

NAIST-IS-MT1551098

修士論文

ノイズあり教師のパーセプトロン学習の統計力学的解析

本田 新

2017 年 3 月 6 日

奈良先端科学技術大学院大学
情報科学研究科 情報科学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士（工学）授与の要件として提出した修士論文である。

本田 新

審査委員：

池田 和司 教授 （主指導教員）

杉本 謙二 教授 （副指導教官）

佐々木 博昭 助教 （副指導教官）

ノイズあり教師のパーセプトロン学習の統計力学的解析*

本田 新

内容梗概

本研究では、単純パーセプトロンと呼ばれる線形二分機械のオンライン学習における汎化性能について議論する。オンライン学習における汎化性能は学習曲線と呼ばれ、その導出に統計力学の計算手法を用いる。ここでは入力次元や学習例題数が十分に大きいと仮定し、統計力学における熱力学的極限と中心極限定理から近似計算を行うことで、学習曲線のダイナミクスが導出される。

教師信号に観測ノイズが重畳する場合の三つのオンライン学習法における学習曲線を導出した。具体的には、一般の学習則について微分方程式を具体的に導出し、それらを数値的に解くことで学習曲線のダイナミクスを求めた。その結果、3 学習のうちパーセプトロン学習とアダptron学習においては従来の問題設定と異なり、学習曲線が非単調にふるまい、ヘブ学習は単調にふるまう結果になった。非単調性を解析するため、線形システムの固有値および固有ベクトルを用いることでその現象について示した。具体的には連立微分方程式を平衡点周りで一次近似することにより、ダイナミクスを連続時間の線形システムで表し、線形システムの状態行列の固有値および固有ベクトルから、非単調性の原因が説明できることを示した。ヘブ学習についてはダイナミクスの漸近解を導出することで単調性を示した。

最後に線形システムを用いることでパーセプトロン学習の収束性の向上を試みたが、収束性の改善はできなかった。

キーワード

単純パーセプトロン, オンライン学習, 非単調学習曲線, 統計力学

*奈良先端科学技術大学院大学 情報科学研究科 情報科学専攻 修士論文, NAIST-IS-MT1551098, 2017 年 3 月 6 日.

Statistical mechanics of Perceptron learning with noisy teacher*

Arata Honda

Abstract

We discuss the generalization performance of linear binary machines called simple perceptrons based on on-line learning in this paper. The generalization performance based on on-line learning is called the learning curve. The learning curve is derived by statistical mechanics methods. The assumption that the number of inputs and examples are sufficiently large derives the simultaneous differential equation about the learning curve.

In this paper, we derive the learning curves of three algorithms based on on-line learning method with noisy teacher, that is, we drive the differential equation about learning rules and numerically calculate the learning curve. We found that the learning curves of the Perceptron learning and the Adatron learning are non-monotonic which is not traditional result while that of the Hebb learning is monotonic. We analyzed why the learning curve behaves non-monotonically using the eigenvalues of the system matrix and the corresponding eigenvectors, that is, we analyzed the behavior around the convergence point. The monotonicity of the Hebb learning is analyzed by asymptotic solution of learning dynamics. Finally, we tried to improve the convergence of learning using the system matrix. However it doesn't improve.

Keywords:

Simple perceptron, on-line learning, non-monotonic learning curve, statistical mechanics

*Master's Thesis, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1551098, 6, 2017.

目次

図目次	v
第 1 章 序論	1
第 2 章 問題定式化	3
第 3 章 統計力学的解析	5
3.1 巨視的変数 u, v の導入	5
3.2 学習のダイナミクスの導出	6
3.3 アンサンブル平均の解析的導出	7
3.3.1 パーセプトロン学習におけるアンサンブル平均の解析的導出	9
3.3.2 先行研究 [13] との比較	10
3.3.3 ヘブ学習におけるアンサンブル平均の解析的導出	11
3.3.4 アダトロン学習におけるアンサンブル平均の解析的導出 . .	12
3.4 計算機実験	12
第 4 章 漸近解析	14
4.1 パーセプトロン学習における非単調性の漸近解析 [13]	14
4.2 アダトロン学習における非単調性の漸近解析	16
4.3 ヘブ学習における単調性の漸近解析	17
第 5 章 学習係数の最適化	19
5.1 数値実験による収束性の改善	19
5.2 線形システム行列を用いた学習係数の調節式	19
5.2.1 遷移行列の導出	20
5.2.2 理論値と実験値の比較	21
第 6 章 結論	23
謝辞	24

参考文献	25
付録. 1 各ベクトルのノルム	27
付録. 2 $\langle f^2 \rangle$ の計算	27

図目次

2.1	教師と生徒.	4
2.2	教師 B と生徒 J	4
3.1	$\langle fv \rangle$ の解析解と [13] の比較 ($\sigma_B^2 = 1.0$).	10
3.2	$\langle fu \rangle$ の解析解と [13] の比較 ($\sigma_B^2 = 1.0$).	10
3.3	$\langle f^2 \rangle$ の解析解と [13] の比較 ($\sigma_B^2 = 1.0$).	11
3.4	パーセプトロン学習における R のダイナミクス ($\sigma_B^2 = 1.0$).	13
3.5	アダプトロン学習における R のダイナミクス ($\sigma_B^2 = 1.0$).	13
3.6	ヘブ学習における R のダイナミクス ($\sigma_B^2 = 1.0$).	13
4.1	(e, d) の描跡 ($\sigma_B^2 = 1.0$).	15
4.2	(e, d) と固有ベクトル ($\sigma_B^2 = 1.0$).	15
4.3	(e, d) の描跡 ($\sigma_B^2 = 1.0$).	17
4.4	(e, d) と固有ベクトル ($\sigma_B^2 = 1.0$).	17
4.5	ヘブ学習における e, d 平面描画 $\eta = 0.4, \sigma_B^2 = 1.0$	18
5.1	パーセプトロン学習における R のダイナミクスと学習係数の数値 的調節結果の比較 ($\sigma_B^2 = 1.0$).	19
5.2	アダプトロン学習における R のダイナミクスと学習係数の数値的調 節結果の比較 ($\sigma_B^2 = 1.0$).	19
5.3	学習率 η と学習時間 t の理論式と経験値の比較.	22
5.4	e のダイナミクス.	22
5.5	線形システム行列 e のダイナミクス.	22

第 1 章 序論

情報統計力学は統計力学の考え方をを用いた情報科学の問題を解析する手法である。統計力学とは巨視的な性質を微視的な物理法則から説明する学問であり、情報統計力学の枠組みから、画像修復 [1] や、誤り訂正符号 [2]、ニューラルネットワーク [3] など、広い範囲での情報処理の問題が解析されている。その解析例の一つに単純パーセプトロンの学習曲線が挙げられる [4, 6, 7, 8, 9, 10, 11, 12, 13]。

単純パーセプトロンとは神経素子 (ニューロン) を組み合わせた、ニューラルネットワークの最も単純なモデルであり、その構造は単一のニューロン自身に学習機能を持たせた非常に簡単な構造となっている。具体的には複数の入力とそれぞれにかかる結合荷重、そして一つの出力から構成され、入力ベクトルと結合ベクトルの積和を内部状態とし、出力関数にて出力決定を行う。単純パーセプトロンの代表的な学習方法としてパーセプトロン学習が挙げられる。パーセプトロン学習は、学習器の出力が望ましい出力と異なっている場合に、自らの結合ベクトルを入力分だけ変化させる学習アルゴリズムである [4]。このアルゴリズムの性質として、解が線形分離可能であれば必ず収束するといった収束定理 [5] や、解が線形分離不可能の場合は収束しないが、結合荷重の大きさは有界であるといった有界定理が一般的によく知られている。しかしながらこれらの性質は、学習例題の集合が与えられたとき、すなわちバッチ学習の性質である。バッチ学習とは対照的に逐次的に学習例題を与えるオンライン学習があり、学習例題を貯蔵するためのメモリの節約や、教師信号を出力する構造が変化しても追従することができる。オンライン学習では、与えられる学習例題数が増えるにつれて汎化性能が上昇する。これは学習曲線と呼ばれ、導出には情報統計力学を用いる。

単純パーセプトロンの学習曲線の導出は Biehl, Riegler により情報統計力学の枠組みで導入され [6]、Inoue, Nishimori らによって学習不可能な場合におけるアダルトン学習の場合が解析されている [7]。また Hara, Okada らによってマージンを用いたパーセプトロン学習の場合 [8]、Miyoshi らによる、教師信号や出力に観測ノイズが含まれる場合や、アンサンブル学習に拡張した場合の学習曲線についても導出が行われてきた [9, 10, 11, 12]。その中でも Hanzawa, Ikeda らにより、教師信号にのみ観測ノイズが含まれる問題設定の下、導出された学習曲線においては、本来単

調に増加する学習曲線が非単調にふるまう [13]. 具体的には, 教師と生徒の正規化重みベクトルの内積が最初は増加するが, 途中から減少に転じて定数に収束する現象が生じる. この現象のメカニズムは学習曲線のダイナミクスを線形システムに近似し, 収束点周りの挙動を調べることで理論的な説明が行われている [13].

