

マージンありパーセプトロン則によるアンサンブル学習について

日大生産工 ○原 一之

1 まえがき

パーセプトロンなどの学習機械をオンラインで学習する枠組み（以下、オンライン学習）において、汎化誤差（未知の入力に対する誤り確率）の収束が最も早い学習則はヘブ則である [4]。それに対し、我々はヘブ則より学習の収束が遅いパーセプトロン則にマージンを導入し、マージンを適応的に変化させることにより、ヘブ則より高速に汎化誤差が収束するマージンありパーセプトロン則 [1] を提案した。

一方、1 個の学習機械で学習できるケースはそれほど多くなく、一般に複数の学習機械を組み合わせる。例えば複数の学習機械を独立に学習した後、それらの多数決を取る方法が提案されている。このような学習法をアンサンブル学習と呼ぶ [2]。

本講演では、ヘブ則、パーセプトロン則を用いたアンサンブル学習を紹介した後、マージン有りパーセプトロン則を用いたアンサンブル学習の学習特性を検討する。具体的には、1 個の学習機械で報告されているような、マージン有りパーセプトロン則の優位性がアンサンブル学習においても見られるか、計算機実験で検討する。

2 モデル

本講演では教師・生徒の定式化を用いるものとする。教師・生徒の定式化とは、学習機械に対して正解を与える学習機械の存在を仮定するもので、その結合係数が既知であるとして議論を行う。以下の議論では教師役の学習機械を教師、学習を行う学習機械を生徒と呼ぶことにする。

本報告では単純パーセプトロンを学習機械として用いる。単純パーセプトロンとは N 入力、1 出力であり、入力と出力の間に結合荷重と呼ばれる、入力をどの程度出力に伝えるかを定めるパラメータを持つ。教師の結合荷重は

$$\mathbf{w}^* = (w_1^*, \dots, w_N^*) \quad (1)$$

であり、生徒の結合荷重は

$$\mathbf{w}^m = (w_1^m, \dots, w_N^m) \quad (2)$$

である。ここで m は学習回数を表す。今、入力ベクトルを $\mathbf{x} = (x_1, x_2, \dots, x_N)$ とすると生徒出力 $\text{sign}(u^m)$ は

$$\text{sign}(u^m) = \text{sign}\left(\sum_{i=1}^N w_i^m x_i^m\right) = \text{sign}(\mathbf{w}^m \cdot \mathbf{x}^m) \quad (3)$$

と計算できる。同様に教師出力 $\text{sign}(v^m)$ も

$$\text{sign}(v^m) = \text{sign}\left(\sum_{i=1}^N w_i^* x_i^m\right) = \text{sign}(\mathbf{w}^* \cdot \mathbf{x}^m) \quad (4)$$

と計算できる。教師の結合荷重は学習しないことに注する。

生徒の結合荷重の各要素 w_i は平均 0、分散 1 の正規分布から独立にサンプルされるものと仮定する。同様に、教師の結合荷重の各要素 w_i^* も平均 0、分散 1 の正規分布から独立にサンプルされるものと仮定する。また、入力の各要素 x_i は平均 0、分散 $1/N$ の正規分布から独立にサンプルされるものと仮定する。これらの仮定は、後に行う理論解析との整合性を保つために用いる。

次に学習をモデル化する。先に述べたように、教師・生徒の定式化を用いる。学習は生徒役の学習機械（以下生徒）によって行われる。教師と生徒は同じ入力を受け取り、それぞれの出力を計算し、規則に応じて結合荷重を変化させることにより学習が行われる。このときの規則を学習則という。学習方程式を以下に示す。

$$\mathbf{w}^{m+1} = \mathbf{w}^m + \eta f(u^m, v^m) \mathbf{x}^m \quad (5)$$

関数 $f(\cdot)$ は学習則によって決まり、ヘブ則では

$$f(u^m, v^m) = \text{sign}(v^m) \quad (6)$$

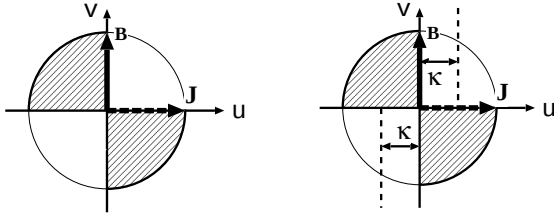
であり、パーセプトロン則では

$$f(u^m, v^m) = \Theta(-u^m \text{sign}(v^m)) \text{sign}(v^m) \quad (7)$$

である。ここで $\text{sign}(x)$ は符号関数であり、 $x > 0$ では 1、 $x \leq 0$ で -1 を返す関数である。また、 $\Theta(x)$ は $x > 0$ で 1、 $x \leq 0$ で 0 を返す階段関数である。 η は 1 回の学習での結合荷重の更新量を定める定数で、学習係数と呼ばれる。式中の m は学習回数である。教師の結合荷重を $\mathbf{w}^*$ とすると、教師出力は

