$$

ここで,

$$\bar{q} = \frac{1}{K^2} \sum_{k=1}^K \sum_{k'=1}^K q_{kk'}, \quad (43)$$

$$\sigma_B^2 = \frac{1}{K} \sum_{k=1}^K \sigma_{Bk}^2, \quad (44)$$

である.

4 結果と議論

以下ではアンサンプル教師と真の教師の方向余弦, アンサンプル教師間の方向余弦, および, それらに重畳される雑音の分散が一様である場合について計算を行う. すなわち,

$$R_{Bk} = R_B, \quad k = 1, \dots, K, \quad (45)$$

$$q_{kk'} = \begin{cases} q, & k \neq k', \\ 1, & k = k', \end{cases} \quad (46)$$

$$\sigma_{Bk}^2 = \sigma_B^2, \quad (47)$$

であるとする. この場合, 式 (43), (44) は,

$$\bar{q} = q + \frac{1-q}{K}, \quad (48)$$

$$\sigma_B^2 = \sigma_B^2, \quad (49)$$

となる.

式 (28), (29), (40)–(49) を用いて理論的に計算される汎化誤差 ϵ_{Jg} のダイナミクスを計算機シミュレシ

ョンの結果と重ねて図 2 に示す. 計算機シミュレーションは $N = 2000$ で実行し, 汎化誤差は各時点で 10^4 個のランダム入力に対する誤差の平均を計算することにより計算した. アンサンプル教師の汎化誤差 ϵ_{Bg} も重ねて描いた. また, このときの R_J と l のダイナミクスを図 3 に示す.

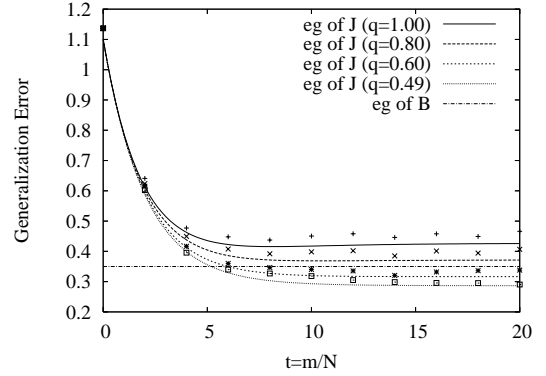


図 2: 汎化誤差 ϵ_{Jg} のダイナミクス. 理論と計算機シミュレーション. q 以外の条件は $\eta = 0.3, K = 3, R_B = 0.7, \sigma_A^2 = 0.0, \sigma_B^2 = 0.1, \sigma_J^2 = 0.2$ である.

これらの図において曲線は理論計算の結果を, +, x, *, □, などの印は計算機シミュレーションの結果を表す. また, q 以外の条件は共通で $\eta = 0.3, K = 3, R_B = 0.7, \sigma_A^2 = 0.0, \sigma_B^2 = 0.1, \sigma_J^2 = 0.2$ である. これらを見ると以下のことがわかる. 図 2 より, q が小さいほど, すなわち, アンサンプル教師の多様性が大きいほど生徒の汎化誤差 ϵ_{Jg} は小さい. 特に $q = 0.6$ と $q = 0.49$ の場合には $t = 5$ 付近で生徒の汎化誤差がアンサンプル教師の汎化誤差よりも小さくなっている. すなわち, このモデルの生徒はアンサンプル教師の入出力だけを例題として学習に用いるにもかかわらず, アンサンプル教師よりも賢くなりうる. また, 図 3 より, アンサンプル教師の多様性が大きいほど真の教師と生徒の方向余弦 R_J は大きく, 生徒の長さ l は小さいことがわかる. なお, 図 2, 図 3 において q の最小値 (0.49) を $R_B(0.7)$ の 2 乗としている理由については後述する.

図 2, 図 3 を見ると, 生徒の汎化誤差 ϵ_{Jg} や R_J, l は $t = 20$ でほぼ定常値に達しているように見えるが, 今回巨視的変数が解析的に得られているのでこれらの $t \rightarrow \infty$ におけるふるまいについては理論的な洞察が可能である. すなわち, 式 (40)–(42) の指数関数のべきの符号に着目することにより $0 < \eta < 2$ でなければ生徒の汎化誤差 ϵ_{Jg} や長さ l は発散することがわかる. $0 < \eta < 2$ の場合については, 式 (40)–(42) において $t \rightarrow \infty$ とすることにより汎化誤差 ϵ_{Jg} や R_J, l の定常値は以下のように

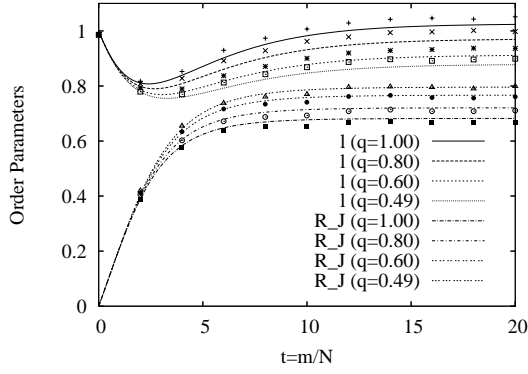


図 3: R_J と l のダイナミクス. 理論と計算機シミュレーション. q 以外の条件は $\eta = 0.3, K = 3, R_B = 0.7, \sigma_A^2 = 0.0, \sigma_B^2 = 0.1, \sigma_J^2 = 0.2$ である.