しかし, この研究では学習曲線のダイナミクスに用いられるアンサンブル平均の導出計算を実験的に行われていることから, 学習曲線の導出が決定的ではない. また, オンライン学習の一つであるパーセプトロン学習のみの学習曲線の解析を行っているものであり, 一般的によく知られている 3 学習のうち, ヘブ学習やアダルトン学習 [3] については解析が行われていない. そこで本研究では, 学習機械が非線形パーセプトロンであり, 教師信号に雑音がある場合のオンライン学習について, 統計力学的手法を用いて理論的に解析を行うことを目的とする. 具体的には, 先行研究で導出ができなかったアンサンブル平均の解析的導出を行い, 残りの 2 学習法についての学習曲線も導出する. また, 学習曲線のダイナミクスの漸近的なふるまいを調べることで, 学習曲線の単調性および非単調性について議論を行う.

本論文の第 2 章からの構成は次の通りである. 第 2 章で問題の定式化を行い, 第 3 章にて情報統計力学による学習曲線の導出を行う. 第 4 章で学習曲線のダイナミクスの漸近的なふるまいを解析する. 第 5 章で学習の収束性の改善を実験的にを行い, 学習の収束性を向上させる理論式を導出する. そして第 6 章にて本論文をまとめる.

第 2 章 問題定式化

学習とは未知の入出力関係を推定することである．そのため学習対象である入出力関係を問題設定者側が定式化することにより，学習機械や学習アルゴリズムの汎化能力について議論することが可能となる．ここでは，オンライン学習における単純パーセプトロンの教師生徒の定式化を行うため，最も簡単なモデルを取り上げることにする．構造として，入力ベクトル (入力) と結合ベクトル (結合荷重) の内積の符号を出力とする．図 2.1 に示すように教師および生徒の二つの単純パーセプトロンを考え，それぞれの結合荷重を $\mathbf{B}$, $\mathbf{J}^m$ とする．ここで m は時刻ステップである．なお本稿では簡単のため，教師の結合荷重，生徒の結合荷重のことをそれぞれ生徒と教師と呼ぶ．教師 $\mathbf{B} = (B_1, \dots, B_N)$ ，生徒 $\mathbf{J}^m = (J_1^m, \dots, J_N^m)$ は N 次元のベクトルであり，教師の各要素 B_i および生徒の初期の各要素 J_i^0 は平均 0，分散 1 のガウス分布に従いそれぞれ独立に生成され，教師は不変である．

時刻 m における入力 $\mathbf{x}^m = (x_1^m, \dots, x_N^m)$ も同様に N 次元のベクトルであり，入力の各要素 x_i は平均 0，分散 $1/N$ のガウス分布に従いそれぞれ独立に生成される．すなわち，

$$\langle B_i \rangle = 0, \quad \langle (B_i)^2 \rangle = 1, \quad (2.01)$$

$$\langle J_i^0 \rangle = 0, \quad \langle (J_i^0)^2 \rangle = 1, \quad (2.02)$$

$$\langle x_i \rangle = 0, \quad \langle (x_i)^2 \rangle = \frac{1}{N}, \quad (2.03)$$

ここで $\langle \cdot \rangle$ は $\cdot$ の平均である．

教師の出力 y^m は，

$$y^m = \text{sgn}(v_m), \quad v_m = \mathbf{B} \cdot \mathbf{x}^m + n_B^m, \quad (2.04)$$

で作成される．ここで $\text{sgn}(\cdot)$ は符号化数であり，観測ノイズ n_B^m は平均 0，分散 σ_B^2 のガウス分布に従う．この観測ノイズは input noise[12] と呼ばれており，入力 $\mathbf{x}$ や教師 $\mathbf{B}$ のノイズであると解釈することができる．

生徒 $\mathbf{J}$ は入力 $\mathbf{x}$ とそれに対する教師 $\mathbf{B}$ の出力 y^m をもとに “学習” をし，結合荷重を

$$\mathbf{J}^{m+1} = \mathbf{J}^m + f^m \mathbf{x}^m. \quad (2.05)$$

のように更新する．ここで， f^m はヘブ学習，パーセプトロン学習，アダptron学習 [3] においてそれぞれ

ヘブ学習

$$f^m = \eta y^m. \quad (2.06)$$

パーセプトロン学習

$$f^m = \eta y^m \Theta(-y^m \mathbf{J}^m \cdot \mathbf{x}^m). \quad (2.07)$$

アダptron学習

$$f^m = -\eta \mathbf{J}^m \cdot \mathbf{x}^m \Theta(-y^m \mathbf{J}^m \cdot \mathbf{x}^m). \quad (2.08)$$

である．ここで $\eta > 0$ は学習係数であり， $\Theta(\cdot)$ はステップ関数

$$\Theta(t) = \begin{cases} 1 & (t \geq 0), \\ 0 & (t < 0). \end{cases} \quad (2.09)$$

である．

図 2.2 に示すように，生徒 $\mathbf{J}^m$ が学習を経て，教師 $\mathbf{B}$ にどれだけ近づいたかの評価には方向余弦

$$R^m = \frac{\mathbf{B} \cdot \mathbf{J}^m}{\|\mathbf{B}\| \|\mathbf{J}^m\|} = \cos \theta_R \quad (2.010)$$

を用いる．すなわち， R^m の上限は 1 で大きいほどよい．

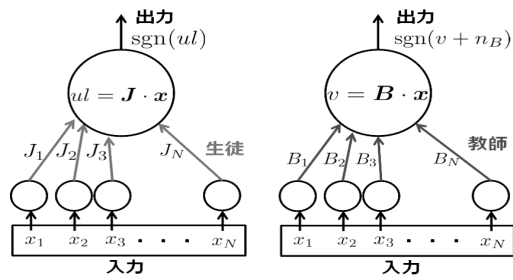


図 2.1 教師と生徒.

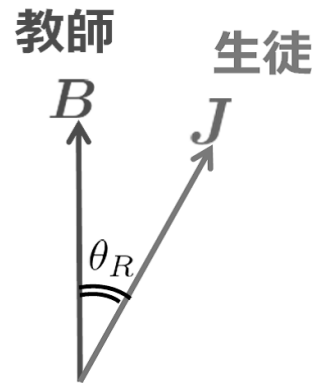


図 2.2 教師 $\mathbf{B}$ と生徒 $\mathbf{J}$.

第 3 章 統計力学的解析

この章では、オンライン学習に統計力学的解析を導入する。統計力学の重要な概念の一つとして、オーダーパラメータと呼ばれるマクロパラメータが存在し、系の構成要素をミクロとして、統計操作を加えることで導出される。また、オーダーパラメータを記述する方程式はオーダーパラメータのみに閉じており、ミクロな量に依存しないことが利点である。“学習”の統計力学でのミクロな量は、入力ベクトルと結合ベクトルの各要素であり、教師と生徒の学習機械の結合荷重の方向余弦 R がオーダーパラメータの一例である [14]。

3.1 巨視的変数 u, v の導入

統計力学的議論のため、入力次元 $N \rightarrow \infty$ の極限 (熱力学的極限) を考える。各ベクトルの各要素は独立なので、大数の法則から各ベクトルの大きさは以下のように計算される (詳細は付録. 1 節を参照)。

$$\|\mathbf{B}\| = \sqrt{N}, \|\mathbf{J}^0\| = \sqrt{N}, \|\mathbf{x}\| = 1, \quad (3.11)$$

生徒のノルム $\|\mathbf{J}\|$ は時刻ステップの経過とともに学習アルゴリズムを経て変化するので、初期値 $\|\mathbf{J}^0\| = \sqrt{N}$ からの変化率を l としてこれを生徒の長さと呼ぶことにする。すなわち、 $\|\mathbf{J}^m\| = l^m \sqrt{N}$ である。

入力と結合荷重の積和を格納する内部状態はそれぞれ、

$$v^m = \mathbf{B} \cdot \mathbf{x}^m, \quad u^m l^m = \mathbf{J}^m \cdot \mathbf{x}^m. \quad (3.12)$$

と計算することができ、熱力学的極限において大数の法則から、 u, v は平均 0, 分散 1, 共分散 R^m の二次元ガウス分布で近似 [3] できる (導出詳細は [15] を参照)。問題の定式化において入力ベクトルと結合荷重ベクトルをガウス分布に設定した理由は式 (2.010) における一致度 R が熱力学的極限において発散しないよう正規化しているのである。

3.2 学習のダイナミクスの導出

ここで、学習曲線を導出するため、オーダーパラメータである l, R のダイナミクスを導出する。ここでは、そのダイナミクスが記述する連立微分方程式を熱力学的極限における自己平均性に基づき決定論的な形で導く。

