

オンライン学習理論に基づく非線形パーセプトロンのアンサンブル学習の解析

Analysis of ensemble learning using nonlinear perceptrons based on online learning theory

三好 誠司 (P)[†], 原 一之^{††}, 岡田 真人^{†††}Seiji Miyoshi (P)[†], Kazuyuki Hara^{††} and Masato Okada^{†††}[†]神戸高専 miyoshi@kobe-kosen.ac.jp, ^{††}都立工業高専, ^{†††}理研 脳総研, † 科技団 さきがけ

Abstract— Online learning with an ensemble of nonlinear perceptrons is analyzed. It decides an output by using the majority vote of K students. It is shown that the AdaTron learning is superior to the Hebbian learning and the perceptron learning in generalization performance as an ensemble.

Keywords— ensemble learning, online learning, nonlinear perceptron, 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) を多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事は一般にアンサンブル学習と呼ばれ注目されている [1–8]. 原と岡田は, 生徒と教師が線形パーセプトロンである場合についてオンライン学習の枠組みで統計力学的な解析を行った [5]. 一方, 符号関数を出力関数とするような非線形パーセプトロンの学習則としてはヘブ学習, 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 [6]. 非線形パーセプトロンの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研究としては [7] があるが, これは K が大きい極限のみを扱っている. また, 上記の三つの学習則をアンサンブル学習に適用した場合の違いについての解析はこれまで行われていない.

そこで本研究では, 符号関数を出力関数とする K 個の単純パーセプトロンによるアンサンブル学習をオンライン学習の枠組みで, かつ, 有限の K について議論する. その結果, 三つの学習則は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, アダトロン学習がもっとも優れているという興味深い事実が明らかになる.

2 モデル

K 個の生徒 J_1, J_2, \dots, J_K からなるアンサンブルを考える. 各生徒の出力は $\text{sgn}(u_1 l), \text{sgn}(u_2 l), \dots, \text{sgn}(u_K l)$ である. ここで, $u_k l = J_k \cdot x$ である. l は結合荷重の長さで, これは後述するように本研究で扱う巨視的変数のひとつである.

教師 B も符号関数を出力関数とするパーセプトロンであるとし, B は不変とする. 教師の出力は $\text{sgn}(v)$ である. ここで, $v = B \cdot x$ である.

教師と各生徒には共通の入力 x が与えられる. 個々の生徒は 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び, 修正の方法を学習則と呼ぶ. 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, x とそれに対する教師の出力だけであるから, 学習は一般に, $J_k^{m+1} = J_k^m + f_k^m x^m$, $f_k^m = f(\text{sgn}(v^m), u_k^m)$ と表せる. ここで m は時刻ステップである.

B, J_k, x は N 次元ベクトルであるとする. x の各要素は平均 0, 分散 $1/N$ のガウス分布に従う独立な確率変数であり, J_k の初期値 J_k^0 と 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るものとする. いま, $N \rightarrow \infty$ の熱力学的極限を考えると, $|x| = 1$, $|B| = \sqrt{N}$, $|J_k| = \sqrt{N}$ となる. $|J_k|$ は学習により変化するため, 初期の大きさに対する比を l とし, これを生徒の結合荷重の長さと呼ぶことに

する. すなわち, $|J_k| = l\sqrt{N}$ である.

3 理論

3.1 教師と生徒の類似度 R と生徒間の類似度 q

本研究はオンライン学習の枠組みで議論を行う. この場合, B, J_k は新たな入力パターン x とは独立になり, v, u_k は平均 0, 分散 1 のガウス分布に従う. ここで, 本研究における第二, 第三の巨視的変数として, 教師 B と生徒 J の類似度 $R = B \cdot J / |B||J|$, 生徒間の類似度 $q = J_\alpha \cdot J_\beta / |J_\alpha||J_\beta|$, $\alpha \neq \beta$ を導入する.

学習が進むにつれ, R と q は 0 から 1 に近づいていく. このとき, R と q は全く無関係な値をとるわけではない. たとえば, $R=1$ のときに $q=0$ とはなり得ないことは R と q の意味合いを考えても明らかである. このように R と q は相互に何らかの拘束関係にあるのであるが, R と比較して q が小さいほど生徒の多様性が維持されていることになるわけだから, アンサンブル学習の効果が大きいと言える. すなわち, アンサンブル学習においては R と q の関係が本質的である. すでに, 線形パーセプトロンについては, この関係が非常に明快な形で定量的に明らかにされている [5] ので, 以下では非線形パーセプトロンについて三つの学習則に関する R と q の関係を解析的に調べることにする.