Ensemble Learning through Perceptron Rule with a Margin

Kazuyuki HARA



(a) パーセプトロン則 (b) マージン有りパーセプトロン則
図1 パーセプトロン則とマージン有りパーセプトロン則の学習領域

$$\text{sign}(v^m) = \text{sign}\left(\sum_{i=1}^N w_i^* x_i^m\right) = \text{sign}(\mathbf{w}^* \cdot \mathbf{x}^m) \quad (8)$$

と計算できる。ここで教師の結合荷重 w_i^* は平均 0、分散 1 の正規分布から独立にサンプルされるものと仮定する。ヘブ則は $f(u^m, v^m) = \text{sign}(v^m)$ なので、常に教師の符号の方向に結合荷重を更新することがわかる。一方、式 (7) よりパーセプトロン則は教師と生徒の符号が異なる場合にのみ結合荷重を更新する。両者の違いを図 1(a) で説明する。図は $P(u, v)$ の分布を表す。横軸は生徒の入力ポテンシャル u 、縦軸は教師の入力ポテンシャル v である。横軸に生徒ベクトル、縦軸に教師ベクトルを配置することから、学習を開始する前の状態に対応する。このとき、出力は入力と結合ベクトルの内積の符号で与えられるので、結合荷重に直交する直線より上では出力は正、下では負となる。図の斜線が教師と生徒の出力の符号が異なる部分である。ヘブ則は入力に関わらず学習するので、円全体を学習する。一方、パーセプトロン則では教師と生徒の出力が異なる場合のみ学習するので、斜線部分を学習する。従って、ヘブ則の方が学習する領域が広いことが分かる。

次にマージンありパーセプトロン則を定式化する [1]。マージンありパーセプトロン則は、教師と生徒の符号が合っていない、生徒出力が小さい間は結合荷重を更新する。

$$f(u^m, v^m) = \Theta(\kappa - u^m \text{sign}(v^m)) \text{sign}(v^m) \quad (9)$$

である。ここで κ はマージンであり、 $\kappa = 0$ ではパーセプトロン則、 $\kappa \rightarrow \infty$ ではヘブ則と等価である。図 1(b) にマージン有りパーセプトロン則の学習領域を示す。図のように κ の分だけ学習領域が広がる事が分かる。従って、マージン有りパーセプトロン則は、パーセプトロン則では学習しない入力、つまり教師と生徒の符号が一致しているが、生徒出力が小さい場合も学習するため、パーセプトロン則より優れた性能を実現できる可能性が高い。

アンサンブル学習を定式化する [3]。アンサンブル学習とは多数の生徒を独立に学習し、その多数決により出力を決定する方法である。多数決は生徒数を K とすると、生徒出力の平均 (式 (10))、もしくは生徒の結合係

数の平均と等価の学習機械の出力 (式 (11)) によって実行できる。

$$\mathbf{w}^{m+1} = \mathbf{w}^m + \eta f\left(\frac{1}{K} \sum_{k=1}^K u_k^m, v\right) \mathbf{x}^m \quad (10)$$

$$\mathbf{w}^{m+1} = \mathbf{w}^m + \eta f\left(\left(\frac{1}{K} \sum_{k=1}^K \mathbf{J}_k\right) \cdot \mathbf{x}^m, v\right) \mathbf{x}^m \quad (11)$$

3 学習機械が 1 個の場合の学習則と収束特性との関係

はじめに学習の評価法について述べる。計算機実験においては、一般に学習の評価は以下の誤差関数が多く用いられている。

$$\epsilon = \frac{1}{M} \sum_{i=1}^M \Theta(-u_i v_i) \quad (12)$$

ここで M は誤差を算出するために用いる入力数であり、一般には $M > N$ である。式より誤差関数は 1 入力当たりの誤り率を表すことが分かる。しかし、この方法では誤差計算が N に依存し、計算量が多いため、本実験では以下の式を用いる [4]。

$$\epsilon = \frac{1}{\pi} \arccos(R), \quad (13)$$

$$R = \mathbf{B} \cdot \frac{1}{K} \sum_{k=1}^K \mathbf{J}_k \quad (14)$$

ここで R は教師の結合係数と生徒の結合係数との方向余弦であり、類似度を表す。この誤差関数はオンライン学習理論で用いられる式で有り、計算量が入力次元 N に因らない点で優れている。