まず l の微分方程式を導出する。式 (2.05) の両辺を二乗する。

$$\mathbf{J}^{m+1} \cdot \mathbf{J}^{m+1} = \mathbf{J}^m \cdot \mathbf{J}^m + (f^m)^2 \mathbf{x}^m \cdot \mathbf{x}^m + 2f^m \mathbf{J}^m \cdot \mathbf{x}^m \quad (3.21)$$

$$N(l^{m+1})^2 = N(l^m)^2 + (f^m)^2 + 2f^m l^m u^m \quad (3.22)$$

この段階では f^m, u^m は確率変数であるので、式 (3.22) は確率的な漸化式であり、 l^m に関する決定論的な方程式になっていない。ここで $\|\mathbf{J}\| \sim O(\sqrt{N})$, $\|\mathbf{x}\| = 1$ という問題定式がダイナミクスの導出に関わってくる。式 (3.22) について、左辺と右辺の第一項は $O(N)$ であり、それ以外の項は $O(1)$ であることが簡単な計算より確認できる。つまり、式 (3.22) から l^m が $O(1)$ 変化するためには、入力 $\mathbf{x}$ は $O(N)$ 個入力されなければならない。そこで、以後は時刻ステップ m を次元 N で正規化した値として、連続変数の学習時間 t を $t = m/N$ とおき、学習過程を表す添字として用いる。時刻ステップが m から $m + Ndt$ になった場合を考えると、学習時間 t は $t + dt$ となる。この時、 dt は微小量であるが N に対しては $O(1)$ であるため、 t から $t + dt$ までの時間変化をまとめると、式 (3.22) が、以下のように縦に Ndt 個並んだ式に対応する [14]。

$$N(l^{m+1})^2 = N(l^m)^2 + (f^m)^2 + 2f^m l^m u^m \quad (3.23)$$

$$N(l^{m+2})^2 = N(l^{m+1})^2 + (f^{m+1})^2 + 2f^{m+1} l^{m+1} u^{m+1}$$

$$\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots$$

$$N(l^{m+Ndt})^2 = N(l^{m+Ndt-1})^2 + (f^{m+Ndt-1})^2 + 2f^{m+Ndt-1} l^{m+Ndt-1} u^{m+Ndt-1} \quad (3.24)$$

式 (3.23), 式 (3.24) を全て足す。このとき、 l が dl だけ変化するために更新が Ndt 回 ($O(N)$ で非常に大きい) 実行されるので、右辺における確率変数の影響をそれに関する平均で置き換えることができることに注意すると、

$$N(l + dl)^2 = Nl^2 + Ndt (\langle f^2 \rangle + 2l \langle fu \rangle) \quad (3.25)$$

となる. この性質は自己平均性と呼ばれる [3]. よって, l のダイナミクスは,

$$\frac{dl}{dt} = \langle fu \rangle + \frac{\langle f^2 \rangle}{2l}. \quad (3.26)$$

次に R のダイナミクスを導出する. 式 (2.05) の両辺に $\mathbf{B}$ をかけると,

$$\mathbf{B} \cdot \mathbf{J}^{m+1} = \mathbf{B} \cdot \mathbf{J}^m + f^m \mathbf{B} \cdot \mathbf{x}^m \quad (3.27)$$

$$Nl^{m+1}R^{m+1} = Nl^mR^m + f^mv^m \quad (3.28)$$

l に関する議論と同様に, 微小時間 dt の間に入力が Ndt 個入り, これによって R が dR , l が dl だけ変化したとすると,

$$N(l + dl)(R + dR) = NlR + Ndt \langle fv \rangle \quad (3.29)$$

となる. よって, R のダイナミクスは,

$$\frac{dR}{dt} = \frac{\langle fv \rangle - \langle fu \rangle R}{l} - \frac{R}{2l^2} \langle f^2 \rangle. \quad (3.210)$$

式 (3.26), 式 (3.210) が l と R のダイナミクスを記述する連立微分方程式である. 具体的な学習則 f に関して式 (3.26), 式 (3.210) に現れるアンサンブル平均 $\langle fu \rangle$, $\langle fv \rangle$, $\langle f^2 \rangle$ を R と l の関数として求めることができれば, 式 (3.26), 式 (3.210) は冒頭にのべたように, オーダーパラメータである R と l に閉じた方程式となる. これらの式を解くことにより, 系の巨視的な挙動を決定論的に議論することができる.

3.3 アンサンブル平均の解析的導出

Hanzawa, Ikeda らはパーセプトロン学習におけるアンサンブル平均を数値的に計算していた [13]. 導出された各アンサンブル平均については非常に精度の良い近似式であるが, 実験的に導出しているため真値かどうか定かではない. そのため, 解析的に求めて比較する必要がある. ここでは, パーセプトロン学習におけるアンサンブル平均の解析的導出を行い, [13] と比較を行う. また, 一般的によく知られている 3 学習のヘブ学習, アダトロン学習 [3] についてのアンサンブル平均も解析的に導出を行う.

なお、以下では誤差関数 $\text{erf}(x)$ および相補誤差関数 $\text{erfc}(x)$ は

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt, \quad (3.31)$$

$$\text{erfc}(x) = 1 - \text{erf}(x), \quad (3.32)$$

である。

例えばアンサンブル平均 $\langle f v \rangle$ は入力 $\mathbf{x}$ に関する平均として以下のように計算される。

$$\langle f v \rangle = \int d\mathbf{x} P(\mathbf{x}) f v \quad (3.33)$$

$$= \int du \int dv \int dn_B P(u, v) P(n_B) f v \quad (3.34)$$

ここで、式 (3.33) 左辺の $\langle \cdot \rangle$ は多くの入力 $\mathbf{x}$ に関する $f v$ のアンサンブル平均を表している。これを $\mathbf{x}$ の確率密度 $P(\mathbf{x})$ を用いて表現したものが右辺にあたるが、これは N 重積分である。いま $N \rightarrow \infty$ を考えているので、このままではこの積分は困難である。しかし、 $\mathbf{x}$ で決まる v と u が平均 0、分散 1、共分散 R の二次元ガウス分布に従う確率変数に注目すると、この積分は積分変数を v と u に置き換えることができ、さらに、式 (2.06)–式 (2.08) における学習則 f に含まれる平均 0、分散 σ_B^2 のガウス分布にしたがう観測ノイズ n_B も考慮されるため、式 (3.34) は三変数の積分式で書くことができる。ここで、

$$P(u, v) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{(u, v)\Sigma^{-1}(u, v)^T}{2}\right) \quad (3.35)$$

$$\Sigma = \begin{pmatrix} 1 & R \\ R & 1 \end{pmatrix} \quad (3.36)$$

$$P(n_B) = \frac{1}{\sqrt{2\pi\sigma_B^2}} \exp\left(-\frac{n_B^2}{2\sigma_B^2}\right) \quad (3.37)$$

である。以下の小々節からは実際に三学習におけるアンサンブル平均を計算する。

3.3.1 パーセプトロン学習におけるアンサンブル平均の解析的導出

パーセプトロン学習の場合, アンサンブル平均のうち $\langle fv \rangle$ および $\langle fu \rangle$ は

$$\langle fv \rangle = \langle \eta \Theta(-u(v + n_B)) \text{sgn}(v + n_B) v \rangle \quad (3.38)$$

$$\begin{aligned} &= \frac{\eta}{2\pi\sqrt{1-R^2}} \int_{-n_B}^{\infty} dv \int_{-\infty}^0 du \int_{-\infty}^{\infty} \frac{dn_B}{\sqrt{2\pi\sigma_B^2}} v \exp\left(-\frac{u^2 + v^2 - 2Ruv}{2(1-R^2)}\right) \exp\left(-\frac{n_B^2}{2\sigma_B^2}\right) \\ &- \frac{\eta}{2\pi\sqrt{1-R^2}} \int_{-\infty}^{-n_B} dv \int_0^{\infty} du \int_{-\infty}^{\infty} \frac{dn_B}{\sqrt{2\pi\sigma_B^2}} v \exp\left(-\frac{u^2 + v^2 - 2Ruv}{2(1-R^2)}\right) \exp\left(-\frac{n_B^2}{2\sigma_B^2}\right) \end{aligned} \quad (3.39)$$

$$\begin{aligned} &= \frac{\eta}{2\pi\sigma_B} \int_{-\infty}^{\infty} dn_B \exp\left(-\frac{n_B^2(1+\sigma_B^2)}{2\sigma_B^2}\right) \text{erfc}\left(-\frac{Rn_B}{\sqrt{2(1-R^2)}}\right) \\ &- \frac{\eta R}{2\pi\sigma_B} \int_{-\infty}^{\infty} dn_B \exp\left(-\frac{n_B^2}{2\sigma_B^2}\right) \text{erfc}\left(-\frac{n_B}{\sqrt{2(1-R^2)}}\right) \end{aligned} \quad (3.310)$$

$$= -\frac{\eta}{\sqrt{2\pi}} R + \frac{\eta}{\sqrt{2\pi(1+\sigma_B^2)}}, \quad (3.311)$$

$$\langle fu \rangle = \frac{\eta}{\sqrt{2\pi(1+\sigma_B^2)}} R - \frac{\eta}{\sqrt{2\pi}}, \quad (3.312)$$