3.2 巨視的変数の微分方程式 (学習則一般の場合)

自己平均性の仮定に基づき, 学習則一般の l と R に関する微分方程式はすでに求められている [6]. 本研究では新たに学習則一般の q に関する微分方程式を導出した. これらをあわせて以下に示す. ここで, $\langle f_\alpha u_\beta \rangle$ は異なる生徒の f と u の積の, $\langle f_\alpha f_\beta \rangle$ は異なる生徒の f と f の積の, $\langle fu \rangle$ は同一の生徒の f と u の積の各サンプル平均を表す.

$$\frac{dl}{dt} = \langle fu \rangle + \frac{\langle f^2 \rangle}{2l}, \quad \frac{dR}{dt} = \frac{\langle fv \rangle - \langle fu \rangle R}{l} - \frac{R}{2l^2} \langle f^2 \rangle, \quad (1)$$

$$\frac{dq}{dt} = \frac{2(\langle f_\alpha u_\beta \rangle - q \langle fu \rangle)}{l} + \frac{\langle f_\alpha f_\beta \rangle - q \langle f^2 \rangle}{l^2}. \quad (2)$$

3.3 汎化誤差

K 個の生徒が多数決により統合出力を決定する場合, 汎化誤差 ϵ_g は R, q を用いて以下のように計算される.

$$\epsilon_g = \int \prod_{k=1}^K du_k dv P(u^T) \Theta \left(-\text{sgn}(v) \sum_{k=1}^K \text{sgn}(u_k) \right), \quad (3)$$

$$P(u^T) = \frac{1}{(2\pi)^{\frac{K+1}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left(-\frac{u^T \Sigma u}{2} \right), \quad (4)$$

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_K \\ v \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & q & \dots & q & R \\ q & 1 & \dots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ q & \dots & q & 1 & R \\ R & \dots & \dots & R & 1 \end{pmatrix}. \quad (5)$$

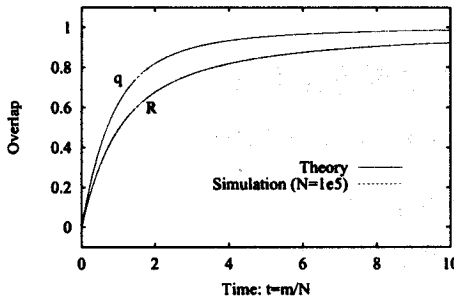


図 1: R と q (ヘブ学習)

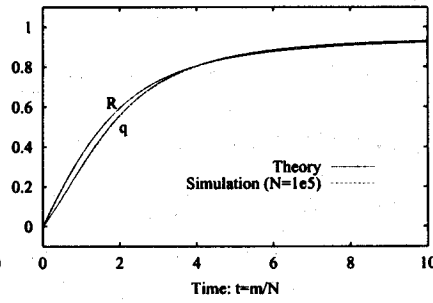


図 2: R と q (パーセプトロン学習)

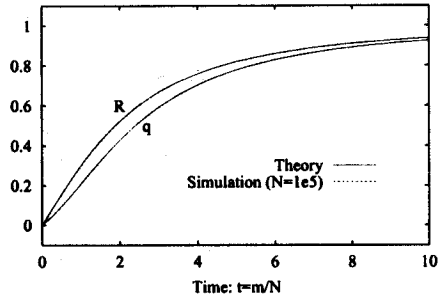


図 3: R と q (アダトロン学習)

4 ダイナミクス

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, $f = \text{sgn}(v)$, $f = \Theta(-uv) \text{sgn}(v)$, $f = -u\Theta(-uv)$ で更新を行う学習則で, その $\langle fu \rangle$, $\langle fv \rangle$, $\langle f^2 \rangle$ はすでに求められている [6]. ここでは新たに各学習則の $\langle f_\alpha u_\beta \rangle$ と $\langle f_\alpha f_\beta \rangle$ を導出した. 具体的な式の形 [8] については紙面の関係で省略する. これらと式 (1), (2) を数値的に解いて得られた結果と計算機実験 ($N = 10^5$) の比較を図 1~図 3 に示す. 理論と計算機実験の結果はほとんど重なっており, 今回導いた理論が計算機実験の結果をよく説明することがわかる. また, これらの図から, 三つの学習則のうち, R と比較して q が最も小さい学習則はアダトロン学習であることがわかる.

