

階層型ニューラルネットワークに対する逐次射影法の改良

奈良先端科学技術大学院大学 *巽 啓司 TATSUMI Keiji
奈良先端科学技術大学院大学 福島 雅夫 FUKUSHIMA Masao

1 はじめに

階層型ニューラルネットワークに対する学習アルゴリズムである逐次型誤差逆伝搬法 [1] はその簡単さゆえに、よく用いられる手法である。しかし、この方法に対して従来提案されてきた高速化手法は学習係数等のパラメータを実験的に求めなければならないなどヒューリスティックな方法がほとんどである。最近、著者らは非線形不等式に対する反復法に基づいた新しい学習アルゴリズム逐次射影法 (successive projection method) を提案した [2, 3, 4]。この学習アルゴリズムでは上記の学習係数を学習の進行にともなって適応的に与えることが可能であり、逐次型誤差逆伝搬法に対する高速化手法とみなすことができる。本報告では、[4] で提案した方法で現れる部分問題をより厳密に解くことを考える。そのため、使用する階層型ニューラルネットワークに若干の修正を加え、非線形性が強いいため扱いが困難であった部分問題をより扱いやすい形に再定式化することを提案する。

2 階層型ニューラルネットワーク

一般に、階層型ニューラルネットワークとは神経細胞を模した多入力1出力のユニットを階層状に構成したものであり、ユニット i の出力は、

$$y_i = f \left(\sum_j w_{ji} y_j - \theta_i \right),$$

y_i : ユニット i の出力, θ_i : 閾値

w_{ji} : ユニット j からユニット i への結合荷重

に従う。ここで出力関数 $f(x)$ は

$$f(x) = \frac{1 - \exp(-\lambda x)}{1 + \exp(-\lambda x)} \quad (\lambda \geq 0 \text{ は定数}) \quad (1)$$

で定義されるシグモイド関数とする。以下、本稿では、3層型ニューラルネットワークを考えることにする。各層のユニット数は入力層 L , 中間層 M , 出力層 N とする。ニューラルネットワークへの入力を $\mathbf{x} \in R^L$ と表し、中間層からの出力を $\mathbf{y} \in R^M$, 出力層からの出力を $\mathbf{z} \in R^N$ と表す。入力層, 中間層間の結合荷重を $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_M\}$ で、中間層, 出力層間の結合荷重を $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N\}$ で表すことにする。た

だし、ここで \mathbf{u}_i は入力層と中間層ユニット i 間の結合荷重を表すベクトル $\mathbf{u}_i = (u_{1i}, u_{2i}, \dots, u_{Li})^T$ であり、 \mathbf{v}_j は中間層と出力層ユニット j 間の結合荷重を表すベクトル $\mathbf{v}_j = (v_{1j}, v_{2j}, \dots, v_{Mj})^T$ である。また、閾値 θ_i は常に -1 の入力にかかる重みと考え、結合荷重と同一視することにし、全結合荷重はベクトル $\mathbf{w} = (\mathbf{u}_1^T, \mathbf{u}_2^T, \dots, \mathbf{u}_M^T, \mathbf{v}_1^T, \dots, \mathbf{v}_N^T)^T$ で表す。

ニューラルネットワークの学習とは与えられた P 個の入力 $\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^P \in R^L$ の各々に対して望ましい出力 $\mathbf{t}^1, \mathbf{t}^2, \dots, \mathbf{t}^P \in R^N$ を実現するような結合荷重 \mathbf{w} を求めることである。以下では、結合荷重が \mathbf{w} であるとき、入力 \mathbf{x} に対するネットワークの出力を表すベクトル関数を $\mathbf{z}(\mathbf{w}, \mathbf{x})$ と書く。さらに与えられた各入力 \mathbf{x}^p に対する出力を $\mathbf{z}^p(\mathbf{w}) = \mathbf{z}(\mathbf{w}, \mathbf{x}^p)$ と表す。

3 逐次射影法

ここでは、パターン分離問題にニューラルネットワークを適用する場合を考える。このとき一般に、望ましい出力 \mathbf{t}^p の各要素は -1 または 1 であり、学習は $-1.0 < T_l < T_u < 1.0$ であるような適当な定数 (たとえば $T_l = -0.9, T_u = 0.9$) を用いて、非線形連立不等式

$$\begin{cases} z_i^p(\mathbf{w}) \geq T_u, & i \in I^p, \\ z_i^p(\mathbf{w}) \leq T_l, & i \in J^p. \end{cases} \quad p \in \{1, \dots, P\} \quad (2)$$

を満たす \mathbf{w} を求めることと考えられる。ただし、

$$I^p = \{i: t_i^p = 1\}, \quad J^p = \{i: t_i^p = -1\}.$$

である。ここで、

$$\bar{z}_i^p(\mathbf{w}) = \begin{cases} -z_i^p(\mathbf{w}), & i \in I^p, \\ z_i^p(\mathbf{w}), & i \in J^p, \end{cases} \quad \bar{t}_i^p = \begin{cases} -T_u, & i \in I^p, \\ T_l, & i \in J^p. \end{cases}$$

を要素とするベクトル関数 $\bar{\mathbf{z}}^p(\mathbf{w})$ および定数ベクトル $\bar{\mathbf{t}}^p$ を定義すると不等式 (2) は

$$\bar{\mathbf{z}}^p(\mathbf{w}) \leq \bar{\mathbf{t}}^p, \quad p \in \{1, \dots, P\}. \quad (3)$$

と変形できる。問題 (3) を解くために、以下の逐次射影法を適用する。

逐次射影法 (SPM)

Step 1 結合荷重の初期値 w^0 を選び, $l := 0, p := 1$ とする.

