Journal of Mathematical Research & Exposition
Vol.21, No.2, 219-228, May, 2001

Convergence of On-Line Gradient Methods for Two-Layer Feedforward Neural Networks *

LI Zheng-xue[1], *WU Wei*[2], *ZHANG Hong-wei*[2]

(1. Dept. of Math., Jilin University, Changchun 130023, China;
2. Dept. of Math., Dalian University of Technology, Dalian 116024, China)

Abstract: A discussion is given on the convergence of the on-line gradient methods for two-layer feedforward neural networks in general cases. The theories are applied to some usual activation functions and energy functions.

Key words: on-line gradient method; feedforward neural network; convergence.

Classification: AMS(1991) 68T01/CLC TP183

Document code: A **Article ID:** 1000-341X(2001)02-0219-10

1. Introduction

The on-line gradient method (OGM) originates from the steepest descent method. It is similar to the Gauss-Seidel method for solving linear equations in that the proceeding results are applied immediately in the computation of the present step in a learning procedure. Due to its rapidity, economy and high efficiency, it has been in the good graces of engineering community and has found a wide application in the computation problems of neural networks. However, in the nonlinear case, not very much is known about the convergence of OGM for a given set of finite training examples. Convergence theorems of OGM are given in [1] for a special error function, the square error function. This paper generalizes the results of [1] to more general cases and obtains corresponding convergence results.

This paper is arranged as follows. In Section 2 we introduce some preliminary knowledge, pose some lemmas, and prove three theorems based on the lemmas. These results are essential to prove the convergence theorems. A weak convergence theorem, a strong convergence theorem and a theorem on the rate of convergence are given in Section 3. In Section 4 the relationship between the assumptions in this paper and those in [1] is discussed and some usual activation functions and energy functions are listed.

*Received date: 1999-12-27
Foundation item: Supported by the Natural Science Foundation of China
Biography: LI Zheng-xue (1962-), male, born in Heilongjiang province, associate professor.
E-mail: Lzxue@student.dlut.edu.cn

2. Preliminary Theorems

What we will discuss are two-layer feedforward neural networks whose structure is n - 1 (see Fig.1). It can be used to solve simple classification problems. We assume that a given set of training examples is $\{\xi^j, O^j\}_{j=1}^J \subset R^n \times R^1$($O^j$ is the desired output), weight vector is $\omega = (\omega_1, \omega_2, \cdots, \omega_n)^T \in R^n$ and the activation function is $g : R^1 \to I_0$ (I_0 is a finite interval). The expression of energy function $E : R^n \to [0, \infty)$ is

$$E(\omega) = \sum_{j=1}^{J} F_j \left[g(\omega \cdot \xi^j)\right] = \sum_{j=1}^{J} E_j \left(\omega \cdot \xi^j\right), \tag{2.1}$$

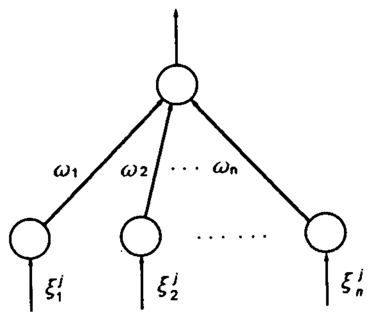

Fig .1 Two-layer feedforward neural networks whose structure is n-1

where we have introduced the function E_j to simplify the expression and the deduction. For given functions E and g, our task is to obtain weight vector ω^* through network learning such that

$$E(\omega^*) = \inf_{\omega \in R^n} E(\omega). \tag{2.2}$$

The gradient of the function $E(\omega)$ is

$$E' = \sum_{j=1}^{J} E_j'(\omega \cdot \xi^j)\xi^j. \tag{2.3}$$

In the steepest descent method or the gradient method , the present weight value ω is revised by

$$\triangle\omega = -\eta E' = \sum_{j=1}^{J}(-\eta)E_j'(\omega \cdot \xi^j)\xi^j, \tag{2.4}$$

where $\eta > 0$ is the learning step size. In the on-line gradient method, we choose

$$\triangle_j\omega = -\eta E_j'(\omega \cdot \xi^j)\xi^j. \tag{2.5}$$

For the convenience of discussion, we train the network by selecting the training examples in a fixed order. Then the learning algorithm of OGM can be express to be

$$\omega^{mJ+j} = \omega^{mJ+j-1} + \triangle_j\omega^{mJ+j-1},\ j = 1, 2, \cdots, J;\ m = 0, 1, \cdots. \tag{2.6}$$