と計算され, R の一次関数となる. $\langle f^2 \rangle$ は

$$\langle f^2 \rangle = \frac{\eta^2}{2} + \frac{\eta^2}{2\pi\sigma_B} \int_{-\infty}^0 du \int_{-\infty}^{\infty} dn_B \exp\left(-\frac{u^2}{2}\right) \text{erf}\left(\frac{n_B + Ru}{\sqrt{2(1-R^2)}}\right) \exp\left(-\frac{n_B^2}{2\sigma_B^2}\right) \quad (3.313)$$

と計算されるが, この積分はそのままでは計算できない. そこで, 式 (3.313) に含まれる誤差関数 $\text{erf}(\cdot)$ を u の関数として原点周りでテイラー展開し, 9 次多項式で近似すると

$$\begin{aligned} \langle f^2 \rangle &\approx \frac{\eta^2}{2} - \frac{\eta^2 R}{\pi\sqrt{1-R^2+\sigma_B^2}} \\ &+ \frac{\eta^2 R^3}{3\pi(1-R^2+\sigma_B^2)\sqrt{1-R^2+\sigma_B^2}} \\ &- \frac{\eta^2 R^5}{5\pi(1-R^2+\sigma_B^2)^2\sqrt{1-R^2+\sigma_B^2}} \end{aligned}$$

$$\begin{aligned}
& + \frac{\eta^2 R^7}{7\pi(1 - R^2 + \sigma_B^2)^3 \sqrt{1 - R^2 + \sigma_B^2}} \\
& - \frac{\eta^2 R^9}{9\pi(1 - R^2 + \sigma_B^2)^4 \sqrt{1 - R^2 + \sigma_B^2}}
\end{aligned} \tag{3.314}$$

$$= \frac{\eta^2}{2} + \sum_{n=1}^5 \frac{(-1)^n \eta^2 R^{2n-1}}{(2n-1)\pi(1 - R^2 + \sigma_B^2)^{\frac{2n-1}{2}}} \tag{3.315}$$

と計算される (詳細は付録. 2 節を参照).

3.3.2 先行研究 [13] との比較

ここで, 解析的に導出したアンサンブル平均と, Ikeda ら [13] の実験的に求めたアンサンブル平均の近似解について比較をおこなう. なお, Ikeda らのアンサンブル平均の導出は $\sigma_B^2 = 1.0$ として変数を一つ減らし, 数値的に求めており, 以下の結果になっている.

$$\langle fv \rangle \approx -0.4\eta R + 0.28\eta \tag{3.316}$$

$$\langle fu \rangle \approx 0.28\eta R - 0.4\eta \tag{3.317}$$

$$\langle f^2 \rangle \approx -0.24\eta^2 R + 0.5\eta^2 \tag{3.318}$$

上記で求めた解析解に $\sigma_B^2 = 1.0$ を代入して比較した (図 3.1, 図 3.2, 図 3.3). その結果, [13] のアンサンブル平均が正確であるということが分かる.

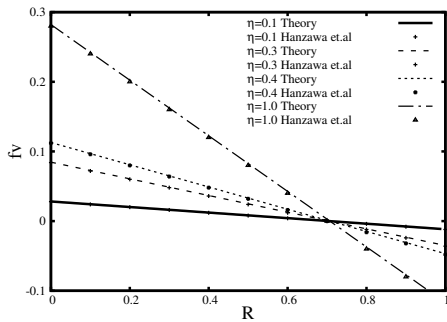


図 3.1 $\langle fv \rangle$ の解析解と [13] の比較 ($\sigma_B^2 = 1.0$).

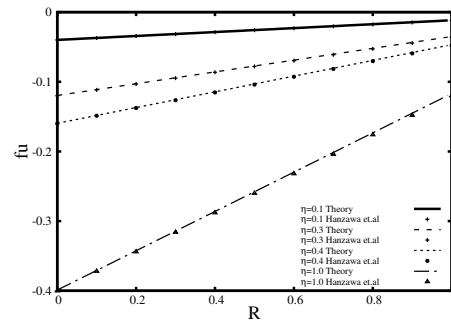


図 3.2 $\langle fu \rangle$ の解析解と [13] の比較 ($\sigma_B^2 = 1.0$).

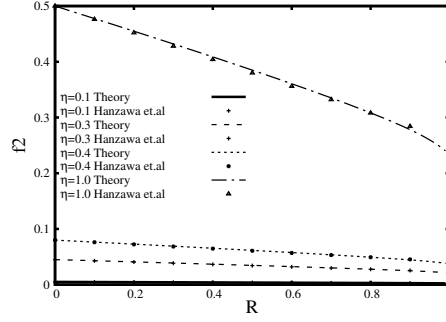


図 3.3 $\langle f^2 \rangle$ の解析解と [13] の比較 ($\sigma_B^2 = 1.0$).

3.3.3 ヘブ学習におけるアンサンブル平均の解析的導出

ヘブ学習則の場合のアンサンブル平均は以下のように解析的に計算できる.

$$\langle fv \rangle = \langle \eta \text{sgn}(v + n_B) v \rangle \quad (3.319)$$

$$\begin{aligned} &= \eta \int_{-n_B}^{\infty} dv \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dn_B P(u, v) P(n_B) v \\ &- \eta \int_{-\infty}^{-n_B} dv \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dn_B P(u, v) P(n_B) v \end{aligned} \quad (3.320)$$

$$= \frac{2\eta}{\sqrt{2\pi(1 + \sigma_B^2)}} \quad (3.321)$$

$$\langle fu \rangle = \langle \eta \text{sgn}(v + n_B) u \rangle \quad (3.322)$$

$$\begin{aligned} &= \eta \int_{-n_B}^{\infty} dv \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dn_B P(u, v) P(n_B) u \\ &- \eta \int_{-\infty}^{-n_B} dv \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dn_B P(u, v) P(n_B) u \end{aligned} \quad (3.323)$$

$$= \frac{2\eta}{\sqrt{2\pi(1 + \sigma_B^2)}} R \quad (3.324)$$

$$\langle f^2 \rangle = \eta^2 \quad (3.325)$$

3.3.4 アダトロン学習におけるアンサンブル平均の解析的導出

アダトロン学習の場合のアンサンブル平均は以下のように解析的に計算できる．

$$\langle fv \rangle = \langle -\eta uv \Theta(-u(v + n_B)) \rangle \quad (3.326)$$

$$\begin{aligned} &= -\eta \int_{-n_B}^{\infty} dv \int_{-\infty}^0 du \int_{-\infty}^{\infty} dn_B P(u, v) P(n_B) uv \\ &\quad - \eta \int_{-\infty}^{-n_B} dv \int_0^{\infty} du \int_{-\infty}^{\infty} dn_B P(u, v) P(n_B) uv \end{aligned} \quad (3.327)$$

$$= \frac{\eta}{\pi} \frac{(1 - R^2)^2}{\sqrt{1 - R^2 + \sigma_B^2}} + \frac{\eta R^2}{\pi} \left(\frac{\sigma_B^2}{1 + \sigma_B^2} \right) \frac{(1 - R^2)}{\sqrt{1 - R^2 + \sigma_B^2}} + R \langle fu \rangle \quad (3.328)$$

$$\langle fu \rangle = \langle -\eta u^2 \Theta(-u(v + n_B)) \rangle \quad (3.329)$$

$$\begin{aligned} &= -\eta \int_{-n_B}^{\infty} dv \int_{-\infty}^0 du \int_{-\infty}^{\infty} dn_B P(u, v) P(n_B) u^2 \\ &\quad - \eta \int_{-\infty}^{-n_B} dv \int_0^{\infty} du \int_{-\infty}^{\infty} dn_B P(u, v) P(n_B) u^2 \end{aligned} \quad (3.330)$$

$$\begin{aligned} &= -\frac{\eta}{2} - \sum_{n=1}^5 \frac{(-1)^n \eta R^{2n-1}}{(2n-1)\pi(1 - R^2 + \sigma_B^2)^{\frac{2n-1}{2}}} + \frac{\eta}{\pi} \frac{R(1 - R^2)}{\sqrt{1 - R^2 + \sigma_B^2}} \\ &\quad + \frac{\eta R^3}{\pi} \left(\frac{\sigma_B^2}{1 + \sigma_B^2} \right) \frac{1}{\sqrt{1 - R^2 + \sigma_B^2}} \end{aligned} \quad (3.331)$$

$$\langle f^2 \rangle = \langle \eta^2 u^2 \Theta(-u(v + n_B)) \rangle \quad (3.332)$$

$$= \langle -\eta \cdot -\eta u^2 \Theta(-u(v + n_B)) \rangle \quad (3.333)$$

$$= -\eta \langle fu \rangle \quad (3.334)$$

3.4 計算機実験

情報統計力学による理論解析の正当性を確認するため、三つの学習法について $N = 10^4$ として R の実験値を求め、各アンサンブル平均の解析解を式 (3.26), 式 (3.210) に代入し、ルンゲクッタ法によって求めた理論値と比較した．その結果を図 3.4, 図 3.5, 図 3.6 に示す．理論値と実験値はよく一致し、パーセプトロン学習およびアダトロン学習では学習曲線が非単調にふるまう．具体的には、各 η で R は最大となってから減少していく．この現象は他の問題設定 [3], [7], [8], [9], [10] では見

