$$E(H_2(t))^2 \le \frac{C}{n\delta} + C\delta\sum_{k=0}^{m_t-1} E(X_n(\tau_t + k\delta) -$$

$$-X(\tau_t + k\delta))^2 + C(K(n,h_n) - (1-2\theta))^2 + \frac{C\delta^2}{h_n}$$

Из леммы 5 вытекает: $E(H_3(t))^2 \le Ch_n + C\delta + \dfrac{C}{n}$.

Окончательно, из оценок $H_0(t)$, $H_1(t)$, $H_2(t)$, $H_3(t)$ получаем следующее неравенство

$$E(X_n(t) - X(t))^2 \le \frac{C\delta^2}{h_n} + CE(X^0{}_n(\tau_t) - x)^2 +$$

$$+\frac{C}{n^4 h_n \delta^2} + \frac{C}{n\delta} + C\delta\sum_{k=0}^{m_t-1} E(X_n(\tau_t + k\delta) - X(\tau_t + k\delta))^2 +$$

$$+Ch_n + \frac{C}{n} + C(K(n,h_n) - (1-2\theta))^2$$

Пользуясь дискретным аналогом неравенства Гронвалла получаем:

$$E(X_n(t) - X(t))^2 \le \frac{C\delta^2}{h_n} + C\delta +$$

$$+CE(X^0{}_n(\tau_t) - x)^2 + \frac{C}{n^4 h_n \delta^2} + \frac{C}{n\delta} +$$

$$+Ch_n + \frac{C}{n} + C(K(n,h_n) - (1-2\theta))^2$$

Если $1/n^2 < h_n < 1/n^{1/2}$, то, положив $\delta = h_n^{1/3}/n^{1/3}$ получим первое неравенство теоремы, а если $h_n \ge 1/n^{1/2}$, тогда $\dfrac{1}{nh_n} \le h_n$ и, взяв $\delta = h_n$, имеем второе неравенство.

Доказательство следствия 1. Получается предельным переходом в неравенстве из условия теоремы 1.

Доказательство теоремы 2. Достаточность доказана в следствии 1. Докажем необходимость. Пусть $X_n(t)$ сходится к некоторому процессу $X(t)$ в требуемом смысле, и пусть $Y_n(t)$ – решение уравнения (4) для $\theta = (1 - K(n,h_n))/2$, то есть, воспользовавшись эквивалентной записью уравнения (4), $Y_n(t)$ – решение уравнения

$$Y_n(t) = x + (I)\int_0^t f(s, Y_n(s))dB(s) +$$

$$+\frac{1-K(n,h_n)}{2}\int_0^t (ff_Y')(s, Y_n(s))ds + \int_0^t g(s, Y_n(s))ds, \quad (48)$$

Из теоремы 1 вытекает справедливость неравенства

$$\sup_{t\in T} E[X_n(t) - Y_n(t)]^2 \le$$

$$\le C \sup_{t\in[0,h_n)} E[X_n^0(t) - x]^2 + \frac{C}{n} + Ch_n + \frac{C}{n^4 h_n^3}$$

Перейдя в этом неравенстве к пределу при $n \to \infty$, $h_n \to 0$ так, что $1/n^2 = o(h_n)$ получим $\sup_{t\in T} E[X_n(t) - Y_n(t)]^2 \to 0$. Таким образом, $Y_n(t)$ сходится к $X(t)$. Поэтому и правая часть равенства (48) сходится. Отсюда вытекает, что и последовательность коэффициентов $(1-K(n, h_n))/2$ сходится, что и влечет за собой сходимость последовательности $K(n, h_n)$. Теорема доказана.