容易に得られる.

$$r_{BkJ} \rightarrow q + \frac{1-q}{K}, \quad (50)$$

$$r_J \rightarrow R_B, \quad (51)$$

$$l^2 \rightarrow \frac{1}{2-\eta} \times \left(2(1-\eta) \left(q + \frac{1-q}{K} \right) + \eta(1 + \sigma_B^2 + \sigma_J^2) \right). \quad (52)$$

式 (50)–(52) より以下のことがわかる. $\eta = 1$ の場合には長さ l の定常値は教師数 K や教師間の類似度 q に依存しない. よってこの場合, 汎化誤差 ϵ_{Jg} や方向余弦 R_J の定常値は K や q に依存しない. $0 < \eta < 1$ の場合には q が小さいほど, また, K が大きいほど l と ϵ_{Jg} の定常値は小さくなり, R_J の定常値は大きくなる. $1 < \eta < 2$ の場合には逆に q が小さいほど, また, K が大きいほど l と ϵ_{Jg} の定常値は大きくなり, R_J の定常値は小さくなる.

すなわち, $\eta < 1$ の場合には教師の数が多いほど, また, 教師の多様性が豊かであるほど生徒は賢くなれるのに対し, $\eta > 1$ の場合には逆に生徒が賢くなるためには教師の数は少ない方が良く, また, 教師の多様性は乏しい方が良い.

式 (52) より $\eta \rightarrow 0, K \rightarrow \infty$ の極限で $l \rightarrow \sqrt{q}$ となるのでこのとき式 (29), (51) より $R_J \rightarrow R_B/\sqrt{q}$ となる. ところで一般に次元が大きいあるベクトル X があって, X との方向余弦が R_0 であるような二個のベクトル Y と Z を独立に生成したとき, Y と Z の方向余弦は $q_0 = R_0^2$ である. このことから, 真の教師との方向余弦が R_B 一定という拘束条件のもとでアンサンブル教師たちが独立に生成されたほど十分な多様性を持っているとき, 学習係数 η が小さく, かつ, 教師数 K

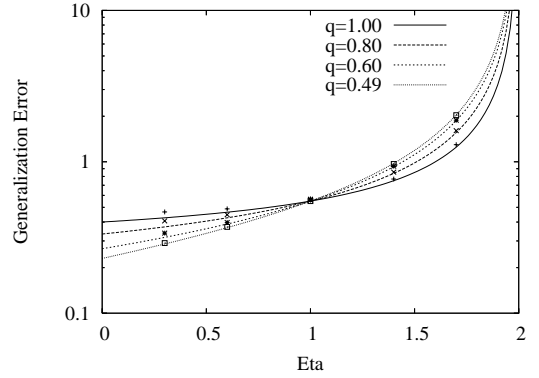


図 4: $K = 3$ の場合の汎化誤差 ϵ_{Jg} の定常値. 理論と計算機シミュレーション. K, q 以外の条件は $R_B = 0.7, \sigma_A^2 = 0.0, \sigma_B^2 = 0.1, \sigma_J^2 = 0.2$ である.

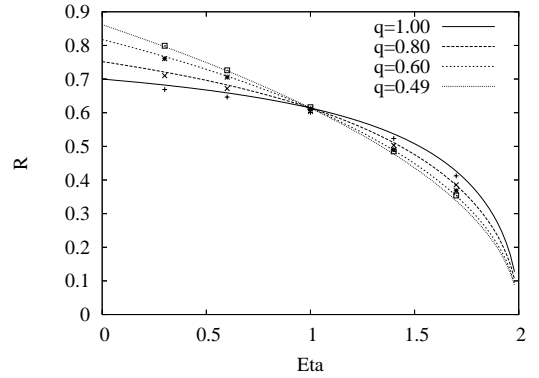


図 5: $K = 3$ の場合の R_J の定常値. 理論と計算機シミュレーション. q 以外の条件は K, q 以外の条件は $R_B = 0.7, \sigma_A^2 = 0.0, \sigma_B^2 = 0.1, \sigma_J^2 = 0.2$ である.

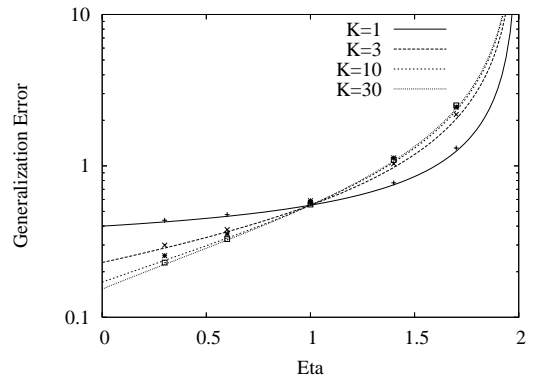


図 6: $q = 0.49$ の場合の汎化誤差 ϵ_{Jg} の定常値. 理論と計算機シミュレーション. q, K 以外の条件は $R_B = 0.7, \sigma_A^2 = 0.0, \sigma_B^2 = 0.1, \sigma_J^2 = 0.2$ である.