From (2.3) and (2.5), the following relationship can be obtained

$$\sum_{j=1}^{J} \triangle_j\omega = -\eta E'(\omega). \tag{2.7}$$

Defining

$$r_{j,m} = \triangle_j \omega^{mJ+j-1} - \triangle_j \omega^{mJ},\ j = 1, 2, \cdots, J;\ m = 0, 1, \cdots.$$

Specifically, we have that

$$r_{1,m} = 0, \quad m = 0, 1, \cdots.$$

Several assumptions used in the paper are given below:

(B1) The input example vectors $\{\xi_j\}_{j=1}^J$ are J linearly independent vectors in R^n.

(B2) The second derivatives of functions $E_j : R^1 \to [0, \infty)$ exist, and $\left|E_j'\right|$ and $\left|E_j''\right|$ $(1 \le j \le J)$ are uniformly bounded.

(B3) The Hessian matrix $H(\omega) = E'' = (\frac{\partial^2 E}{\partial \omega_i \partial \omega_j})_{n\times n}$ is a positive definite matrix, i.e., there exist constants C_1 and C_2 $(C_2 \ge C_1 > 0)$ such that for arbitrary vectors y and ω in R^n, there holds

$$C_1 \|y\|^2 \le y^T H(\omega) y \le C_2 \|y\|^2,$$

where $\|\cdot\|$ is the Euclidean norm in R^n.

Remark 1 Only conditions (B1)–(B2) are needed to deduce the convergence of the iterative sequence $\{\omega^i\}$ of weights (see Theorem 3.1 later). For most nonlinear activation functions, this is the best result we can expect. Condition (B3) is more stronger. Theorem 3.2 indicates that we can obtain a strong convergence of OGM if (B3) is satisfied, similarly as the standard gradient method for the minimum problems of quadratic functionals.

Lemma 2.1 *Suppose that (B1) is satisfied. Then there exists a positive constant* $\alpha \in (0, \frac{1}{J}]$ *such that for arbitrary* $\delta = (\delta_1, \delta_2, \cdots, \delta_J)^T \in R^J$, *there holds that*

$$\|\sum_{j=1}^J \delta_j \xi^j\|^2 \ge \alpha \sum_{j=1}^J \|\delta_j \xi^j\|^2. \tag{2.8}$$

Remark 2 The proofs of Lemma 2.1 and some theorems below are omitted because they can be proved easily or be found in [1].

Lemma 2.2 *For* $j = 1, 2, \cdots, J;\ m = 0, 1, \cdots$, *there holds*

$$\omega^{mJ+j} = \omega^{mJ} + \sum_{k=1}^j (\triangle_k \omega^{mJ} + r_{k,m}). \tag{2.9}$$

Lemma 2.3 *Suppose that (B2) is satisfied. Then there exists a constant* $C > 0$ *such that for* $m = 0, 1, \cdots$, *there holds*

$$\sum_{j=1}^J \|r_{j,m}\| \le C\eta \sum_{j=1}^J \|\triangle_j \omega^{mJ}\|. \tag{2.10}$$

Lemma 2.4 *Let (B2) be satisfied and let* $\omega_d^m = \omega^{(m+1)J} - \omega^{mJ}$, $m = 0, 1, \cdots$, *then there exists a constant* $C > 0$ *such that*

$$\|\omega_d^m\| \le C \sum_{j=1}^J \|\triangle_j \omega^{mJ}\|. \tag{2.11}$$

Lemma 2.5 *By the Cauchy-Schwarz inequality, we have that*

$$(\sum_{j=1}^{J}\|\triangle_j\omega^{mJ}\|)^2 \le J\sum_{j=1}^{J}\|\triangle_j\omega^{mJ}\|^2,\ m=0,1,\cdots. \tag{2.12}$$

Theorem 2.1 *Suppose that (B2) is satisfied, then there exists a positive constant γ independent of η, j and m such that for $j=1,2,\cdots,J$; $m=0,1,\cdots$, there holds*

$$E(\omega^{(m+1)J}) \le E(\omega^{mJ}) - \frac{1}{\eta}\|\sum_{j=1}^{J}\triangle_j\omega^{mJ}\|^2 + \gamma\sum_{j=1}^{J}\|\triangle_j\omega^{mJ}\|^2. \tag{2.13}$$

Proof Set

$$M_0 = \max_{1\le j\le J}\|\xi^j\|,\ M_k = \max_{1\le j\le J}\sup_{x\in R^1}|E_j^{(k)}(x)|,\ k=1,2. \tag{2.14}$$