Результаты настоящей статьи доложены 08.09.2003г. на международной конференции «AMADE – 2003» [11]

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Itô K. On a stochastic integral equations //Proc. Imp. Acad. Tokyo. – 1946. – Vol. 22. – P. 32 – 36.
2. Стратонович Р.Л. Избранные вопросы теории флюктуаций в радиотехнике. – М.: Советское радио, 1961. – 558 с.
3. Пугачев В.С., Синицын И.Н. Стохастические дифференциальные системы. Анализ и фильтрация. – М.:Наука, 1990. – 630 с.
4. Wong E., Zakai M. On the relationship between ordinary and stochastic differential equations // Internat. J. Engin. Sci. – 1965. – Vol. 3. – P. 213 – 229.
5. Мацкявичюс В. S^p-устойчивость решений симметрических стохастических дифференциальных уравнений // Лит. мат. сб. – 1985. – Т. 25, № 4. – С. 72 – 84.
6. Лазакович Н.В. Стохастические дифференциалы в алгебре обобщенных случайных процессов //Доклады АН Беларуси. – 1994. – Т.38, №5. – С. 23–27.
7. Лазакович Н.В., Яблонский О.Л. О приближении решений одного класса стохастических уравнений // Сиб. матем. журнал. – 2001. – Т. 42, № 1. – С. 87–102.
8. Лазакович Н.В., Юферова И.В. Аппроксимация стохастических интегралов Ито и Стратоновича в алгебре обобщенных случайных процессов //Известия АН Беларуси. Сер. физ.-мат. наук. – 1994. – №2. – С. 28–32.
9. Яблонский О.Л. Классификация способов аппроксимации стохастических интегралов в алгебре обобщенных случайных процессов //Доклады НАН Беларуси. – 2000. – Т.44, №2. – С. 22–26.
10. Егоров Ю.В. К теории обобщенных функций //Успехи математических наук. – 1990. – Т.45, вып.5 (275). – С. 3–40.
11. Русина Т.И., Яблонский О.Л. Неоднородные уравнения в дифференциалах в алгебре обобщенных случайных процессов.// Тезисы докл. междунар. конф. «Аналитические методы анализа и дифференц. уравн.» – Мн.:ИМ НАНБ, 2003. С. 154.

УДК 681.324:519.711.7

Головко В.А., Маньяков Н.В.

МАТРИЧНЫЙ НЕЙРОСЕТЕВОЙ МЕТОД ОБУЧЕНИЯ МНОГОСЛОЙНОЙ СЕТИ С ИСПОЛЬЗОВАНИЕМ АДАПТИВНОГО ШАГА

Основным методом обучения многослойных гетерогенных нейронных сетей прямого распространения без обратных связей является метод обратного распространения ошибки [1], наиболее эффективно представимый в его матричной форме для программной реализации [2]:

Модификация синаптических связей и порогов многослойной гетерогенной нейронной сети производится в соответствии с формулами:

Головко Владимир Адамович*, д.т.н., профессор каф. ЭВМ и С Брестского государственного технического университета. Беларусь, БГТУ, 224017, г. Брест, ул. Московская, 267.*

$$w_{j_{n-1}j_n}^{(n)}(t+1)=w_{j_{n-1}j_n}^{(n)}(t)-\alpha^{(n)}\cdot\frac{1}{L}\cdot\sum_{k=1}^{L}C^{(n)}\cdot M_{j_nj_{n-1}}^{(n)}\cdot Y^{(n-1),k}$$

$$T_{j_n}^{(n)}(t+1)=T_{j_n}^{(n)}(t)-\alpha^{(n)}\cdot\frac{1}{L}\cdot\sum_{k=1}^{L}C^{(n)}\cdot M_{j_n(m_{n-1}+1)}^{(n)}\cdot Y^{(n-1),k},$$

где $C^{(n)}$ вычисляется рекуррентно:

$$C^{(n)}=C^{(n+1)}\cdot W^{(n+1)}\cdot MF_n', \quad C^{(N)}=\varepsilon_N^k\cdot MF_N',$$

$$\varepsilon_N^k=\left(\left(y_1^{(N),k}-t_1^k\right)\ \left(y_2^{(N),k}-t_2^k\right)\ \dots\ \left(y_{m_2}^{(N),k}-t_{m_2}^k\right)\right),$$

а $$MF_n'=\begin{pmatrix} F_n'\left(S_1^{(n),k}\right) & 0 & \dots & 0\\ 0 & F_n'\left(S_2^{(n),k}\right) & \dots & 0\\ \dots & \dots & \dots & \dots\\ 0 & 0 & \dots & F_n'\left(S_{m_n}^{(n),k}\right)\end{pmatrix}$$

– матрица размерности $m_n\times m_n$, а матрица $M_{j_nj_{n-1}}^{(n)}$ размерности $m_n\times(m_{n-1}+1)$ состоит из числа 1 на позиции j_nj_{n-1} и нулей в качестве остальных элементов матрицы.