られない点が特徴的である。この学習曲線の非単調性の影響により、学習時間 t の経過とともに R を最大化する学習係数 η が変化することがわかる。具体的には、学習係数 η が小さくなるほど、非単調性は大きくなる。つまりこの現象を解明することで、学習係数の制御による性能向上が期待できる。そのため次の章では、この学習曲線の非単調性の原因を理論的に示す。

また、二つの学習法とは異なり学習曲線が単調にふるまうヘブ学習についても、議論を行う。

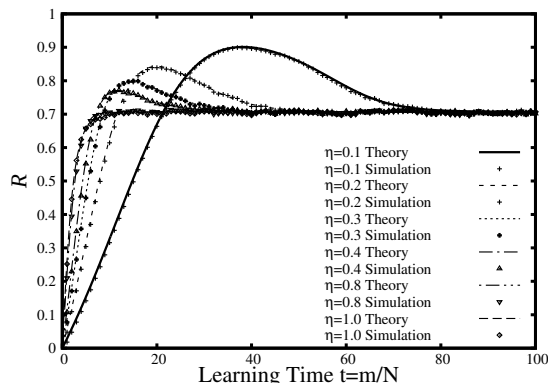


図 3.4 パーセプトロン学習における R のダイナミクス ($\sigma_B^2 = 1.0$).

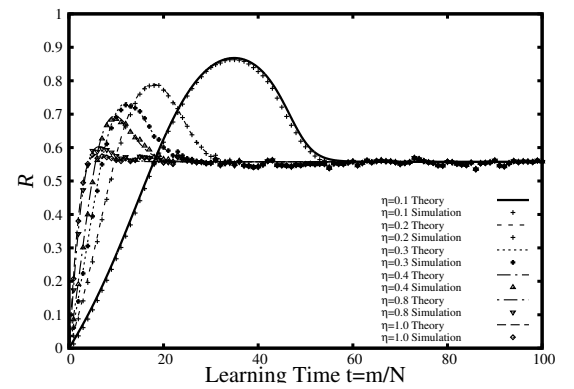


図 3.5 アダトロン学習における R のダイナミクス ($\sigma_B^2 = 1.0$).

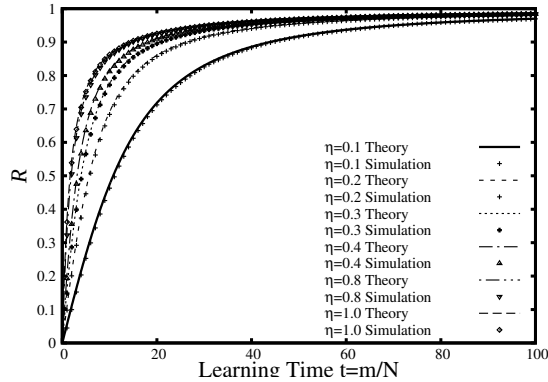


図 3.6 ヘブ学習における R のダイナミクス ($\sigma_B^2 = 1.0$).

第 4 章 漸近解析

パーセプトロン学習における非単調性については、学習曲線のダイナミクスを線形システムに近似し、収束点周りの挙動を調べることで理論的な説明が行われている [13]. 具体的には学習時間 t が $t \rightarrow \infty$ と十分に時間を経て、 R, l の変化量が小さくなった極限での漸近系を調べる. ここでは, [13] を踏まえ、アダルトロン学習における非単調性についても理論的に説明を行う. また、ヘブ学習についての単調性についても同様に極限における漸近系を調べる.

以後、解析の簡単のため教師出力に重畳する観測ノイズ n_B の分散 $\sigma_B^2 = 1.0$ と固定する.

4.1 パーセプトロン学習における非単調性の漸近解析 [13]

第 3 章において, [13] の近似的なアンサンブル平均が正しいことを示した. そこで、この章では解析の簡単のため, [13] のアンサンブル平均を使用する. この場合、パーセプトロン学習における R, l の微分方程式はそれぞれ、

$$\frac{dl}{dt} = 0.28\eta R - 0.4\eta + \frac{-0.24\eta^2 R + 0.5\eta^2}{2l}, \quad (4.11)$$

$$\frac{dR}{dt} = \frac{0.28\eta(1 - R^2)}{l} - \frac{R}{2l^2}(-0.24\eta^2 R + 0.5\eta^2), \quad (4.12)$$

となる. 非単調性の解析を行うために, $t \rightarrow \infty$ のときに, R, l が 0 に収束するような変数 e, d を以下におく.

$$e = (1 - R) - 0.30, \quad d = 1/l - 1.23/\eta, \quad (4.13)$$

式 (4.11), 式 (4.12) は, $e, d \ll 1$ の条件の時, 以下のように変数変換される.

$$\begin{aligned} \frac{de}{dt} = & 0.1162\eta^2 d^2 + 0.1428\eta d - 0.593542\eta e d - 0.605574e \\ & - 0.082\eta^2 e d^2 - 0.01494\eta e^2 d + 0.163172e^2 - 0.12\eta^2 e^2 d^2, \end{aligned} \quad (4.14)$$

$$\begin{aligned} \frac{dd}{dt} = & -0.408\eta d^2 - 0.250699d - 0.16241\eta e d^2 + 0.144509e d \\ & - 0.166\eta^2 d^3 + 0.200152\frac{e}{\eta} - 0.12\eta^2 e d^3. \end{aligned} \quad (4.15)$$

さらに式 (4.14), 式 (4.15) を固定点 $(e, d) = (0, 0)$ 周りで 1 次テイラー展開を行うと,

$$\begin{pmatrix} de/dt \\ dd/dt \end{pmatrix} = \begin{pmatrix} -0.606 & 0.143\eta \\ 0.200/\eta & -0.251 \end{pmatrix} \begin{pmatrix} e \\ d \end{pmatrix} \quad (4.16)$$

となり, この行列は線形システムと呼ばれており, 線形システムの固有値, 固有ベクトルが収束点回りの挙動を決定する. 式 (4.16) の固有値, 固有ベクトルは, 以下に計算される.

$$\lambda_1 = -0.67, \quad u_1 = \begin{pmatrix} \eta \\ -0.43 \end{pmatrix}, \quad (4.17)$$

$$\lambda_2 = -0.18, \quad u_2 = \begin{pmatrix} \eta \\ 3.07 \end{pmatrix}. \quad (4.18)$$

理論の正当性を確認するために, 式 (4.14), 式 (4.15) をルンゲクッタ法で求め, e, d 平面上で描いたものが図 4.1 であり, 理論曲線と計算機実験の結果がよく一致している. 非単調性は e が減少していき, 途中から増加するという形で現れる.

まず, 状態ベクトル $(e, d)^T$ は固有値の関係から, $\lambda_1 < \lambda_2 < 0$ の速さで収束する. すなわち, 固有ベクトル u_2 が u_1 よりもゆっくりと減少する. 図 4.2 において, (e, d) 平面の u_1 は第 1/3 象限に, u_2 は第 2/4 象限に位置する. 固有値の関係上, 曲線 (実線) は第 4 象限から第 3 象限に移動し, u_1 が収束点に収束したのち, ゆっくりと u_2 方向に沿って原点に収束する挙動となる. このように非単調性が, 固有値の大きさの差および固有ベクトルの方向の変化により起因することがわかる.

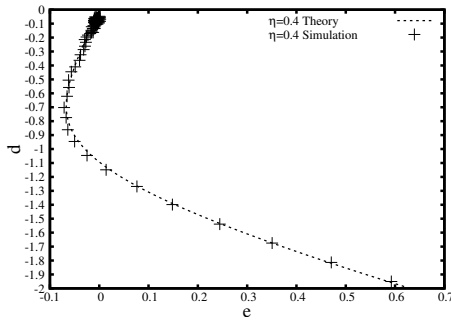


図 4.1 (e, d) の描跡 ($\sigma_B^2 = 1.0$).

