1　階層型ニューラルネットワーク

階層型ニューラルネットワークは心理学者ローゼンブラットによって提案されたパーセプトロンから発展したものである。パーセプトロンはパターンを学習・識別することができるニューラルネットワークであり、形式ニューロンとシナプスの可塑性を用いている。また、小脳においてパーセプトロンと類似した機能を有する部分があると指摘されている。本章ではパーセプトロンはら発展した階層型ニューラルネットワークについて述べる。ニューラルネットワークによる応用のうちで最も多いのが、階層型ニューラルネットワークのバックプロパゲーションである。

1.1　パーセプトロン

1.1.1　単純パーセプトロン

パーセプトロンは 3 つの層からなる階層型ニューラルネットワークである。パーセプトロンを構成する 3 つの層はそれぞれ感覚ユニット (sensory unit)、連合ユニット (associate unit)、反応ユニット (response unit) と呼ばれる。パーセプトロンの基本形である単純パーセプトロンは感覚ユニットから連合ユニット、連合ユニットから反応ユニットへと一方向に繋がっている。また反応ユニットは 1 つのニューロンによって構成され、連合ユニットの全てのニューロンと接続される。

感覚ユニットの各ニューロンは入力を $x \in \{0,1\}$ としたとき出力値 z は

$$z = x$$

と表わされる。連合ユニットと反応ユニットのニューロンはマッカロとピッツの形式ニューロンであり、入力 (神経細胞が受け取るシナプス前神経細胞の出力) を $x_k \in \{0,1\}$ $(k = 1,2,\cdots,n)$、$w_k \in \mathbb{R}$ をシナプス結合荷重、y を膜電位の変化量としたとき、出力値 z は

$$y = \sum_{k=1}^{n} w_k x_k \tag{1.1}$$

$$f(u) = \begin{cases} 1, & u > 0 \\ 0, & u \leq 0 \end{cases} \tag{1.2}$$

$$z = f(y - h) \tag{1.3}$$

という計算式で表わされる。ここで感覚ユニットのニューロン数を m とし、i 番目 $(i = 1,2,\cdots,m)$ のニューロンの出力を $o_1^I,\cdots,o_m^I$、連合ユニットのニューロン数を n として感覚ユニットの i 番目のニューロンと連合ユニットの j 番目 $(j = 1,2,\cdots n)$ のユニットとのシナプス結合荷重を $w_{i,j}^{I,M}$ とすると連合ユニットの j 番目のニューロンへの入力は

$$\sum_{i=1}^{m} w_{i,j}^{I,M} o_i^I$$

と表わすことができる。この式は式 (1.1) に対応し連合ユニットの各ニューロンの出力値は式 (1.2)、式 (1.3) によって計算されるため、θ_j^M を連合ユニットの j 番目のニューロンのしきい値とすると連合ユニットのニューロンの出力値 $o_1^M, \cdots, o_n^M$ は

$$o_j^M = \begin{cases} 1, & \sum_{i=1}^m w_{i,j}^{I,M} o_i^I - \theta_j^M > 0 \\ 0, & \sum_{i=1}^m w_{i,j}^{I,M} o_i^I - \theta_j^M \leq 0 \end{cases}$$

となる。同様に連合ユニットの j 番目のニューロンと反応ユニットのニューロンとのシナプス結合荷重を $w_j^{M,O}$、反応ユニットのニューロンのしきい値を θ^O とすれば、反応ユニットのニューロンの出力値 o^O は

$$o^O = \begin{cases} 1, & \sum_{j=1}^n w_j^{M,O} o_j^M - \theta^O > 0 \\ 0, & \sum_{j=1}^n w_j^{M,O} o_j^M - \theta^O \leq 0 \end{cases} \tag{1.4}$$

となる。また、この単純パーセプトロンにおいて学習するパラメータは $w_j^{M,O}$ $(j = 1, 2, \cdots, n)$ と θ^O のみであり、$w_{i,j}^{I,M}$ $(i = 1, 2, \cdots, m, j = 1, 2, \cdots, n)$ と θ_j^M は固定して学習を行う。

ここで $w_{n+1}^{M,O} = -\theta^O$、$o_{n+1}^M = 1$ とすれば

$$\begin{aligned} \sum_{j=1}^n w_j^{M,O} o_j^M - \theta^O &= \sum_{j=1}^n w_j^{M,O} o_j^M + w_{n+1}^{M,O} o_{n+1}^O \\ &= \sum_{j=1}^{n+1} w_j^{M,O} o_j^M \end{aligned}$$

と書きなおすことができる。また、新たに $\mathbf{w}^{M,O} \in \mathbb{R}^{n+1}, \mathbf{o}^M \in \{0,1\}^{n+1}$ を

$$\mathbf{w}^{M,O} = \left(w_1^{M,O}, w_2^{M,O}, \ldots, w_n^{M,O}, w_{n+1}^{M,O}\right) = \left(w_1^{M,O}, w_2^{M,O}, \ldots, w_n^{M,O}, -\theta^O\right)$$

$$\mathbf{o}^M = \left(o_1^M, o_2^M, \ldots, o_n^M, o_{n+1}^M\right) = \left(o_1^M, o_2^M, \ldots, o_n^M, 1\right)$$