次に 1 個の学習機械をヘブ則、パーセプトロン則、マージン有りパーセプトロン則で学習する場合の収束特性を計算機実験で求める。

ヘブ則は常に教師の符号の方向に結合荷重を更新するため効率が悪いが、収束が早いことが知られている [4]。一方、パーセプトロン則は生徒が誤ったときだけ学習するため効率はよいが、収束が遅いことが知られている [4]。図 2 に計算機実験の結果を示す。図では入力次元 $N = 1000$ 、学習係数 $\eta = 1$ 、 $\kappa = 1.0$ とした。図の横軸は学習回数を入力次元 N で正規化した学習時間であり、縦軸は誤差である。図より学習の初期ではマージン有りパーセプトロン則の収束が最も早い、 $t = 100$ ではヘブ則とマージン有りパーセプトロン則の誤差は同等であることが分かった。

一方、 κ を変化させたときの効果を次に示す。図 3 ではマージン $\kappa = 0, 0.5, 1.0, 5.0, 20.0$ そしてヘブ則の誤差曲線を示す。横軸は入力次元 N で学習回数を正規化した時間 t 、縦軸は誤差 (式 (14)) である。

図の最も上の曲線が $\kappa = 0$ 、つまりパーセプトロン則による学習結果を表す。 κ を大きくして行くに従って、

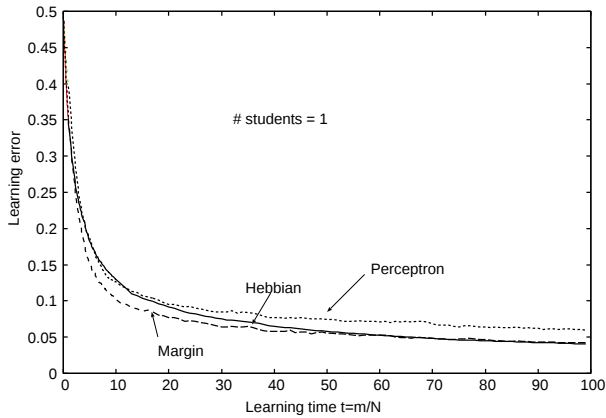


図2 ヘブ則、パーセプトロン則、マージン有りパーセプトロン則の誤差曲線

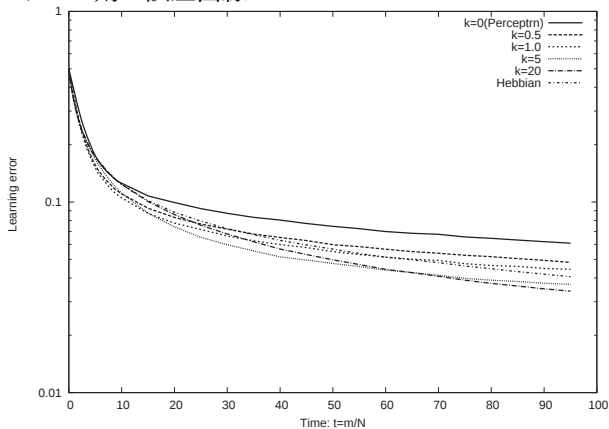


図3 $\kappa = 0, 0.5, 1, 5, 20$ としたときの誤差曲線

誤差は小さくなり、学習の初期（図の左側）では $\kappa = 5.0$ の誤差が最も小さくなった。しかし、学習が進むに従って $\kappa = 5.0$ の誤差は劣化し、 $t = 100$ （図の右端）では $\kappa = 20.0$ が最も誤差が小さくなった。従って、学習機械が1個の場合はマージン κ が大きい方が漸近的な誤差が小さくなることが分かった。

また、図3では明確に示されていないが、マージンが無限大のとき、マージン有りパーセプトロン則はヘブ則と等価であることから、 t が十分大きな極限ではヘブ則が最も誤差が小さくなることが示唆される [2]。