By the Taylor expansion we derive

$$E_j(\omega^{(m+1)J}\cdot\xi^j) = E_j(\omega^{mJ}\cdot\xi^j) + E_j'(\omega^{mJ}\cdot\xi^j)\,\omega_d^m\cdot\xi^j + \delta_{j,m}, \tag{2.15}$$

where $\delta_{j,m} = \frac{1}{2}E_j''(t)(\omega_d^m\cdot\xi^j)^2$, $t=(1-\theta)(\omega^{mJ}\cdot\xi^j)+\theta(\omega^{(m+1)J}\cdot\xi^j), 0<\theta<1$.Noting (2.5) and summing both sides of (2.15) from 1 to J, we have

$$\sum_{j=1}^{J}E_j(\omega^{(m+1)J}\cdot\xi^j) = \sum_{j=1}^{J}E_j(\omega^{mJ}\cdot\xi^j) - \frac{1}{\eta}(\sum_{j=1}^{J}\triangle_j\omega^{mJ})\cdot\omega_d^m + \sum_{j=1}^{J}\delta_{j,m}.$$

Also, by (2.1) and (2.9), we can obtain

$$E(\omega^{(m+1)J}) = E(\omega^{mJ}) - \frac{1}{\eta}\|\sum_{j=1}^{J}\triangle_j\omega^{mJ}\|^2 - \frac{1}{\eta}(\sum_{j=1}^{J}\triangle_j\omega^{mJ})\cdot(\sum_{j=1}^{J}r_{j,m}) + \sum_{j=1}^{J}\delta_{j,m}.$$

Using (2.10) and (2.12), we have that

$$\begin{aligned}\frac{1}{\eta}|(\sum_{j=1}^{J}\triangle_j\omega^{mJ})\cdot(\sum_{j=1}^{J}r_{j,m})| &\le \frac{1}{\eta}\sum_{j=1}^{J}\|\triangle_j\omega^{mJ}\|(C_1\eta\sum_{j=1}^{J}\|\triangle_j\omega^{mJ}\|)\\ &\le C_1J\sum_{j=1}^{J}\|\triangle_j\omega^{mJ}\|^2.\end{aligned}$$

From (2.14), (2.11) and (2.12), we get

$$\sum_{j=1}^{J}|\delta_{j,m}| \le \frac{1}{2}\sum_{j=1}^{J}|E_j''(t)|\|\omega_d^m\|^2\|\xi^j\|^2 \le \frac{1}{2}C_2^2M_0^2M_2J^2\sum_{j=1}^{J}\|\triangle_j\omega^{mJ}\|^2.$$

Finally (2.13) is proved by choosing $\gamma = (C_1+\frac{1}{2}C_2^2M_0^2M_2J)J$. □

Theorem 2.2 *Suppose that $(B1)$ and $(B2)$ are satisfied, that positive constants α and γ are defined in Lemma 2.1 and Theorem 2.1 respectively, and that the step size η satisfies*

$$0 < \eta < \frac{\alpha}{\gamma}, \tag{2.16}$$

then for $m = 0, 1, \cdots$, there holds

$$E(\omega^{(m+1)J}) \leq E(\omega^{mJ}) - \beta \sum_{j=1}^{J} \|\triangle_j \omega^{mJ}\|^2, \tag{2.17}$$

where the positive constant β is defined by

$$\beta = \frac{\alpha}{\eta} - \gamma. \tag{2.18}$$

Moreover, there also holds

$$\sum_{m=0}^{\infty} \sum_{j=1}^{J} \|\triangle_j \omega^{mJ}\|^2 < \infty, \tag{2.19}$$

$$\lim_{m\to\infty} \sum_{j=1}^{J} \|\triangle_j \omega^{mJ}\| = 0, \tag{2.20}$$

$$\lim_{m\to\infty} \|\omega^{mJ+j} - \omega^{mJ}\| = 0, \quad j = 1, 2, \cdots, J. \tag{2.21}$$

Proof Let $\delta_j = -\eta E_j'(\omega^{mJ} \cdot \xi^j)$, then from (2.5) we have

$$\triangle_j \omega^{mJ} = \delta_j \xi^j.$$

Combining with (2.8) gives

$$\|\sum_{j=1}^{J} \triangle_j \omega^{mJ}\|^2 \geq \alpha \sum_{j=1}^{J} \|\triangle_j \omega^{mJ}\|^2. \tag{2.22}$$

Again, making use of (2.18), we derive

$$-\frac{1}{\eta}\|\sum_{j=1}^{J} \triangle_j \omega^{mJ}\|^2 + \gamma \sum_{j=1}^{J} \|\triangle_j \omega^{mJ}\|^2 \leq -\beta \sum_{j=1}^{J} \|\triangle_j \omega^{mJ}\|^2.$$

(2.17) can be obtained by substituting the above expression into (2.13).