とするとベクトル同士の内積によって

$$\sum_{j=1}^n w_j^{M,O} o_j^M - \theta^O = \sum_{j=1}^{n+1} w_j^{M,O} o_j^M = \mathbf{w}^{M,O} \cdot \mathbf{o}^M \tag{1.5}$$

が成立する。

1.1.2 単純パーセプトロンの学習

単純パーセプトロンの学習法のうち、代表的なものである誤り訂正法 (learning by error-correction) について説明する。誤り訂正法は、複数の入力と教師信号 (入力を与えた際に求められる出力値) の組を用いて学習を行う。おおまかな手順は以下の通りである。

1. パーセプトロンに入力を与えて出力値を計算する
2. 得られた出力値と教師信号を比較してシナプス結合荷重としきい値によるベクトル $\mathbf{w}^{M,O}$ を修正する
3. 修正がなくなるまで 1 へ戻る

シナプス結合荷重としきい値によるベクトル $\mathbf{w}^{M,O}$ を修正した $\mathbf{w}'^{M,O}$ は

$$\Delta \mathbf{w}^{M,O} = \eta \left(t^O - o^O \right) \mathbf{o}^M$$

によって計算される $\Delta \mathbf{w}^{M,O}$ を用いて

$$\mathbf{w}'^{M,O} = \mathbf{w}^{M,O} + \Delta \mathbf{w}^{M,O} \tag{1.6}$$

と表わされる。ここで $t^O \in \{0, 1\}$ は入力に対する教師信号、$\eta \in (0, 1]$ は学習率 (learning rate) である。学習率は学習に大きく影響を与える定数であり、実際に学習する際には学習率 η は適切に決める必要がある。

t^O, o^O 共に 0, 1 のどちらかであるため $t^O - o^O$ がとりうる値は

$$t^O - o^O = \begin{cases} 1, & t^O = 1, o^O = 0 \\ -1, & t^O = 0, o^O = 1 \\ 0, & t^O = o^O \end{cases}$$

であり $\Delta \mathbf{w}^{M,O}$ は

$$\Delta \mathbf{w}^{M,O} = \begin{cases} \eta \mathbf{o}^M, & t^O = 1, o^O = 0 \\ -\eta \mathbf{o}^M, & t^O = 0, o^O = 1 \\ \mathbf{0}, & t^O = o^O \end{cases}$$

と書くことができる。また、教師信号 t^O は入力に対して o^O がとるべき値であるため、式 (1.6) はパーセプトロンが入力に対して誤った値を出力したときにシナプス結合荷重としきい値を修正するという式である。さらに修正後の $\mathbf{w}'^{M,O}$ を用いて同じ入力に対する式 (1.5) の値を計算すると

$$\begin{aligned} \mathbf{w}'^{M,O} \cdot \mathbf{o}^O &= \left(\mathbf{w}^{M,O} + \Delta \mathbf{w}^{M,O} \right) \cdot \mathbf{o}^M \\ &= \mathbf{w}^{M,O} \cdot \mathbf{o}^O + \Delta \mathbf{w}^{M,O} \cdot \mathbf{o}^M \\ &= \mathbf{w}^{M,O} \cdot \mathbf{o}^O + \left[\eta \left(t^O - o^O \right) \mathbf{o}^M \right] \cdot \mathbf{o}^M \\ &= \mathbf{w}^{M,O} \cdot \mathbf{o}^O + \eta \left(t^O - o^O \right) \mathbf{o}^M \cdot \mathbf{o}^M \\ &= \mathbf{w}^{M,O} \cdot \mathbf{o}^O + \eta \left(t^O - o^O \right) \left\| \mathbf{o}^M \right\|^2 \end{aligned}$$

となり $t^O - o^O$ がとりうる値から、$t^O = 0, o^O = 1$ のときは式 (1.5) の値は $\eta \|\mathbf{o}^M\|^2$ だけ減ることで o^O が 0 になりやすくなり、$t^O = 1, o^O = 0$ のときは式 (1.5) の値は $\eta \|\mathbf{o}^M\|^2$ だけ増えることで o^O が 1 になりやすくなり、それ以外の場合は式 (1.5) の値と o^O の値は変化しないことがわかる。

学習をつづけることを時間の経過とらえることで式 (1.5) はヘブの強化則ということができ、式 (1.5) はシナプス結合荷重を変化させているためシナプスの可塑性を利用しているともいうことができる。

式 (1.6) によるシナプス結合荷重の修正を複数の入力と教師信号の組に対して繰り返し行なうことで学習を進める。誤り訂正法による単純パーセプトロンの学習の詳細な手順は以下の通りである。

Step1 複数の入力パターン $\mathbf{i}_p^I = (i_{p,1}^I, i_{p,2}^I, \cdots, i_{p,m}^I) \in \mathbb{R}^m$ $(p = 1, 2, \cdots, P)$ と、p 番目の入力パターンと対応する教師信号 $t_p^O \in \{0, 1\}$ を用意する。

Step2 シナプス結合荷重 $\mathbf{w}^{M,O} = \left(w_1^{M,O}, w_2^{M,O}, \cdots, w_{n+1}^{M,O}\right)$ の初期値を乱数によって小さい値に設定する。そして、学習率 $\eta \in (0, 1]$ を与える。