4 アンサンブル学習における学習則と収束特性の関係

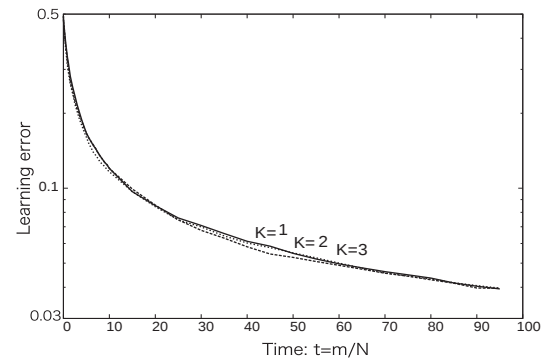
ここでは式 (11) を用いたアンサンブル学習の実験結果を示す。計算機実験の条件は $N = 1000$ 、学習係数 $\eta = 1$ 、マージン $\kappa = 1$ とした。横軸は入力次元 N で学習回数を正規化した時間 t 、縦軸は誤差（式 (14)）である。

図4(a) はヘブ則を用いた結果である。結果より、生徒の数に因らず誤差曲線は同じになった。ヘブ則は式 (6) より、教師の出力の符号の方向に生徒の結合荷重を更新を行う。従って生徒の多様性が影響を与えないため、多数決の効果が発揮されず、生徒の数が誤差に反映されなかったものと考えられる。

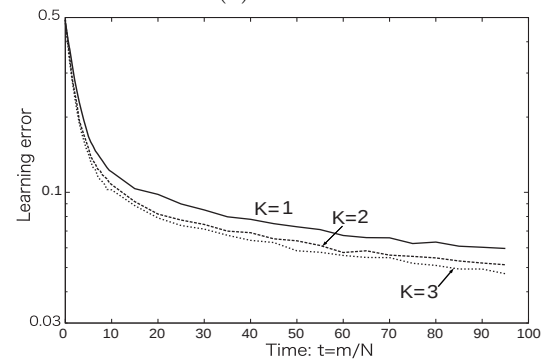
図4(b) はパーセプトロン則を用いた結果である。図中で上から生徒数が1, 2, 3に対する誤差曲線である。

このように、パーセプトロン則を用いた場合には生徒数が増えるに従って、誤差は小さくなることが分かった。パーセプトロン則では式 (7) より、生徒と教師の出力の違いによって結合荷重を更新するため、生徒の多様性が保存され、それによって多数決の効果が発揮されたと考えられる。

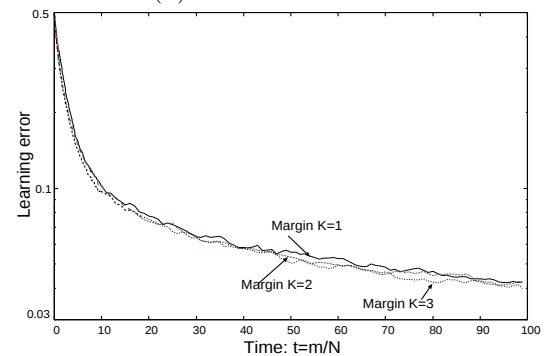
図4(c) はマージン有りパーセプトロン則を用いた結果である。結果より、ヘブ則と同様に、生徒の数に因らず誤差曲線は同じになった。マージン有りパーセプトロン則では式 (9) を用いて、各生徒の結合荷重を更新する。従って、生徒の多様性が学習に反映され、生徒数に因って誤差曲線が変わるものと考えられる。しかし、図4(c)からはその様な結果が得られなかった。



(a) ヘブ則



(b) パーセプトロン則



(c) マージン有りパーセプトロン則

図4 アンサンブル学習の誤差曲線。生徒数は1, 2, 3とした。

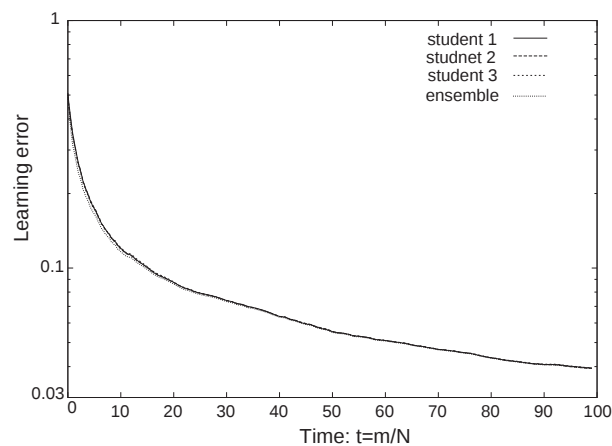
マージン有りパーセプトロン則を用いた場合、アンサンブル学習において、生徒数が異なった場合でも誤差曲