From (2.17), there holds

$$E(\omega^{(M+1)J}) \leq E(\omega^0) - \beta \sum_{m=0}^{M} \sum_{j=1}^{J} \left\|\triangle_j \omega^{mJ}\right\|^2.$$

Since $E(\omega^{(M+1)J}) \geq 0$ for any nonnegative integer M, we have

$$\sum_{m=0}^{M} \sum_{j=1}^{J} \|\triangle_j \omega^{mJ}\|^2 \leq \frac{1}{\beta} E(\omega^0),$$

so (2.19) is also satisfied.

(2.20) can be easily proved from (2.19), (2.12) and the necessary condition for the convergence of number series.

It follows from (2.9) and (2.10) that

$$\|\omega^{mJ+j} - \omega^{mJ}\| \leq C \sum_{j=1}^{J} \|\triangle_j \omega^{mJ}\|, \ j = 1, 2, \cdots, J,$$

and (2.21) is thus satisfied. This completes the proof. □

Theorem 2.3 *Suppose that (B1) and (B2) are satisfied. Then there exist positive constants μ and ν such that for $m = 0, 1, \cdots$, there holds*

$$\sum_{j=1}^{J} \|r_{j,m}\| \leq \mu\eta^2 \|E'(\omega^{mJ})\| \text{ and } \|\omega_d^m\| \leq \nu\eta \|E'(\omega^{mJ})\|.$$

Proof It follows from (2.22) and (2.7) that

$$\sum_{j=1}^{J} \|\triangle_j \omega^{mJ}\|^2 \leq C_1 \|\sum_{j=1}^{J} \triangle_j \omega^{mJ}\|^2 = C_1 \eta^2 \|E'(\omega^{mJ})\|^2. \tag{2.23}$$

By (2.10), (2.12) and (2.23), we conclude that

$$\begin{aligned} \sum_{j=1}^{J} \|r_{j,m}\| &\leq C_2 \eta [(\sum_{j=1}^{J} \|\triangle_j \omega^{mJ}\|)^2]^{\frac{1}{2}} \leq C_2 \eta (J \sum_{j=1}^{J} \|\triangle_j \omega^{mJ}\|^2)^{\frac{1}{2}} \\ &\leq C_2 J^{\frac{1}{2}} \eta (C_1 \eta^2 \|E'(\omega^{mJ})\|^2)^{\frac{1}{2}} = \mu \eta^2 \|E'(\omega^{mJ})\|, \end{aligned}$$

where $\mu = C_1^{\frac{1}{2}} C_2 J^{\frac{1}{2}}$. Finally, from (2.11), (2.12) and (2.23) we get

$$\begin{aligned} \|\omega_d^m\| &\leq C_3 \sum_{j=1}^{J} \|\triangle_j \omega^{mJ}\| \leq C_3 J^{\frac{1}{2}} (\sum_{j=1}^{J} \|\triangle_j \omega^{mJ}\|^2)^{\frac{1}{2}} \\ &\leq C_3 J^{\frac{1}{2}} (C_1 \eta^2 \|E'(\omega^{mJ})\|^2)^{\frac{1}{2}} = \nu\eta \|E'(\omega^{mJ})\|, \end{aligned}$$

where $\nu = C_1^{\frac{1}{2}} C_3 J^{\frac{1}{2}}$. This completes the proof. □

3. Convergence Theorems

Based on the above preliminaries, it is not difficult to obtain the following convergence theorems whose proofs are the same as the corresponding results in [1].

Theorem 3.1 (weak convergence theorem) *Suppose that (B1), (B2), (2.16) and (2.18) are satisfied and the sequence $\{\omega^i\}$ is generated from the learning algorithm (2.6) of OGM, then there exists a constant $E^* \geq 0$ such that*

$$\lim_{i\to\infty} E(\omega^i) = E^* \ \textit{and} \ \lim_{i\to\infty} \|E'(\omega^i)\| = 0.$$

Moreover, if $\bar{\omega}$ is a limit point of the sequence $\{\omega^i\}$, then $E'(\bar{\omega}) = 0$.