Step3 ある $p \in \{1, 2, \cdots, P\}$ に対して、p 個目の入力パターン $\mathbf{i}_p^I$ に対する連合ユニットの出力値 $o_{p,j}^M$ $(j = 1, 2, \cdots, n)$ を

$$o_{p,j}^M = \begin{cases} 1, & \sum_{i=1}^m w_{i,j}^{I,M} i_{p,i}^I - \theta_j^M > 0 \\ 0, & \sum_{i=1}^m w_{i,j}^{I,M} i_{p,i}^I - \theta_j^M \leq 0 \end{cases}$$

によって計算する。

Step4 得られた連合ユニットのニューロンの出力 $o_{p,j}^{M,O}$ $(j = 1, 2, \cdots, n)$ によって表わされるベクトル $\mathbf{o}_p^M = \left(o_{p,1}^M, o_{p,2}^M, \cdots, o_{p,n}^M, 1\right)$ とシナプス結合荷重から反応ユニットのニューロンの出力値 o_p^O を

$$o_p^O = \begin{cases} 1, & \mathbf{w}^{M,O} \cdot \mathbf{o}_p^M > 0 \\ 0, & \mathbf{w}^{M,O} \cdot \mathbf{o}_p^M \leq 0 \end{cases}$$

によって計算する。

Step5 反応ユニットのニューロンの出力値の計算に用いた入力パターンベクトル $\mathbf{t}_p^I$ と対応する教師信号 t_p^O と得られた反応ユニットの出力値 o_p^O を用いて計算される $\Delta \mathbf{w}^{M,O}$ を

$$\Delta \mathbf{w}^{M,O} = \eta \left(t_p^O - o_p^O\right) \mathbf{o}_p^M$$

によって計算する。そして、シナプス結合荷重 $\mathbf{w}^{M,O}$ に $\Delta \mathbf{w}^{M,O}$ を加算することでシナプス結合荷重を修正する。

Step6 任意の $p \in \{1, 2, \cdots, P\}$ に対して、入力パターンベクトル $\mathbf{i}^I_p$ から計算される反応ユニットの出力値 o^O_p と t^O_p が等しいならば終了する。そうでない場合は Step3 から Step5 までを繰り返す。

実際に学習を行う際には Step6 において、シナプス結合荷重の変化 $\Delta \mathbf{w}^{M,O}$ が一定の値より小さくなったら終了としたり、もしくは繰り返し回数に上限を設けて繰り返し回数が上限を越えたときに終了とするなどの条件も用いる。Step2 において学習率 η の値は大きいほど学習が早く進むがシナプス結合荷重や、教師信号と計算された反応ユニットの出力値との差が発散、振動しやすくなる。そのため η には適切な値を用いる必要がある。適切な η の値は問題の対象によって変化する。

単純パーセプトロンは入力パターンを 2 つの集合に分類することができる機械であり、分類する事を認識という。単純パーセプトロンにおける認識とは、入力パターンに対して反応ユニットのニューロンの出力値を計算することであり、出力値が 0 か 1 かによって入力パターンが 2 つの集合に分類される。式 (1.4) から反応ユニットのニューロンの出力値は式 (1.5) の値が正か負かによって決まることがわかる。式 (1.5) が

$$\sum_{j=1}^{n} w^{M,O}_j o^M_j - \theta^O = \sum_{j=1}^{n+1} w^{M,O}_j o^M_j = \mathbf{w}^{M,O} \cdot \mathbf{o}^M = 0$$

という $n+1$ 次元の超平面を表わしていると考えると、認識することは入力パターンと対応する $\mathbf{o}^M$ を超平面によって 2 つの集合にわけることと言いかえることができる。認識することは、入力パターン $\mathbf{i}^I \in \{0, 1\}^m$ を座標変換によって $\mathbf{o}^M \in \{0, 1\}^{n+1}$ へと変換し $\mathbf{o}^M$ を n 次元の超平面によって 2 つの集合のどちらかに分類わけすることと言いかえることができる。

また、単純パーセプトロンにおいて学習するパラメータはシナプス結合荷重 $\mathbf{w}^{M,O}$ のみであるため、入力パターン $\mathbf{i}^I$ から $\mathbf{o}^M$ への座標変換のパラメータは変化せずに n 次元の超平面の係数のみが変化することとなる。したがって、単純パーセプトロンの学習は入力パターンを適切に識別可能な n 次元超平面を求めることと同値となる。問題によってはどのような超平面を用いても正しく識別することができないものが存在し、逆に正しく識別することができる超平面が存在するとき線形分離可能という。

定義 1 : 線形分離可能 (linear separable) $F^+, F^- \in \{0, 1\}^n$ $(n \in \mathbb{N})$ としたとき任意の $\mathbf{o}^+ \in F^+, \mathbf{o}^- \in F^-$ に対して

$$\mathbf{w} \cdot \mathbf{o}^+ > 0$$
$$\mathbf{w} \cdot \mathbf{o}^- \leq 0$$

となる $\mathbf{w} \in \mathbb{R}^n$ が存在することを F^+ と F^- は線形分離可能であるという。

パーセプトロンでは線形分離可能でない問題に対して適切な識別を行なうことはできない。したがってパーセプトロンが扱うのは線形分離可能な問題である。また線形分離可能

な問題に対してであれば誤り訂正法を用いて必ず適切なパラメータを学習することができることが示される。

定理 1 : パーセプトロンの収束定理 (perceptron convergence theorem)