Изменение синаптических связей и порогов сети производится начиная с последнего N-ого до первого слоя сети.

При таком обучении шаг $\alpha^{(n)}$ может оставаться как постоянным, так и адаптивным. В последнем случае предлагается использовать послойное обучение, являющиеся расширением предложенного для двухслойной сети [3], и основанное на следующих предпосылках. Сначала, после подачи всех элементов обучающего множества, изменяются синаптические связи последнего слоя. Затем, после того как на модифицированную сеть были поданы все элементы обучающего множества, изменяются синаптические связи предшествующего слоя. И так далее, до достижения первого слоя сети. Затем осуществляется переход к началу алгоритма. Процедура обучения останавливается после того, как ошибка сети не превышает заданную после двух последовательных итераций. Шаги обучения $\alpha^{(n)}$ выбираются для наилучшей минимизации ошибки сети при изменении синаптических связей соответствующих слоев.

Теорема. При послойном обучении многослойной гетерогенной нейронной сети прямого распространения без обратных связей адаптивный шаг для каждого слоя определяется в соответствии с формулами:

$$\alpha^{(n)}=\frac{L\cdot\sum\limits_{j_{n-1}=1}^{m_{n-1}}\sum\limits_{j_n=1}^{m_n}\left(\sum\limits_{k=1}^{L}C^{(n)}\cdot M_{j_nj_{n-1}}^{(n)}\cdot Y^{(n-1),k}\right)}{\sum\limits_{j_{n-1}=1}^{m_{n-1}}\sum\limits_{j_n=1}^{m_n}\sum\limits_{l_{n-1}=1}^{m_{n-1}}\sum\limits_{l_n=1}^{m_n}\left(\sum\limits_{k=1}^{L}\left(\left(K_{l_{n-1}l_n}^{(n),k}\right)^T\cdot U^{(n),k}\cdot\left(K_{j_{n-1}j_n}^{(n),k}\right)\right)\right)},$$

где

$$K_{j_{n-1}j_n}^{(n),k}=M_{j_nj_{n-1}}^{(n)}\cdot Y^{(n-1),k},$$

а

$$U^{(n),k}=\left(W^{(n+1)}\cdot MF_n'\right)^T\cdot U^{(n+1),k}\times$$
$$\times\left(W^{(n+1)}\cdot MF_n'\right)+W^{(n+1)}\cdot MF_n''$$

вычисляется рекуррентно с начальным условием

$$U^{(N),k}=\left(MF_N'\right)^2+DE^{(N),k}\cdot MF_n'.$$

Доказательство.

В соответствии с [2] имеем:

$$\frac{\partial E_s^{(k)}}{\partial w_{j_{n-1}j_n}^{(n)}}=\frac{\partial\left(\sum\limits_{i_N=1}^{m_N}\frac{1}{2}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)^2\right)}{\partial w_{j_{n-1}j_n}^{(n)}}=\sum_{i_N=1}^{m_N}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)\cdot F_N'\times$$
$$\times\left(S_{i_N}^{(N),k}\right)\cdot\sum_{i_{N-1}=1}^{m_{N-1}}w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\ldots F_{n+1}'\left(S_{i_{n+1}}^{(n+1),k}\right)\times$$
$$\times\sum_{i_n=1}^{m_n}w_{i_ni_{n+1}}^{(n+1)}\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}=C^{(n)}\cdot M_{j_nj_{n-1}}^{(n)}\cdot Y^{(n-1),k}$$