Theorem 3.2 (strong convergence theorem) *Suppose that (B1)–(B3), (2.16) and (2.18) are satisfied and the sequence $\{\omega^i\}$ is generated from the learning algorithm (2.6) of OGM, then there exists a unique minimum point $\omega^* \in R^n$ such that*

$$E(\omega^*) = \inf_{\omega \in R^n} E(\omega) \quad and \quad \lim_{i \to \infty} \|\omega^i - \omega^*\| = 0.$$

In order to present Theorem 3.3 for the estimation of the convergence rate, we require the step size

$$\eta < \min\{\frac{\alpha}{\gamma}, \frac{2}{2\mu + \nu^2 B_2}, \frac{B_2}{2B_1^2}\}, \tag{3.1}$$

where α and γ are the constants that appear in Lemma 2.1 and Theorem 2.1 respectively, and constants μ and ν are chosen according to Theorem 2.3. Let

$$\lambda = 1 - (\mu + \frac{1}{2}\nu^2 B_2)\eta. \tag{3.2}$$

Then we have

$$0 < \lambda < 1. \tag{3.3}$$

Combining (3.3) with (3.1) gives

$$\eta < \frac{B_2}{2B_1^2} < \frac{B_2}{2B_1^2 \lambda}.$$

Write

$$q = 1 - \frac{2B_1^2 \lambda}{B_2}\eta, \tag{3.4}$$

then there holds that

$$0 < q < 1.$$

Theorem 3.3 *Suppose (B1)–(B3) and (3.1)-(3.4) are satisfied, ω^* is the unique minimum point of function $E(\omega)$ in R^n and the sequence $\{\omega^i\}$ is generated from the learning algorithm (2.6) of OGM, then for $j = 0, 1, \cdots, J-1$; $m = 0, 1, \cdots$, there holds that*

$$|E(\omega^{mJ+j}) - E(\omega^*)| \le q^m |E(\omega^j) - E(\omega^*)|,$$

$$\|\omega^{mJ+j} - \omega^*\| \le (\frac{B_2}{B_1})^{\frac{1}{2}} q^{\frac{m}{2}} \|\omega^j - \omega^*\|.$$

4. Applications of Convergence Theorems

This paper is based on the work of [1] as mentioned above. So in this section we first have a look at the relationship between the assumptions in the two papers. [1]

has given four assumptions (A1)–(A4), in which conditions (A1) and (A4) are the same with (B1) and (B3) in this paper respectively. Condition (A2) requires that there holds $\omega^i \in \Omega$ $(i = 0, 1, \cdots)$ for an arbitrary initial weight vector $\omega^0 \in \Omega$, where Ω is a region in R^n and the meaning of the sequence $\{\omega^i\}$ is the same as that in Theorem 3.1. Condition ($A3$) requires that the functions $|g^{(k)}(\omega \cdot \xi^j)|$ $(k = 0, 1, 2;\ 1 \le j \le J)$ are uniformly bounded for arbitrary $\omega \in \Omega$. When a network is used for classification problems, Ω becomes R^n(cf. the discussion after condition (A3) in [1]). Then conditions (A2) and (A3) virtually demand that the functions $|g^{(k)}(x)|$ $(k = 0, 1, 2)$ be uniformly bounded for arbitrary $x \in R^1$. Hence, to analyze the relationship between these two conditions and (B2) in the paper becomes the key problem. To this end we have (the proof is omitted)

Proposition 4.1 *If conditions (A2) and (A3) are satisfied with $\Omega = R^n$ and the functions $E_j(x) = \frac{1}{2}[O^j - g(x)]^2 (1 \le j \le J)$, then condition (B2) is satisfied.*

Compared with (A2) and (A3), condition (B2) is suitable not only for more error functions, but also for some unbounded activation functions.

Proposition 4.2 *Suppose that $g(x) = (x^2 + \beta^2)^{\frac{1}{2}} (\beta > 0), E_j(x) = \ln\{1 + [O^j - g(x)]^2\}$ $(1 \le j \le J)$, then the function $E_j(x)$ $(1 \le j \le J)$ satisfies (B2).*

The proof to the above proposition is straightforward and thus is omitted.

For the convenience of applications, we list below some usual activation functions and energy functions (the activation function $g(x)$ in the expression of the energy function $E(\omega)$ can be chosen from the following three types). It can be proved that the suitably chosen functions $E_j(x)$ $(1 \le j \le J)$ will satisfy condition (B2) (where usually the activation function (4.7) is only used for (4.11) and (4.12), and the activation function $g(x)$ in energy functions (4.10) and (4.15) is usually chosen from (4.2), while the others can be combined arbitrarily.)