定理 *1* の証明. 正しい識別を与えるシナプス結合荷重のうち任意の $\mathbf{o} \in F^+ \cup F^-$ に対して

$$\mathbf{o} \cdot \mathbf{w}^* \neq \mathbf{0}$$

となる $\mathbf{w}^* \in \mathbb{R}^{n+1}, \mathbf{w}^* \neq \mathbf{0}$ をひとつ選ぶ。実数空間が連続であることから $\mathbf{w}^*$ は必ず存在する。

また、シナプス結合荷重の初期値を $\mathbf{w}_0$、t 回修正を繰り返したシナプス結合荷重を $\mathbf{w}_t$ とすると式 (1.6) から、ある $\mathbf{o}^M \in \{0,1\}^{n+1}$ が存在して

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \eta \left(t^O - o^O\right) \mathbf{o}^M \tag{1.7}$$

と書くことができる。ここで t^O は $\mathbf{o}^M$ と対応する教師信号であり、o^O は $\mathbf{o}^M$ を連合ユニットのニューロンの出力値からなるベクトル、$\mathbf{w}_t$ を連合ユニットと反応ユニット間のシナプス結合荷重として計算された反応ユニットの出力値である。

式 (1.7) の両辺と $\mathbf{w}*$ の内積をとり、変形すると

$$\begin{aligned}\mathbf{w}^* \cdot \mathbf{w}_{t+1} &= \mathbf{w}^* \cdot \left[\mathbf{w}_t + \eta \left(t^O - o^O\right) \mathbf{o}^M\right] \\ &= \mathbf{w}^* \cdot \mathbf{w}_t + \eta \left(t^O - o^O\right) \mathbf{w}^* \cdot \mathbf{o}^M \end{aligned}\tag{1.8}$$

となる。式 (1.8) は修正が発生したときの式であるため $t^O \neq o^O$ であり $\left(\mathbf{w} * \cdot \mathbf{o}^M\right)\left(\mathbf{w}_t \cdot \mathbf{o}^M\right) \leq 0$ である。したがって式 (1.4) から

$$o^O = \begin{cases} 0, & \mathbf{w} * \cdot \mathbf{o}^M > 0 \\ 1, & \mathbf{w} * \cdot \mathbf{o}^M \leq 0 \end{cases}$$

$$t^O = \begin{cases} 1, & \mathbf{w} * \cdot \mathbf{o}^M > 0 \\ 0, & \mathbf{w} * \cdot \mathbf{o}^M \leq 0 \end{cases}$$

であり、$\eta > 0$ であることと $\mathbf{w}^*$ の定義から

$$\eta \left(t^O - o^O\right) \mathbf{w}^* \cdot \mathbf{o}^M > 0$$

が成立する。ここで δ を

$$\delta = \min_{\mathbf{o} \in F^+ \cup F^-} \mathbf{w}^* \cdot \mathbf{o}^M$$

とすると δ は非ゼロな実数となり、任意の $\mathbf{o} \in F^+ \cup F^-$ に対して

$$\eta \left(t^O - o^O\right) \mathbf{w}^* \cdot \mathbf{o}^M \geq \eta\delta \tag{1.9}$$

となる。そして式 (1.8) と式 (1.9) から、任意の t ($t = 1, 2, \cdots$) に対して

$$\begin{aligned} \mathbf{w}^* \cdot \mathbf{w}_t + \eta \left(t^O - o^O \right) \mathbf{w}^* \cdot \mathbf{o}^M &\geq \mathbf{w}^* \cdot \mathbf{w}_t + \eta\delta \\ \rightleftharpoons \mathbf{w}^* \cdot \mathbf{w}_{t+1} &\geq \mathbf{w}^* \cdot \mathbf{w}_t + \eta\delta \end{aligned}$$

となる。したがって、初期値 $\mathbf{w}_0$ に n 回修正を行なった $\mathbf{w}_n$ に対して

$$\mathbf{w}^* \cdot \mathbf{w}_n \geq \mathbf{w}^* \cdot \mathbf{w}_0 + n\eta\delta \tag{1.10}$$

が成立する。

また式 (1.6) から $\|\mathbf{w}_t\|^2$ は $\mathbf{w}_t$ によって

$$\begin{aligned} \|\mathbf{w}_{t+1}\|^2 &= \left\| \mathbf{w}_t + \eta \left(t^O - o^O \right) \mathbf{o}^M \right\|^2 \\ &= \|\mathbf{w}_t\|^2 + 2\eta \left(t^O - o^O \right) \mathbf{w}_t \cdot \mathbf{o}^M + \eta^2 \left(t^O - o^O \right)^2 \left\| \mathbf{o}^M \right\|^2 \end{aligned} \tag{1.11}$$

と表わすことができ、式 (1.8) と同様に式 (1.11) は修正が発生したときの式であるため $t^O \neq o^O$ であり $\left(\mathbf{w}^* \cdot \mathbf{o}^M \right) \left(\mathbf{w}_t \cdot \mathbf{o}^M \right) \leq 0$ である。したがって式 (1.4) から

$$o^O = \begin{cases} 1, & \mathbf{w}_t \cdot \mathbf{o}^M > 0 \\ 0, & \mathbf{w}_t \cdot \mathbf{o}^M \leq 0 \end{cases} \tag{1.12}$$

$$t^O = \begin{cases} 0, & \mathbf{w}_t \cdot \mathbf{o}^M > 0 \\ 1, & \mathbf{w}_t \cdot \mathbf{o}^M \leq 0 \end{cases} \tag{1.13}$$