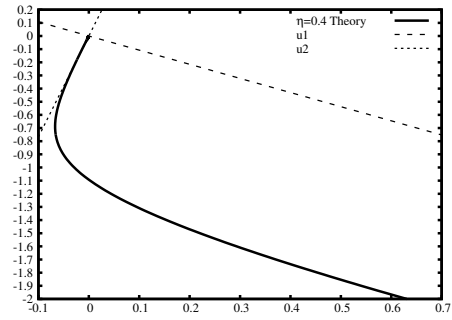


図 4.2 (e, d) と固有ベクトル ($\sigma_B^2 = 1.0$).

4.2 アダptron学習における非単調性の漸近解析

アダptron学習の非単調性についても同様に、線形システムを用いることで理論的に説明を行う。まず解析の簡単のためアダptron学習のアンサンブル平均を R の収束点である $R = 0.56$ 周りで一次テイラー展開を行うと、

$$\langle fu \rangle \approx 0.41\eta R - 0.48\eta \quad (4.21)$$

$$\langle fv \rangle \approx -0.30\eta R + 0.17\eta \quad (4.22)$$

$$\langle f^2 \rangle \approx -0.41\eta^2 R + 0.48\eta^2, \quad (4.23)$$

と近似することができる。学習時間 t が $t \rightarrow \infty$ と十分に時間を経て、 R, l の変化量が小さくなった極限での漸近系を調べると、 $R \rightarrow 0.56, l = 2.0/\eta$ であることが分かった。パーセプトロン学習の時と同様に変数 e, d を以下に用いる。

$$e = (1 - R) - 0.44, \quad d = 1/l - 2.0/\eta. \quad (4.24)$$

$e, d \ll 1$ の条件の時、 e, d の微分方程式は

$$\begin{aligned} \frac{de}{dt} = & 0.0701122\eta^2 d^2 + 0.138224\eta d - 0.3208\eta ed - 0.6e \\ & - 0.0104\eta^2 ed^2 - 0.41\eta e^2 d - 0.205\eta^2 e^2 d^2, \end{aligned} \quad (4.25)$$

$$\begin{aligned} \frac{dd}{dt} = & -0.5008\eta d^2 - 0.5008d - 0.82\eta ed^2 - 0.82ed \\ & - 0.1252\eta^2 d^3 - 0.205\eta^2 ed^3, \end{aligned} \quad (4.26)$$

さらに式 (4.25), 式 (4.26) を固定点 $(e, d) = (0, 0)$ 周りで 1 次テイラー展開を行うと、

$$\begin{pmatrix} de/dt \\ dd/dt \end{pmatrix} = \begin{pmatrix} -0.60 & 0.138224\eta \\ 0 & -0.5008 \end{pmatrix} \begin{pmatrix} e \\ d \end{pmatrix} \quad (4.27)$$

と計算され、この線形システム行列の固有値、固有ベクトルはそれぞれ

$$\lambda_1 = -0.6, \quad u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad (4.28)$$

$$\lambda_2 = -0.5008, \quad u_2 = \begin{pmatrix} 1.39332\eta \\ 1 \end{pmatrix}. \quad (4.29)$$

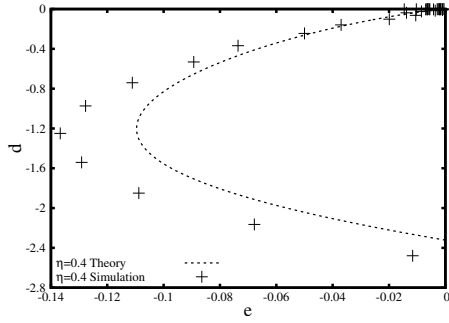


図 4.3 (e, d) の描跡 ($\sigma_B^2 = 1.0$).

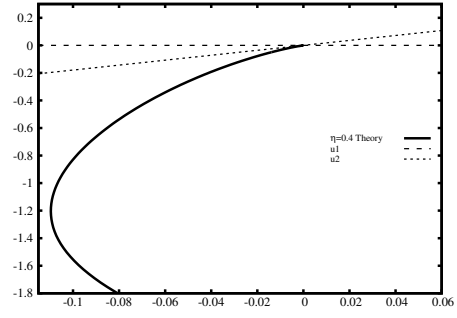


図 4.4 (e, d) と固有ベクトル ($\sigma_B^2 = 1.0$).

となる．理論の正当性を確認するために，式 (4.25)，式 (4.26) をルンゲクッタ法で求め， e, d 平面上で描いたものが図 4.3 であり，理論曲線と計算機実験の結果がおおよそ一致している．また導出した固有値の値と固有ベクトルの方向はパーセプトロンの時と同様に非単調性について説明している (図 4.4)．

4.3 ヘブ学習における単調性の漸近解析

ヘブ学習における単調性を理論的に説明する．まず学習時間 t が $t \rightarrow \infty$ と十分に時間を経て， R, l の変化量が小さくなった極限での漸近系を調べる．パーセプトロン学習の時と同様に変数 e, d を以下に用いる．

$$e = 1 - R, \quad d = 1/l. \quad (4.31)$$

e, d の微分方程式は，

$$\frac{de}{dt} = -\frac{2\eta}{\sqrt{\pi}}de + \frac{\eta^2}{2}d^2 + \frac{\eta}{\sqrt{\pi}}de^2 - \frac{\eta^2}{2}d^2e, \quad (4.32)$$

$$\frac{dd}{dt} = -\frac{\eta}{\sqrt{\pi}}d^2 + \frac{\eta}{\sqrt{\pi}}d^2e - \frac{\eta^2}{2}d^3, \quad (4.33)$$

と変数変換される．

式 (4.32)，式 (4.33) は 1 次の項を持たないため，線形システムに近似することができない．その代わり，これらの微分方程式の漸近解を [3] と同様に，解析的に求めることで単調性を示す．

$e, d \ll 1$ の条件の時, 式 (4.32), 式 (4.33) は [3] と同様に以下の式に以下のように近似される.

$$\frac{de}{dt} \approx -\frac{2\eta}{\sqrt{\pi}}de + \frac{\eta^2}{2}d^2, \quad (4.34)$$

$$\frac{dd}{dt} \approx -\frac{\eta}{\sqrt{\pi}}d^2, \quad (4.35)$$

式 (4.35) から, d の漸近解は,

$$d \approx \sqrt{\pi} \frac{1}{\eta} t^{-1}, \quad (4.36)$$

となる. 式 (4.36) を式 (4.34) に代入すると,

$$\frac{de}{dt} = -2t^{-1}e + \frac{\pi}{2}t^{-2}, \quad (4.37)$$

式 (4.37) を得ることにより, e の漸近解は,

$$e \approx \frac{\pi}{2}t^{-1}, \quad (4.38)$$

とそれぞれ解析的に求まる. 式 (4.38) の e の漸近解から e は単調に減少することがわかる. すなわち R は単調に増加する. 最後に理論の正当性を確認するために, 導出した漸近解と数値実験により得られた e, d について比較を行った (図 4.5). 理論曲線と計算機実験の結果がよく一致しており, e, d がそれぞれ単調に減少していることがわかる.

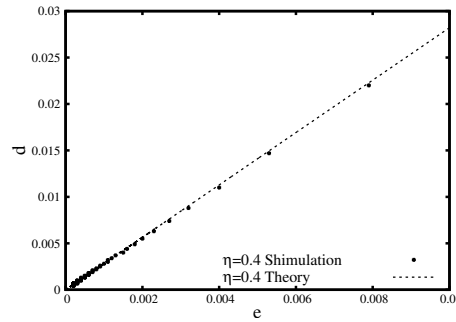


図 4.5 ヘブ学習における e, d 平面描画 $\eta = 0.4, \sigma_B^2 = 1.0$

第 5 章 学習係数の最適化

5.1 数値実験による収束性の改善

第 4 章において線形システム行列を用いて、パーセプトロン学習およびアダルトン学習における学習曲線の非単調性について理論的に説明することができた。ここでは、実際に学習曲線の非単調性を利用することで、学習係数の制御による性能向上が可能か確認してみる。学習係数 $\eta = 2.0$ から始め、図 3.4 の各交点において方向余弦 R をより 1 に近づけるような η に変動させた。具体的には各 η における R の極大点を取りうる学習時間 t を実験的に調べ、 η を変動させた。その計算機実験の結果が図 5.1 と図 5.2 である。各 η と実験的に調節した η の学習曲線を比べると、学習の収束性が改善されていることがわかる。

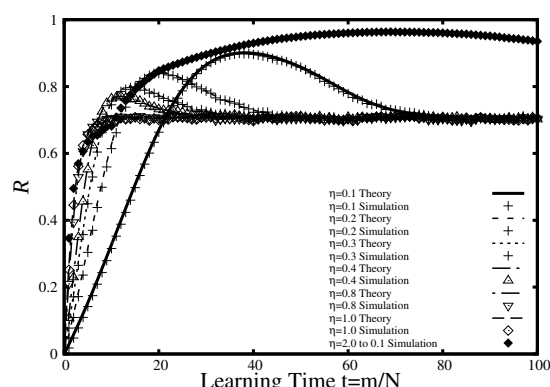


図 5.1 パーセプトロン学習における R のダイナミクスと学習係数の数値的調節結果の比較 ($\sigma_B^2 = 1.0$).