Step 2 パターン p に関する (3) 式を現在の結合荷重 w^l において線形近似した制約条件をもつ, 次の2次計画部分問題の解を求め, w^{l+1} とする.

$$\begin{aligned} \min \quad & \frac{1}{2} \|w - w^l\|^2 \\ \text{s.t.} \quad & \hat{z}^p(w^l) + \nabla \hat{z}^p(w^l)^\top (w - w^l) \leq \hat{t}^p. \end{aligned} \quad (4)$$

Step 3 終了条件が満たされなければ, $p := p + 1 \pmod{P}$, $l := l + 1$ として **Step 1** に戻る.

問題 (4) の最適解 w^{l+1} は, その実行可能性領域に w^l を射影した点となっている. この凸2次計画問題は, 双対法やレムケ法などの最適化のアルゴリズムを適用して解くことができる. しかし, (1) 式で定義されたシグモイド関数 f は, その値 $f(x)$ が -1 または 1 に近づくにつれ微分係数が 0 に近づいていく. このとき $\nabla \hat{z}^p(w^l)$ の要素も 0 に近づいて行き, 問題を扱う上で困難が生じる. そこで, 中間層, 出力層の各々について, 次節で述べるような修正を加える.

4 ネットワークの修正

(1) 式のシグモイド関数 f は狭義単調増加関数であるから, (3) 式の連立不等式は次のように書きかえることができる.

$$v_i^\top \begin{pmatrix} f(u_1^\top x^p) \\ \vdots \\ f(u_M^\top x^p) \end{pmatrix} \leq f^{-1}(\hat{t}_i^p), \quad \begin{matrix} p \in \{1, \dots, P\}, \\ i \in \{1, \dots, N\}. \end{matrix} \quad (5)$$

このとき, 出力層の各ユニットの出力関数の微分係数の値が 0 に近づくことによる困難を回避できる.

つぎに, 中間層のユニットの出力関数を以下のような微分可能な関数で置きかえる.

$$h(x) = \begin{cases} f(X_u) + f'(X_u)(x - X_u), & (x \geq X_u) \\ f(x), & (X_l < x < X_u) \\ f(X_l) + f'(X_l)(x - X_l), & (x \leq X_l) \end{cases}$$

この関数 $h(x)$ は, シグモイド関数において x の値が X_u 以上および X_l 以下の部分を線形関数で近似したものである (図 1). さらに, この関数の出力値を $-1 \leq h(x) \leq 1$ に抑えるために各々の2次計画部分問題に対して以下の制約条件を加える.

$$h^{-1}(-1) \leq u_i^\top x^p \leq h^{-1}(1), \quad i \in \{1, \dots, M\}.$$

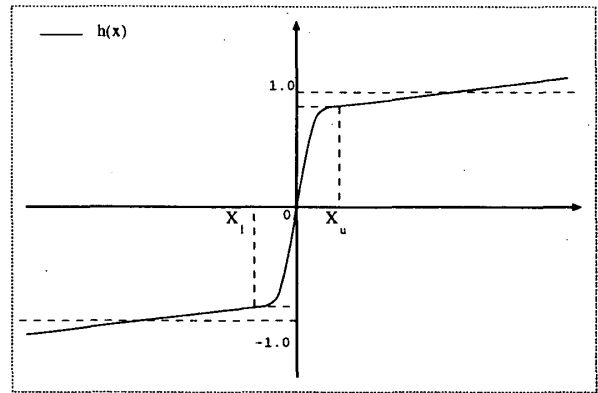


図 1: $h(x)$

さらに, $\hat{z}^p(w)$ と \hat{t}^p の各要素を

$$\hat{z}_i^p(w) = v_i^\top \begin{pmatrix} h(u_1^\top x^p) \\ \vdots \\ h(u_M^\top x^p) \end{pmatrix}, \quad \hat{t}_i^p = f^{-1}(\hat{t}_i^p)$$

と定義すると, (4) に対応する2次計画部分問題は

$$\begin{aligned} \min \quad & \frac{1}{2} \|w - w^l\|^2 \\ \text{s.t.} \quad & \hat{z}^p(w^l) + \nabla \hat{z}^p(w^l)^\top (w - w^l) \leq \hat{t}^p, \\ & h^{-1}(-1) \leq u_i^\top x^p \leq h^{-1}(1), \quad i \in \{1, \dots, M\}. \end{aligned}$$

となる.

5 おわりに

本報告では [3] で提案した方法をさらに拡張し, 部分問題をより厳密に解くことを考えた. 計算機実験の結果は当日発表する.

参考文献

- [1] D.E. Rumelhart, G.E. Hinton, R.J. Williams, "Learning representations by back-propagating errors", *Nature*, Vol. 323, 1986, pp. 533 - 536.
- [2] 茨木, 福島: 最適化の手法, 共立出版, 1993, pp. 181-188.
- [3] 巽, 福島: "逐次射影法としての誤差逆伝搬法," システム制御情報学会論文誌 Vol. 8, 1995, pp. 204-211.
- [4] K. Tatsumi and M. Fukushima, "A Successive Projection Method for Multilayer Neural Networks," Technical Report IS-95016, Nara Institute of Science and Technology, March, 1994.