であり、$\eta > 0$ であることとから

$$2\eta \left(t^O - o^O \right) \mathbf{w}_t \cdot \mathbf{o}^M \leq 0 \tag{1.14}$$

が成立する。ここで M を

$$M = \max_{\mathbf{o} \in F^+ \cup F^-} \left\| \mathbf{o}^M \right\|$$

とすると $0 < \eta \leq 1$ と式 (1.12)、式 (1.13)、式 (1.14) から、任意の t ($t = 1, 2, \cdots$) と $\mathbf{o}^M$ に対して

$$\begin{aligned} \|\mathbf{w}_t\|^2 + 2\eta \left(t^O - o^O \right) \mathbf{w}_t \cdot \mathbf{o}^M + \eta^2 \left(t^O - o^O \right)^2 \left\| \mathbf{o}^M \right\|^2 &\leq \|\mathbf{w}_t\|^2 + \eta^2 \left(t^O - o^O \right)^2 \left\| \mathbf{o}^M \right\|^2 \\ \|\mathbf{w}_t\|^2 + \eta^2 \left(t^O - o^O \right)^2 \left\| \mathbf{o}^M \right\|^2 &\leq \|\mathbf{w}_t\|^2 + \eta^2 \left(t^O - o^O \right)^2 M^2 = \|\mathbf{w}_t\|^2 + \eta^2 M^2 \end{aligned}$$

となり、式 (1.11) から

$$\|\mathbf{w}_{t+1}\|^2 \leq \|\mathbf{w}_t\|^2 + \eta^2 M^2$$

が得られる。したがって、初期値 $\mathbf{w}_0$ に n 回修正を行なった $\mathbf{w}_n$ に対して

$$\|\mathbf{w}_n\|^2 \leq \|\mathbf{w}_0\|^2 + n\eta^2 M^2 \tag{1.15}$$

が成立する。

さらに $\mathbf{w}^*$ が非ゼロベクトルであることから、式 (1.10) は $\mathbf{w}^*$ と $\mathbf{w}_n$ のなす角を θ とすると

$$\begin{aligned}
\mathbf{w}^* \cdot \mathbf{w}_n &\geq \mathbf{w}^* \cdot \mathbf{w}_0 + n\eta\delta \\
\rightleftharpoons \|\mathbf{w}^*\|\|\mathbf{w}_n\| \cos\theta &\geq \mathbf{w}^* \cdot \mathbf{w}_0 + n\eta\delta \\
\rightleftharpoons \|\mathbf{w}^*\|\|\mathbf{w}_n\| &\geq \mathbf{w}^* \cdot \mathbf{w}_0 + n\eta\delta \\
\rightleftharpoons \|\mathbf{w}_n\| &\geq \frac{\mathbf{w}^* \cdot \mathbf{w}_0 + n\eta\delta}{\|\mathbf{w}^*\|} \qquad (1.16)
\end{aligned}$$

と書きかえることができる。

式 (1.15) と式 (1.16) から

$$\frac{(\mathbf{w}^* \cdot \mathbf{w}_0 + n\eta\delta)^2}{\|\mathbf{w}^*\|^2} \leq \|\mathbf{w}_n\|^2 \leq \|\mathbf{w}_0\|^2 + n\eta^2 M^2 \tag{1.17}$$

となる。ここで式 (1.17) を見たす $\|\mathbf{w}_n\|^2$ が存在するためには

$$\frac{(\mathbf{w}^* \cdot \mathbf{w}_0 + n\eta\delta)^2}{\|\mathbf{w}^*\|^2} \leq \|\mathbf{w}_0\|^2 + n\eta^2 M^2 \tag{1.18}$$

が成立している必要があり、式 (1.18) を変形すると

$$\begin{aligned}
&(\mathbf{w}^* \cdot \mathbf{w}_0 + n\eta\delta)^2 \leq \|\mathbf{w}^*\|^2\|\mathbf{w}_0\|^2 + n\eta^2 M^2\|\mathbf{w}^*\|^2 \\
\rightleftharpoons\ & n^2\eta^2\delta^2 + n\left[2\eta\delta\,(\mathbf{w}^* \cdot \mathbf{w}_0) - \eta^2 M^2\|\mathbf{w}^*\|^2\right] + (\mathbf{w}^* \cdot \mathbf{w}_0)^2 - \|\mathbf{w}^*\|^2\|\mathbf{w}_0\|^2 \leq 0 \qquad (1.19)
\end{aligned}$$

となる。また、式 (1.19) の左辺は下に凸な放物線であるため

$$n^2\eta^2\delta^2 + n\left[2\eta\delta\,(\mathbf{w}^* \cdot \mathbf{w}_0) - \eta^2 M^2\|\mathbf{w}^*\|^2\right] + (\mathbf{w}^* \cdot \mathbf{w}_0)^2 - \|\mathbf{w}^*\|^2\|\mathbf{w}_0\|^2 = 0 \tag{1.20}$$