Activation functions

(i) Sigmoid functions

$$g(x) = \frac{1}{1 + e^{-x}}, \tag{4.1}$$

$$g(x) = \tanh(x), \tag{4.2}$$

$$g(x) = \frac{2}{\pi} \arctan(x), \tag{4.3}$$

$$g(x) = erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du. \tag{4.4}$$

(ii) Radial basis functions

$$g(x) = e^{-\frac{x^2}{\sigma^2}}, \quad \sigma > 0, \tag{4.5}$$

$$g(x) = (x^2 + \beta^2)^{-\frac{1}{2}}, \quad \beta > 0, \tag{4.6}$$

$$g(x) = (x^2 + \beta^2)^{\frac{1}{2}}, \quad \beta > 0. \tag{4.7}$$

(iii) Segment function

$$g(x) = \begin{cases} 0, & x \leq -2, \\ (2+x)^3, & -2 < x \leq -1, \\ (2+x)^3 - 4(1+x)^3, & -1 < x \leq 0, \\ (2-x)^3 - 4(1-x)^3, & 0 < x \leq 1, \\ (2-x)^3, & 1 < x \leq 2, \\ 0, & x > 2. \end{cases} \tag{4.8}$$

Energy function

(i) Square error function

$$E = \frac{1}{2}\sum_{j=1}^{J}[O^j - g(\omega \cdot \xi^j)]^2. \tag{4.9}$$

(ii) Measure of the cross entropy

$$E = \sum_{j=1}^{J}[\frac{1}{2}(1+O^j)\ln\frac{1+O^j}{1+g(\omega \cdot \xi^j)} + \frac{1}{2}(1-O^j)\ln\frac{1-O^j}{1-g(\omega \cdot \xi^j)}]. \tag{4.10}$$

(iii) Cauchy error function

$$E = \sum_{j=1}^{J}\ln\{1 + [O^j - g(\omega \cdot \xi^j)]^2\}. \tag{4.11}$$

(iv) Logistical error function

$$E = \frac{\beta}{\alpha}\sum_{j=1}^{J}\ln\{\cosh[\alpha(O^j - g(\omega \cdot \xi^j))]\},\ \alpha > 0, \beta > 0. \tag{4.12}$$

(v) Generalized error function

$$E = \sum_{j=1}^{J}\{O^j[O^j - g(\omega \cdot \xi^j)] + \frac{\lambda}{2}[g^2(\omega \cdot \xi^j) - (O^j)^2]\},\ 0 \leq \lambda \leq 1. \tag{4.13}$$

(vi) Linear combination error functions

$$E = \frac{1}{4}\sum_{j=1}^{J}\{[O^j - g(\omega \cdot \xi^j)]^2 + A_0[1 - \frac{g(\omega \cdot \xi^j)}{O^j + \varepsilon^j}]^2\},\ A_0 > 0. \tag{4.14}$$

$$E = \sum_{j=1}^{J}\{\frac{1}{2}[O^j - g(\omega \cdot \xi^j)]^2 + \frac{1}{20}[(1+O^j)\ln\frac{1+O^j}{1+g(\omega \cdot \xi^j)} + (1-O^j)\ln\frac{1-O^j}{1-g(\omega \cdot \xi^j)}]\}. \tag{4.15}$$

References:

[1] WU W & XU Y S. *Convergence of the On-Line Gradient Method for Neural Networks* [M]. Proceedings to the Sixth Annual National Meeting of Computational Mathematics, 1999, 165.

[2] PAN D. *Research on the activity function in ANN and its application* [J]. Journal of South China University of Technology, 1998, **26**: 49–54.

[3] LU J G. *BP algorithm used in multilayer neural network* [J]. Computer Engineering, 1994, **20**: 17–19.

[4] XU B Z. *Theory and Application of Neural Networks* [M]. South China University of Technology Press, Guangzhou, 1994.

二层前传神经网络中在线梯度法的收敛性

李 正 学[1], 吴 微[2], 张 宏 伟[2]

(1. 吉林大学数学系，长春 130023; 2. 大连理工大学应用数学系， 116024)

摘 要: 本文给出了一般情况下二层前传神经网络中的在线梯度法的收敛性定理，并将其应用于一些常用的活化函数和能量函数。