5 議論

すでに述べたように, アンサンブル学習においては R と q の関係が本質的である. そこで, 両者の関係をより明確にするため, R と q を軸にとって描き直したグラフを図 4 に示す. この図を見ると, アダトロン学習の特性が最も下になっている. すなわち, R と比較して q が最も小さい学習則はアダトロン学習であることが明確になった. 言い換えるならば, q の立ち上がり方が最も遅く, 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. よってこの段階で, 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る.

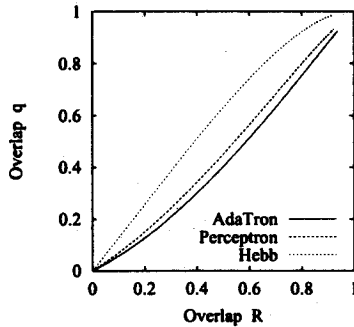


図 4: R vs. q (理論)

この予想を確認するため, 前節で求めた R , q の値と式 (3)-(5) から $K=3$ における汎化誤差を数値的に求め, 計算機実験の結果と比較するとともに, 三つの学習則を比較した. 結

果を図 5~図 7 に示す. 計算機実験では $N=10^4$ とし, 各時点で 10^5 個のランダム入力によるテストにより汎化誤差を計算した. アンサンブル学習の効果を明らかにするための比較対象として $K=1$ の場合の理論計算結果もプロットしている.

これらの図より, 今回理論計算により求めた汎化誤差が計算機実験の結果をよく説明することがわかる. また, K を 1 から 3 にすることにより汎化誤差が最も大きく改善される学習則はアダトロン学習であることがわかる. すなわち, 図 4 の R と q の関係から予想されたとおり, アンサンブルの効果が最も大きいのはアダトロン学習である. もともとアダトロン学習は三つの学習則の中でもっとも高速な漸近特性を示す [6] という長所を持つが, 学習の初期段階では遅いという短所も持つ. 今回, アダトロン学習が「生徒の多様性維持」という点でアンサンブル学習との相性の良さを有することが明らかにされ, 学習初期段階での短所もアンサンブル学習と組み合わせることにより大きく改善できることが明らかになった.

謝辞 本研究の一部は科学研究費補助金 (課題番号 13780313, 14580438, 15500151) によるものである.

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] 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).
- [6] 西森 秀俊, “スピニングラス理論と情報統計力学,” 岩波書店, 東京, 1999.
- [7] Urbanczik, R., “Online learning with ensembles,” Phys. Rev. E, 62(1), 1448-1451 (2000).
- [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,” 信学技報 NC, (2003 年 7 月).

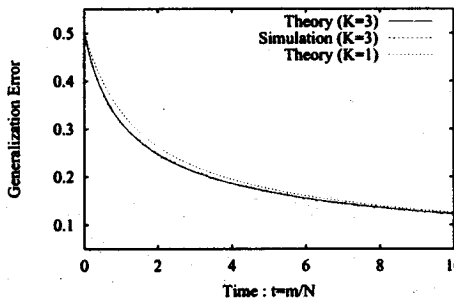


図 5: 汎化誤差 (ヘブ学習)

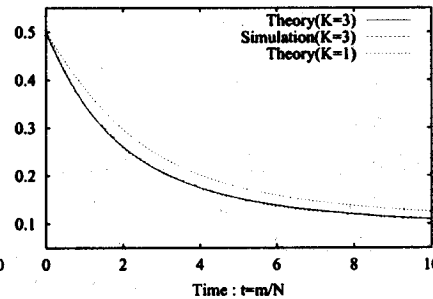


図 6: 汎化誤差 (パーセプトロン学習)

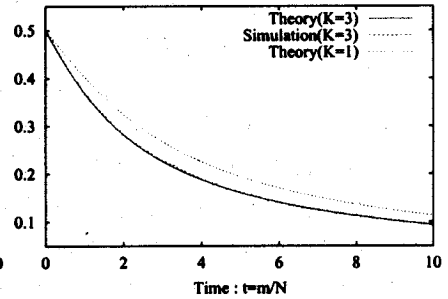


図 7: 汎化誤差 (アダトロン学習)