の解を n_-, n_+ $(n_- \leq n_+)$ とすると式 (1.19) が成立することと

$$n_- \leq n \leq n_+$$

が成立することは同値である。

ここで式 (1.20) の判別式 D は

$$\begin{aligned}
D &= \left[2\eta\delta\,(\mathbf{w}^* \cdot \mathbf{w}_0) - \eta^2 M^2\|\mathbf{w}^*\|^2\right]^2 - 4\eta^2\delta^2\left[(\mathbf{w}^* \cdot \mathbf{w}_0)^2 - \|\mathbf{w}^*\|^2\|\mathbf{w}_0\|^2\right] \\
&= \eta^4 M^4\|\mathbf{w}^*\|^4 - 4\eta^3\delta M^2\,(\mathbf{w}^* \cdot \mathbf{w}_0)\,\|\mathbf{w}^*\|^2 + 4\eta^2\delta^2\|\mathbf{w}^*\|^2\|\mathbf{w}_0\|^2 \\
&= \left[\eta^4 M^4\|\mathbf{w}^*\|^2 - 4\eta^3\delta M^2\,(\mathbf{w}^* \cdot \mathbf{w}_0) + 4\eta^2\delta^2\|\mathbf{w}_0\|^2\right]\|\mathbf{w}^*\|^2 \qquad (1.21)
\end{aligned}$$

となり、式 (1.21) は $\mathbf{w}^* = \frac{2\delta}{\eta M^2}\mathbf{w}_0$ のときのみ 0 となり、それ以外の場合に正となることがわかる。さらに $(\mathbf{w}^* \cdot \mathbf{w}_0)^2 - \|\mathbf{w}^*\|^2\|\mathbf{w}_0\|^2 \leq 0$ であるため

$$n_- \leq 0 \leq n_+$$

が必ず成立し $\mathbf{w}^* = \frac{2\delta}{\eta M^2}\mathbf{w}_0$ でないとき n は有限の値 n_+ によってバウンドされる。また、$\mathbf{w}^* = \frac{2\delta}{\eta M^2}\mathbf{w}_0$ であるとき $n_- = n_+ = 0$ であり、$\mathbf{w}^*$ の正の実数倍ベクトルは正しく分離するため $n = 0$ となる。

以上より、シナプス結合荷重の初期値 $\mathbf{w}_0$ がゼロベクトルであるときに誤り訂正法によるパーセプトロンの学習が有限回の修正で収束することが示された。 □

線形分離可能な問題に対しては単純パーセプトロンは適切に学習することができることが示された。実際に単純パーセプトロンを用いる場合、感覚ユニットと連合ユニット間のシナプス結合荷重は学習せずに固定する。感覚ユニットへ線形分離可能でない入力群が与えられる場合でも、感覚ユニットと連合ユニット間でランダムに定められたシナプス結合荷重による座標変換によって、連合ユニットの出力ベクトルの空間上では線形分離可能な問題へ変換することができるようになる。連合ユニットのニューロン数 n が十分に大きいとき、入力パターン数が $2n$ 以下であるどのような問題であっても線形分離可能となることが示されている。

また、パーセプトロンの学習則である誤り訂正法をデルタ則 (delta rule) ということもある。

1.2 バックプロパゲーション

1.2.1 一般化デルタ則

前節で示したデルタ則 (誤り訂正法) を一般化した、一般化デルタ則を導出する。まず入力から出力を計算する過程を一般化する。

入力層でない層のあるニューロン j は、そのニューロンよりも前の層の任意のニューロンの出力 o_i (i は 1 から前の層のニューロン数までの整数) を受けとる。このとき、ニューロン j は前の層のニューロンとのシナプス結合荷重を前の層のニューロンに乗じ和をとっている。よってニューロン j への入力 u_j を

$$u_j = \sum_i w_{i,j} o_i \tag{1.22}$$

と定める。ここで $w_{i,j}$ は前の層のニューロン i とニューロン j のシナプス結合荷重とした。またしきい値を用いていないが、前節で用いたように前の層に出力値が 1 で固定されたニューロンを仮定することでしきい値を表現することはできる。ニューロン j の出力値 o_j は非線形な実数上の 1 変数関数 f によって

$$o_j = f(u_j) \tag{1.23}$$

として求められる。この関数 f は活性化関数もしくは入出力関数といわれる。ここで f をステップ関数とすることで形式ニューロンの式と等価とすることができる。また、

$$f(x) = \frac{1}{1 + e^{-\epsilon x}}$$

と定義されるシグモイド関数は定義域が実数全体であり、終域が実数上の開区間 $(0, 1)$ かつ単調増加連続関数であるという性質からよく用いられる。

次に学習する過程を一般化する。ここで、ある入力パターン p (p は 1 から入力パターン数までの整数) に対してパーセプトロンの出力層のニューロン j が出力すべき値を $t_{p,j}$、実際のニューロン j (j は 1 から出力層のニューロン数までの整数) の出力値を $o_{p,j}$ とし、誤差関数 E を二乗誤差によって

$$E = \frac{1}{2} \sum_{p,j} \left(o_{p,j} - t_{p,k}\right)^2 \tag{1.24}$$

と定める。この誤差関数 E の値が 0 となれば入力パターン全てを正しく識別できていることとなる。したがって、誤差関数 E の値を小さくすることを学習ということができる。