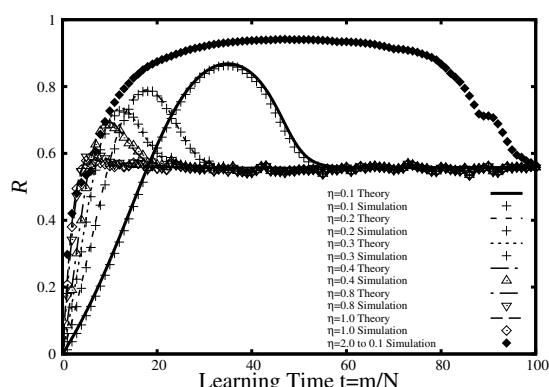


図 5.2 アダルトン学習における R のダイナミクスと学習係数の数値的調節結果の比較 ($\sigma_B^2 = 1.0$).

5.2 線形システム行列を用いた学習係数の調節式

先ほどの数値的な実験から、各 η における R の極大点における学習時間 t と η の関係式を導出することができれば、収束性の改善につながるという仮説がたつ。しかしながら、 R の極大点を求めるためには式 (3.210) を解析的に解く必要があり、それは困難である。そこで、学習曲線の非単調性を示した線形システム行列を代わり

に用いることでその関係式を導出することを試みる．すなわち e の極大点を求めるときの学習時間 t を導出する．

まず、式 (4.16) の左辺を $\dot{c}$, 2×2 行列を A と置くと、

$$\dot{c} = Ac, \quad (5.21)$$

と置くことができ、この微分方程式を解くと、

$$c(t) = \exp(At)c(0). \quad (5.22)$$

e の極大点を求めるためには、一階微分の関係式が以下になればよい．

$$\dot{c}(t) = A \exp(At)c(0) = \begin{pmatrix} 0 \\ C \end{pmatrix}. \quad (5.23)$$

ここで、 C は不定定数である．また初期条件 $t = 0$ の時に $l = 1, R = 0$ であることから、初期値 $c(0)$ は、

$$c(0) = \begin{pmatrix} 0.70 \\ 1 - 1.23/\eta \end{pmatrix}, \quad (5.24)$$

であるため、 e の極大点を導出するために求めるものは、遷移行列 $\exp(At)$ である．

5.2.1 遷移行列の導出

遷移行列である $\exp(At)$ を対角化を用いることで導出を行う．固有ベクトルを並べた行列を S , 対角化した行列を Λ と置くと、関係式 $S^{-1}AS = \Lambda$ となる． $A^k = S\Lambda^k S^{-1}$ を用いると、遷移行列 $\exp(At)$ は $S \exp(\Lambda t) S^{-1}$ となる．行列 A を式 (4.17), 式 (4.18) より対角化を行うと、 Λ および $\exp(\Lambda t)$ は

$$\Lambda = \begin{pmatrix} -0.67 & 0 \\ 0 & -0.18 \end{pmatrix} \exp(\Lambda t) = \begin{pmatrix} \exp(-0.67t) & 0 \\ 0 & \exp(-0.18t) \end{pmatrix} \quad (5.25)$$

となる．従って、遷移行列 $\exp(At)$ は

$$\exp(At) = S \exp(\Lambda t) S^{-1} \quad (5.26)$$

$$\begin{aligned} &= \begin{pmatrix} \eta & \eta \\ -0.437 & 2.95 \end{pmatrix} \begin{pmatrix} \exp(-0.67t) & 0 \\ 0 & \exp(-0.18t) \end{pmatrix} \left\{ \frac{1}{3.423\eta} \begin{pmatrix} 2.95 & -\eta \\ 0.437 & \eta \end{pmatrix} \right\} \\ &= \frac{\exp(-0.18t)}{3.423\eta} \begin{pmatrix} 2.95\eta \exp(-0.49t) + 0.473\eta & -\eta^2 \exp(-0.49t) + \eta^2 \\ -1.395 \exp(-0.49t) + 1.395 & 0.473\eta \exp(-0.49t) + 2.95\eta \end{pmatrix} \end{aligned} \quad (5.27)$$

e の極大点を求めため、遷移行列 $\exp(At)$, $c(0)$, A を式 (5.23) に代入すると、以下に計算される.

$$\dot{c}(t) = A \exp(At) c(0) \quad (5.28)$$

$$\begin{aligned} &= \frac{\exp(-0.18t)}{3.423\eta} \begin{pmatrix} -0.606 & 0.143\eta \\ 0.200/\eta & -0.251 \end{pmatrix} \\ &\times \begin{pmatrix} 2.95\eta \exp(-0.49t) + 0.473\eta & -\eta^2 \exp(-0.49t) + \eta^2 \\ -1.395 \exp(-0.49t) + 1.395 & 0.473\eta \exp(-0.49t) + 2.95\eta \end{pmatrix} \\ &\times \begin{pmatrix} 0.70 \\ 1 - 1.23/\eta \end{pmatrix} \\ &= \exp(-0.18t) \\ &\times \begin{pmatrix} -0.581 \exp(-0.49t) - 0.025 & 0.197\eta \exp(-0.49t) - 0.054\eta \\ (0.275 \exp(-0.49t) - 0.075)/\eta & -0.093 \exp(-0.49t) - 0.158 \end{pmatrix} \\ &\times \begin{pmatrix} 0.70 \\ 1 - 1.23/\eta \end{pmatrix} = \begin{pmatrix} 0 \\ C \end{pmatrix}. \end{aligned} \quad (5.29)$$

上段の式より t の導出を行うと、

$$0.197\eta \exp(-0.67t) - 0.054\eta \exp(-0.18t) = 0.649 \exp(-0.67t) - 0.048 \exp(-0.18t) \quad (5.210)$$

$$t = \log \left\{ \frac{0.197\eta - 0.649}{0.054\eta - 0.048} \right\} \quad (5.211)$$

と学習時間 t と学習係数 η の関係式を導出することができる.

5.2.2 理論値と実験値の比較

導出した理論式と経験的に求めた η と t の関係図を図 5.3 に示す. その結果, 経験的な実験値と導出した理論値がかけ離れていることがわかる. また η が大きくなるほど t が小さくなるはずであるがこの理論式はその逆であることもわかる.

理論式が実験値とかけ離れている原因を調べるため, 線形システムに近似する前の e の値および式 (4.14) をそれぞれ学習率 $\eta = 0.1$ から 1.0 まで 0.01 刻みで描画した図を図 5.4, 図 5.5 に示す.

これらの図から, 近似する前の e の極大点における学習時間 t の数値と近似した後の e の数値が非常に異なる値をとることがわかる. 考察として, 非線形なダイナ

ミクスに対して線形に近似していることからこの誤差は生じるものだと考えられる。また、学習曲線の解析については、学習時間 $t = 0$ の初期値から学習係数 η を定数と仮定しているため、学習の途中で学習係数 η を変動させたとしても数値実験の結果と一致する保証はないという点も考察の一つである。これらの考察から線形システムで非単調性について説明はできるものの、線形システムを利用して学習の収束性の改善を行うことはできない可能性があるということが分かった。

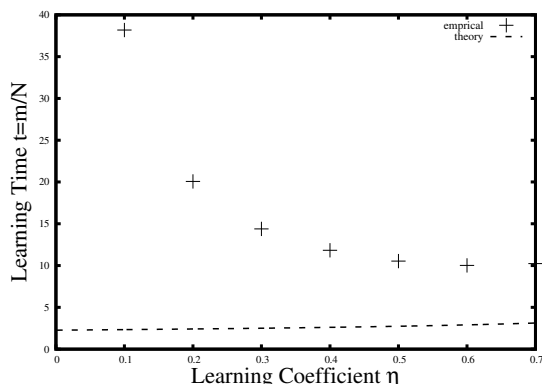


図 5.3 学習率 η と学習時間 t の理論式と経験値の比較.

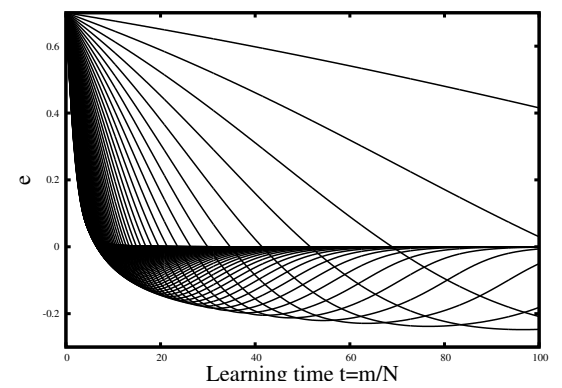


図 5.4 e のダイナミクス.