線が同じになった。そこで、マージン有りパーセプトロン則を用いた場合のアンサンブル学習の効果を確認した。具体的には、各生徒の平均誤差とアンサンブル学習の誤差を比較した。アンサンブル学習の効果がなければ、両者の誤差は同じになるはずである。

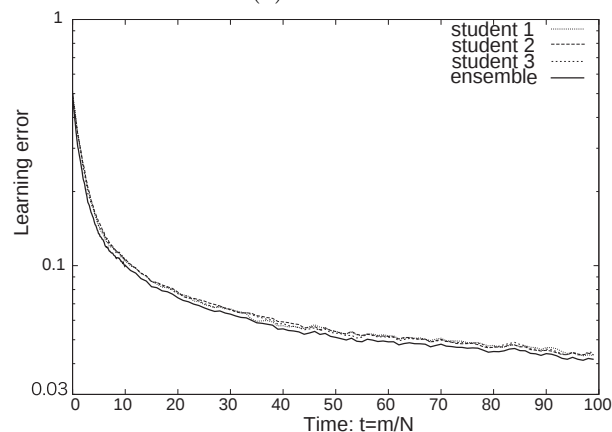
計算機実験の条件は $N = 1000$ 、学習係数 $\eta = 1$ 、マージン $\kappa = 1$ とした。横軸は入力次元 N で学習回数を正規化した時間 t 、縦軸は誤差 (式 (14)) である。

図 5(a) ではヘブ則を用いたアンサンブル学習における、各生徒の誤差とアンサンブル学習の誤差を合わせてプロットした。図に示すように、各生徒の誤差とアンサンブル学習の誤差は一致しており、アンサンブル学習のこうかがないことが分かる。

図 5(b) ではマージン有りパーセプトロン則を用いたアンサンブル学習における、各生徒の誤差とアンサンブル学習の誤差を合わせてプロットした。図中、最も下の曲線がアンサンブル学習の誤差曲線であり、その上で重なって見えるのが各生徒の誤差曲線である。このようにマージン有りパーセプトロン則を用いたアンサンブル学習では、各生徒の誤差より、アンサンブル学習の誤差の方が小さいことから、アンサンブル学習の効果があることが分かった。



(a) ヘブ則



(b) マージン有りパーセプトロン則

図 5 $\kappa = 1$ 、生徒数 3 のときの各生徒とアンサンブル学習の誤差曲線

5 まとめ

本講演では、マージン有りパーセプトロン則を用いたアンサンブル学習の効果を計算機実験によって明らかにした。1 個の学習機械による学習の場合、マージン有りパーセプトロン則は、最も誤差の収束が早いヘブ則と同等の性能の性能を有する。一方、マージン有りパーセプトロン則をアンサンブル学習に適用した場合、アンサンブル学習の効果は有るが、生徒数に応じて誤差が小さくなる、という現象は確認できなかった。マージン有りパーセプトロン則がパーセプトロン則とヘブ則の中間的な学習則であること、マージンが大きい場合にはマージン有りパーセプトロン則はヘブ則と同等の振る舞いをすることを考え合わせると、今回の実験で用いた $\kappa = 1$ が十分大き過ぎた、という可能性は否定できない。

以上のことを考慮すると、アンサンブル学習の効果が有り、かつ漸近的な誤差が最小となるような最適なマージンを求めることが重要である。しかし、計算機実験ではアンサンブル学習の効果を厳密に議論することは難しいため、マージン有りパーセプトロン則を用いたアンサンブル学習を統計力学的な手法を用いて理論的に解析することが今後の課題である。

参考文献

- [1] Hara, K., and Okada, M., "On-line learning through simple perceptron learning with a margin", *Neural Networks*, vol. 17, (2004) pp. 215-223 .
- [2] Hara, K., and Okada, M., "Ensemble learning of Linear Perceptrons: On-Line Learning theory", *Journal of Physical Society of Japan*, vol. 74, no. 11, (2005) pp. 2966-2972 .
- [3] Miyoshi, S., Hara, K., and Okada, M., "Analysis of ensemble learning using simple perceptrons based on online learning theory", *Physical Review E*, vol. 71, (2005) 036116.
- [4] Nishimori, H., *Spin Glass Theory and Statistical Mechanics of Information*, Oxford University Press, Oxford, UK (2001).