誤差関数 E は出力層のニューロンの値 $o_{p,j}$ に関する関数である。しかし出力層のニューロンの値 $o_{p,j}$ は出力層と前の層との間のシナプス結合荷重 $w_{i,j}$ の関数であるため、誤差関数 E も $w_{i,j}$ に関する (陰 (implicit) に定義された) 関数であるといえる。したがって、各シナプス結合荷重 $w_{i,j}$ に適切な $\eta > 0$ を用いた

$$\Delta w_{i,j} = -\eta \frac{\partial E}{\partial w_{i,j}} \tag{1.25}$$

を加算することで誤差関数 E の極小値 (もしくは鞍点) へ近づくことができる。これは最急降下法 (steepest descent method もしくは gradient decent method) と呼ばれる。

また、式 (1.22) と式 (1.23) から連鎖律により

$$\frac{\partial E}{\partial w_{i,j}} = \sum_{p} \frac{\partial E}{\partial o_{p,j}} \frac{\partial o_{p,j}}{\partial u_{p,j}} \frac{\partial u_{p,j}}{\partial w_{i,j}}$$

が得られる。ここで $u_{p,j}$ は入力パターン p に対して計算された出力層のニューロン j の入力値である。さらに式 (1.22) と式 (1.23) と式 (1.24) から

$$\frac{\partial E}{\partial o_{p,j}} = o_{p,j} - t_{p,j}$$

$$\frac{\partial o_{p,j}}{\partial u_{p,j}} = \left.\frac{df}{dx}\right|_{x=u_{p,j}}$$

となり、さらに $o_{p,i}$ を出力層の前の中間層の入力パターン p に対するニューロン i の出力値とすると

$$\frac{\partial u_{p,j}}{\partial w_{i,j}} = o_{p,i}$$

となる。よって式 (1.25) の各因子へ代入することで

$$\Delta w_{i,j} = -\eta \sum_{p} \left(o_{p,j} - t_{p,j}\right) \left.\frac{df}{dx}\right|_{x=u_{p,j}} o_{p,i} \tag{1.26}$$

が得られる。

式 (1.26) は一般化デルタ則と呼ばれる。一般化デルタ則では全ての入力パターンに対する誤差を計算してから、シナプス結合荷重の修正量を決め修正している。しかし、η が非常に小さい場合は入力パターン毎に修正量を決めて修正する方法でも、全体の修正量はほぼ等しくなる。

1.2.2　誤差逆伝搬法

誤差逆伝搬法はフィードフォワード型の階層型ニューラルネットワークの教師あり学習 (supervised learning) の代表的な方法である。入力パターンに対して計算された出力値と教師信号との誤差が小さくなるように、シナプス結合荷重を修正する。

誤差逆伝播法を用いることで入力層と中間層間での高次への写像も学習することができるため、線形分離可能でない問題に対しても適切なパラメータを学習することができる。

実際の学習は、まず入力パターンに対して各ニューロンの出力値を計算し、その後計算された誤差から逆向きにシナプス結合荷重の修正値を計算する。生物のニューラルネットワークでは信号が逆向きに伝わることはないため、誤差逆伝搬法は実際の生物のモデルとしては適切でない。しかし工学上の応用がききやすいため広く利用されている。

以下の表に誤り訂正法と誤差逆伝搬法の違いをまとめる。

表 1: 誤り訂正法と誤差逆伝搬法の違い

	誤り訂正法	誤差逆伝搬法
対象	単純パーセプトロン	フィードフォワード型の階層型ニューラルネットワーク
出力値	2 値	実数上の開区間 $(0, 1)$
入出力関数	ステップ関数	シグモイド関数
学習する値	中間層と出力層の間のシナプス結合荷重と出力層のしきい値	任意のシナプス結合荷重としきい値
終了条件	一定回数繰り返しても収束しない場合	2 乗誤差が一定値以下になること

m 層のフィードフォワード型の階層型ニューラルネットワークを学習する場合を考える。学習に用いる入力パターンは P 個のときを考える。パターン p に対しての k 層目の j 番目のニューロンの出力値を $o^k_{p,j}$、入力値を $i^k_{p,j}$、$k-1$ 層目の i 番目のニューロンと k 層目の

j 番目のニューロンの間のシナプス結合荷重 $w_{i,j}^{k-1,k}$、k 層目の j 番目のニューロンの入出力関数を f_j^k、しきい値を θ_j^k、k 層目のニューロン数を n_k とすると

$$o_{p,j}^k = f_j^k(i_{p,j}^k) \tag{1.27}$$

$$i_{p,j}^k = \sum_{i=1}^{n_{k-1}} w_{i,j}^{k-1,k} o_{p,i}^{k-1} - \theta_j^k \tag{1.28}$$

という関係式が成立する。ここで k 層目に出力値が常に 1 である $n_k + 1$ 番目のニューロンの存在を仮定することで式 (1.28) は

$$i_{p,j}^k = \sum_{i=1}^{n_{k-1}+1} w_{i,j}^{k-1,k} o_{p,i}^{k-1}$$

と書きかえることができる。

パターン p に対する m 層目の j 番目のニューロンの教師信号を $t_{p,i}^{mk}$ し、誤差関数 E を

$$E = \sum_p E_p$$

$$E_p = \frac{1}{2}\sum_{i=1}^{n_m}\left(t_{p,i}^m - o_{p,i}^m\right)^2$$

と定義する。ここで E_p はパターン p に対する誤差ということができる。

最急降下法によってシナプス結合荷重に

$$\Delta_p w_{i,j}^{k-1,k} = -\eta\frac{\partial E_p}{\partial w_{i,j}^{k-1}} \tag{1.29}$$