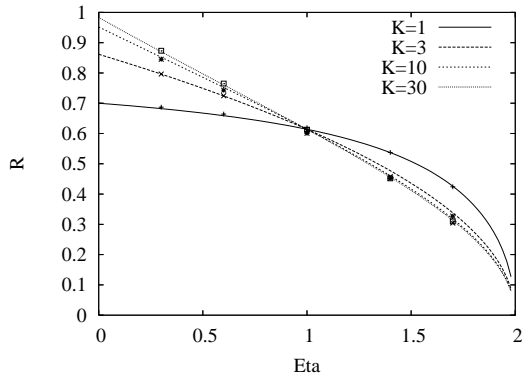


図 7: $q = 0.49$ の場合の R_J の定常値. 理論と計算機シミュレーション. q, K 以外の条件は K, q 以外の条件は $R_B = 0.7, \sigma_A^2 = 0.0, \sigma_B^2 = 0.1, \sigma_J^2 = 0.2$ である.

が多い極限で, またそのときに限り, 生徒と真の教師の方向余弦 R_J は雑音の分散によらず 1 になる.

生徒の学習係数 η と汎化誤差 ϵ_{Jg} , 方向余弦 R_J の定常値の関係の計算例を図 4 ~ 図 7 に示す. これらのうち, 図 4, 5 は $K = 3$ 固定で q を変えた場合であり, 図 6, 7 は $q = 0.49$ 固定で K を変えた場合である. K, q 以外の条件は図 2, 図 3 と同様 $R_B = 0.7, \sigma_A^2 = 0.0, \sigma_B^2 = 0.1, \sigma_J^2 = 0.2$ である. また, 計算機シミュレーションは $\eta = 0.3, 0.6, 1.0, 1.4, 1.7$ で実行し, いずれの場合においても十分定常に達していると判断される $t = 20$ の値をプロットしている.

これらの図から以下のことがわかる. 学習係数 η が小さいほど汎化誤差 ϵ_{Jg} は小さく, 方向余弦 R_J は大きい. また, $\eta = 2$ で ϵ_{Jg} が発散し, R_J がゼロになる相転移現象が起こることが確認される. $\eta < 1$ の場合には, アンサンブル教師の数 K が多いほど, また, アンサンブル教師の類似度 q が小さいほど (アンサンブル教師の多様性が豊かであるほど) 汎化誤差 ϵ_{Jg} は小さくなり, 方向余弦 R_J は大きくなる. これに対し, $\eta > 1$ の場合には逆になっている.

5 むすび

真の教師, そのまわりに存在する複数のアンサンブル教師, 生徒のいずれもが雑音の重畳された線形なパーセプトロンである場合を考え, オンライン学習の枠組みで統計力学的手法を用いることにより汎化誤差を解析的に求めた結果, 生徒の学習係数 η が 1 より小さい場合には教師数 K が多いほど, また, 教師の多様性が豊かであるほど生徒の汎化誤差が小さくなるが, η が 1 より大きい場合には逆であることが明らかになった. また, 教師の多様性が十分に豊かであるときには, $\eta \rightarrow 0, K \rightarrow \infty$ の極限で真の教師と生徒の方向余弦が 1 になることが明

らかになった.

謝辞

本論文の一部は科学研究費補助金 特定領域研究 (課題番号 14084212), 同 基盤研究 (C) (課題番号 14580438, 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] Saad, D. (ed.), On-line Learning in Neural Networks, Cambridge University Press, (1998)
- [2] 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).
- [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] Krogh, A. and Sollich, P., “Statistical mechanics of ensemble learning,” Phys. Rev. E, 55(1), 811-825 (1997).
- [6] Urbanczik, R., “Online learning with ensembles,” Phys. Rev. E, 62(1), 1448-1451 (2000).
- [7] 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, 113-118 (2002).
- [8] 岡田 真人, 原 一之, 三好 誠司, “[チュートリアル講演] アンサンブル学習”, 信学技報, NC2003-35, pp.7-12, 2003.7
- [9] Miyoshi, S., Hara, K. and Okada, M., “Analysis of ensemble learning using simple perceptrons based on on-line learning theory”, Phys. Rev. E, 71, 036116. March 2005.
- [10] 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).
- [11] 三好 誠司, 岡田 真人, “教師が真の教師のまわりをまわる場合のオンライン学習”, 信学技報, NC2005-10, pp.19-24, 2005.6
- [12] 山口 幸, 前田 圭子, 上江洲 達也, “2種類の教師信号からのオンライン学習”, 日本物理学会年次大会, 24aYB-9, 2005.3
- [13] 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.
- [14] Nishimori, H., “Statistical Physics of Spin Glasses and Information Processing: An Introduction,” Oxford University Press, (2001)