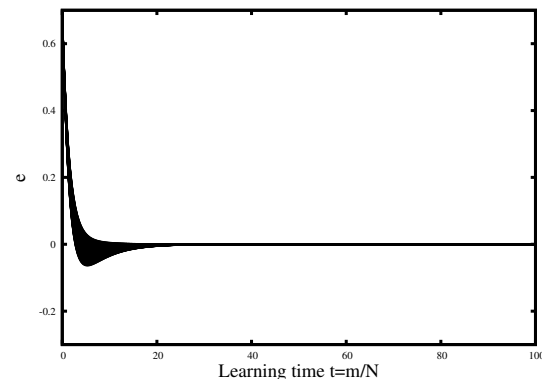


図 5.5 線形システム行列 e のダイナミクス.

第 6 章 結論

本研究では、教師信号に観測ノイズが重畳する場合の単純パーセプトロンについて、パーセプトロン学習、アダptron学習、ヘブ学習の学習曲線を情報統計力学によって理論的に導出した。その結果、パーセプトロン学習、アダptron学習では学習曲線が非単調にふるまい、 R は 1 より小さい定数に収束するが、ヘブ学習では非単調性は生じず、 $R = 1$ に単調に収束することがわかった。またパーセプトロン学習、アダptron学習においては、学習時間 t の極限において導出される線形システムにより非単調性について理論的に説明できることを示した。ヘブ学習においては、学習曲線のダイナミクスの漸近形を解析的に解くことで単調性について理論的に示した。最後にパーセプトロン学習、アダptron学習においては、線形システム行列を用いることで学習の収束性の改善を試みたが、改善はできない結果となった。

謝辞

本研究を進めるにあたり、ご指導を頂いた池田和司教授に感謝致します。本論文を作成するにあたり、指導教官の池田和司教授, 吉本潤一郎准教授, 久保孝富特任准教授, 爲井智也助教, 佐々木博昭助教から、丁寧かつ熱心なご指導を賜りました。ここに感謝の意を表します。

数理情報学研究室の秘書である谷本史さんには、日常の生活や事務的な手続きで多大な助力をいただきました。

日常の議論を通じて多くの知識や示唆を頂いた数理情報学研究室の皆様には感謝します。

参考文献

- [1] Carlucci DM. Inoue J.: Image restoration using the chiral Potts spin glass: Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics. 2547-53.(1999)
- [2] NICOLAS S.: Spin-glass models as error-correcting codes: Nature 339, 693 - 695(1989)
- [3] Nishimori, H.: Statistical Physics of Spin Glasses and Information Processing: An Introduction. Oxford Univ. Press, Oxford, UK (2001)
- [4] Rosenblatt, F.: Principle of Neurodynamics. Spartan, Washington, D. C. (1961)
- [5] S.Haykin, Neural Networks—A Comprehensive Foundation—, Macmillan College Publishing Company, New York, 1994.
- [6] M. Biehl, P. Riegler: On-Line Learning with a perceptron. Europhysics Letters, Volume 28, no. 7 (1994)
- [7] Inoue, J., Nishimori, H.: On-line AdaTron Learning of Unlearnable Rules. Physical Review E 55(4), 4544–4551 (1997)
- [8] Hara, K., Okada, M.: On-line Learning through Simple Perceptron Learning with a Margin. Neural Networks 17, 215–223 (2004)
- [9] Miyoshi, S., Hara, K., Okada, M.: Analysis of Ensemble Learning Using Simple Perceptrons Based on Online Learning Theory. Physical Review E 71, 036116 (2005)
- [10] Miyoshi, S., Okada, M.: Analysis of On-line Learning When a Moving Teacher Goes around a True Teacher. J. Physical Society of Japan 75(2), 024003 (2006)
- [11] Miyoshi, S., Okada, M.: Statistical Mechanics of Online Learning for Ensemble Teachers. J. Physical Society of Japan 75(4), 044002 (2006)
- [12] Uezu, T., Miyoshi, S., Izuo, M., Okada, M.: Theory of Time Domain Ensemble On-line Learning of Perceptron under the Existence of External Noise. J. Physical Society of Japan 76(11), 114006 (2007)

- [13] Ikeda, K., Hanzawa, H., Miyoshi, S.: Convergence Properties of Perceptron Learning with Noisy Teacher, Intelligent Science and Intelligent Data Engineering, LNCS, Vol. 7751, pp. 417–424, 2013.
- [14] 渡辺澄夫, 萩原克幸, and 赤穂昭太郎. 学習システムの理論と実現. 森北出版, 2005.
- [15] 三好誠司, オンライン学習の統計力学的解析; システム制御情報学会誌, 51(5), 216-223, (2007)

付録. 1 各ベクトルのノルム

まず, N 次元ベクトル $\mathbf{B}, \mathbf{J}^0, \mathbf{x}^m$ の l_2 ノルムを計算する. いま, 熱力学的極限 ($N \rightarrow \infty$) を仮定しているため, 大数の法則から,

$$\|\mathbf{B}\| = \sqrt{\sum_{i=1}^N B_i^2} = \sqrt{N \frac{\sum_{i=1}^N B_i^2}{N}} \xrightarrow{N \rightarrow \infty} \sqrt{N \langle B_i^2 \rangle} = \sqrt{N} \quad (\text{付録. 11})$$

が成り立つ, 同様に計算することで $\mathbf{J}^0, \mathbf{x}^m$ のノルムが式 (3.11) に求まる.

付録. 2 $\langle f^2 \rangle$ の計算

式 (3.313) の誤差関数 $\text{erf}(\cdot)$ を u の関数として原点周りで 9 次までテイラー展開を行うと, その奇数次の項は

$$\begin{aligned} \text{erf}(au + b) &\approx \frac{u}{1} \frac{2}{\pi} a \exp(-b^2) \\ &+ \frac{u^3}{3!} \frac{2 \cdot (-2)}{\pi} a^3 \exp(-b^2) \{1 - 2b^2\} \\ &+ \frac{u^5}{5!} \frac{2 \cdot (-2)^2}{\pi} a^5 \exp(-b^2) \{3 - 12b^2 + 4b^4\} \\ &+ \frac{u^7}{7!} \frac{2 \cdot (-2)^3}{\pi} a^7 \exp(-b^2) \\ &\times \{15 - 90b^2 + 60b^4 - 8b^6\} \\ &+ \frac{u^9}{9!} \frac{2 \cdot (-2)^4}{\pi} a^9 \exp(-b^2) \\ &\times \{105 - 840b^2 + 840b^4 - 224b^6 + 16b^8\} \end{aligned} \quad (\text{付録. 21})$$

となる (偶数次の項は積分で消える). ただし

$$a = \frac{R}{\sqrt{2(1 - R^2)}}, \quad b = \frac{n_B}{\sqrt{2(1 - R^2)}}, \quad (\text{付録. 22})$$

と置いた. 式 (付録. 21) を式 (3.313) に代入して部分積分すると, 各次数の積分は, 1 次

$$\frac{2}{1} a \frac{-1}{\sqrt{c}}, \quad (\text{付録. 23})$$

3 次

$$\frac{2 \cdot -2}{3!} a^3 \frac{2 \cdot -1}{\sqrt{c}} \left\{ 1 - 2 \left(\frac{1}{\sqrt{2(1-R^2)}} \right)^2 \left(\frac{1}{2c} \right) \right\}, \quad (\text{付録. 24})$$

5 次

$$\begin{aligned} & \frac{2 \cdot (-2)^2}{5!} a^5 \frac{4 \cdot 2 \cdot -1}{\sqrt{c}} \cdot 3 \\ & \times \left\{ 1 - 2 \left(\frac{1}{\sqrt{2(1-R^2)}} \right)^2 \left(\frac{1}{2c} \right) \right\}^2, \end{aligned} \quad (\text{付録. 25})$$

7 次

$$\begin{aligned} & \frac{2 \cdot (-2)^3}{7!} a^7 \frac{6 \cdot 4 \cdot 2 \cdot -1}{\sqrt{c}} \cdot 5 \cdot 3 \\ & \times \left\{ 1 - 2 \left(\frac{1}{\sqrt{2(1-R^2)}} \right)^2 \left(\frac{1}{2c} \right) \right\}^3, \end{aligned} \quad (\text{付録. 26})$$

9 次

$$\begin{aligned} & \frac{2 \cdot (-2)^4}{9!} a^9 \frac{8 \cdot 6 \cdot 4 \cdot 2 \cdot -1}{\sqrt{c}} \cdot 7 \cdot 5 \cdot 3 \\ & \times \left\{ 1 - 2 \left(\frac{1}{\sqrt{2(1-R^2)}} \right)^2 \left(\frac{1}{2c} \right) \right\}^4, \end{aligned} \quad (\text{付録. 27})$$

となる。ただし

$$c = \frac{(1 - R^2 + \sigma_B^2)}{2\sigma_B^2(1 - R^2)} \quad (\text{付録. 28})$$

である。これにより式 (3.315) が得られる。