を加算することで修正する。η は半開区間 $(0,1]$ 上の実数である。

式 (1.29) の右辺の偏微分は連鎖律によって

$$\frac{\partial E_p}{\partial w_{i,j}^{k-1}} = \frac{\partial E_p}{\partial i_{p,j}^k}\frac{\partial i_{p,j}^k}{\partial w_{i,j}^{k-1}}$$

と書くことができ、

$$\frac{\partial i_{p,j}^k}{\partial w_{i,j}^{k-1}} = o_{p,i}^{k-1}$$

であるため

$$\frac{\partial E_p}{\partial w_{i,j}^{k-1}} = \frac{\partial E_p}{\partial i_{p,j}^k} o_{p,i}^{k-1}$$

となり、

$$\delta_{p,j}^k = -\frac{\partial E_p}{\partial i_{p,j}^k}$$

と置くことで

$$-\frac{\partial E_p}{\partial w_{i,j}^{k-1}} = \delta_{p,j}^k o_{p,i}^{k-1}$$

となる。最急降下法を用いるために非ゼロな 1 以下の実数定数を学習率 η として、$w_{i,j}^{k-1,k}$ のパターン p に対する修正量 $\Delta_p w_{i,j}^{k-1,k}$ は

$$\Delta_p w_{i,j}^{k-1,k} = \eta \delta_{p,j}^k o_{p,i}^{k-1}$$

と表わすことができる。

また、$\delta_{p,j}^k$ は連鎖律と式 (1.27) によって

$$\begin{aligned}\delta_{p,j}^k &= -\frac{\partial E_p}{\partial i_{p,j}^k} \\ &= -\frac{\partial E_p}{\partial o_{p,j}^k}\frac{\partial o_{p,j}^k}{\partial i_{p,j}^k} \\ &= -\frac{\partial E_p}{\partial o_{p,j}^k}\left.\frac{df_j^k(x)}{di_{p,j}^k}\right|_{x=i_{p,j}^k}\end{aligned}$$

と書くことができる。ここで

$$\frac{\partial E_p}{\partial o_{p,j}^k} = \begin{cases} o_{p,j}^m - t_{p,j}^m & k = m \\ \sum_{l=1}^{n_k+1}\left(\frac{\partial E_p}{\partial i_{p,l}^{k+1}}\frac{\partial i_{p,j}^{k+1}}{\partial o_{p,j}^k}\right) = -\sum_{l=1}^{n_k+1}\left(\delta_{p,l}^{k+1} w_{j,l}^{k,k+1}\right) & \text{otherwise} \end{cases}$$

となり、まとめると層数 m の階層型ニューラルネットワークの学習は

$$\Delta_p w_{i,j}^{k-1,k} = \eta \delta_{p,j}^k o_{p,i}^{k-1}$$

$$i = 1, 2, \cdots, n_k + 1, \quad j = 1, 2, \cdots, n_k, \quad k = 2, 3, \cdots, m$$

$$\delta_{p,j}^k = -\frac{\partial E_p}{\partial o_{p,j}^k}\left.\frac{df_j^k(x)}{di_{p,j}^k}\right|_{x=i_{p,j}^k}$$

$$\frac{\partial E_p}{\partial o_{p,j}^k} = \begin{cases} o_{p,j}^m - t_{p,j}^m & k = m \\ \sum_{l=1}^{n_k+1}\left(\frac{\partial E_p}{\partial i_{p,l}^{k+1}}\frac{\partial i_{p,j}^{k+1}}{\partial o_{p,j}^k}\right) = -\sum_{l=1}^{n_k+1}\left(\delta_{p,l}^{k+1} w_{j,l}^{k,k+1}\right) & \text{otherwise} \end{cases}$$

として表わされる $\Delta_p w_{i,j}^{k-1,k}$ を $w_{i,j}^{k-1,k}$ に各パターンに対して加算することで行われる。

したがって、誤差逆伝搬法において入出力関数は C^1 級関数である必要がある。誤差逆伝搬法で一般に用いられるシグモイド関数は

$$f(x) = \frac{1}{1 + e^{-\epsilon x}}$$

と定められる実数全体から実数区間 $(0,1)$ への単調増加連続関数である。また、シグモイド関数の 1 階導関数は

$$\frac{df(x)}{dx} = \epsilon f(x)(1 - f(x))$$

である。

学習の際に修正するとき、各パターン毎に修正する方法と全パターンの修正量をまとめて後から修正する方法の 2 つがある。前者を逐次修正法 (またはオンライン学習)、後者を一括修正法 (またはバッチ学習もしくはフルバッチ学習) という。また、逐次修正法と一括修正法を合わせて、いくつかのパターン分の修正をまとめて行なうミニバッチ学習という方法も用いられる。厳密に逐次修正法は全パターンに対する誤差の極小値を得るわけではないが学習率 η を小さくすることで全パターンに対する誤差の極小値を得ることができる。

1.2.3 応用例

誤差逆伝搬法は広く応用に使われている。応用されている事例をいくつか示す。

英語の発音学習システム a

1.2.4 ニューラルネットワークの構造とパラメータの与え方

1.2.5 誤差逆伝搬法の改良