$$\frac{\partial E_s^{(k)}}{\partial T_{j_n}^{(n)}}=\frac{\partial\left(\sum\limits_{i_N=1}^{m_N}\frac{1}{2}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)^2\right)}{\partial T_{j_n}^{(n)}}=\sum_{i_N=1}^{m_N}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)\times$$
$$\times F_N'\left(S_{i_N}^{(N),k}\right)\sum_{i_{N-1}=1}^{m_{N-1}}w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\ldots\cdot F_{n+1}'\left(S_{i_{n+1}}^{(n+1),k}\right)\times$$
$$\sum_{i_n=1}^{m_n}w_{i_ni_{n+1}}^{(n+1)}\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot(-1)\cdot\delta_{j_n}^{i_n}=C^{(n)}\cdot M_{j_n(m_{n-1}+1)}^{(n)}\cdot Y^{(n-1),k}$$

Найдем частные производные второго порядка от слагаемых функции ошибки:

$$\frac{\partial^2E_s^{(k)}}{\partial w_{j_{n-1}j_n}^{(n)}\partial w_{l_{n-1}l_n}^{(n)}}=\frac{\partial}{\partial w_{l_{n-1}l_n}^{(n)}}\left(\sum_{i_N=1}^{m_N}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)\times\right.$$
$$\times F_N'\left(S_{i_N}^{(N),k}\right)\cdot\sum_{i_{N-1}=1}^{m_{N-1}}w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\ldots\cdot$$
$$\left.\cdot\ldots\cdot F_{n+1}'\left(S_{i_{n+1}}^{(n+1),k}\right)\cdot\sum_{i_n=1}^{m_n}w_{i_ni_{n+1}}^{(n+1)}\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}\right)=$$
$$=\left(\sum_{i_N=1}^{m_N}\frac{\partial\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)}{\partial w_{l_{n-1}l_n}^{(n)}}\cdot F_N'\left(S_{i_N}^{(N),k}\right)\cdot\sum_{i_{N-1}=1}^{m_{N-1}}w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\ldots\cdot\right.$$
$$\cdot\ldots\cdot F_{n+1}'\left(S_{i_{n+1}}^{(n+1),k}\right)\cdot\sum_{i_n=1}^{m_n}w_{i_ni_{n+1}}^{(n+1)}\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}+$$
$$+\sum_{i_N=1}^{m_N}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)\cdot\frac{\partial F_N'\left(S_{i_N}^{(N),k}\right)}{\partial w_{l_{n-1}l_n}^{(n)}}\cdot\sum_{i_{N-1}=1}^{m_{N-1}}w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\ldots\cdot$$
$$\cdot\ldots\cdot F_{n+1}'\left(S_{i_{n+1}}^{(n+1),k}\right)\cdot\sum_{i_n=1}^{m_n}w_{i_ni_{n+1}}^{(n+1)}\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}+$$
$$+\sum_{i_N=1}^{m_N}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)\cdot F_N'\left(S_{i_N}^{(N),k}\right)\cdot\sum_{i_{N-1}=1}^{m_{N-1}}w_{i_{N-1}i_N}^{(N)}\cdot\frac{\partial F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)}{\partial w_{l_{n-1}l_n}^{(n)}}\cdot\ldots\cdot$$
$$\cdot\ldots\cdot F_{n+1}'\left(S_{i_{n+1}}^{(n+1),k}\right)\cdot\sum_{i_n=1}^{m_n}w_{i_ni_{n+1}}^{(n+1)}\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}+$$
$$+\ldots+$$
$$+\sum_{i_N=1}^{m_N}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)\cdot F_N'\left(S_{i_N}^{(N),k}\right)\cdot\sum_{i_{N-1}=1}^{m_{N-1}}w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\ldots\cdot$$
$$\left.\cdot\ldots\cdot F_{n+1}'\left(S_{i_{n+1}}^{(n+1),k}\right)\cdot\sum_{i_n=1}^{m_n}w_{i_ni_{n+1}}^{(n+1)}\cdot\frac{\partial F_n'\left(S_{i_n}^{(n),k}\right)}{\partial w_{l_{n-1}l_n}^{(n)}}\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}\right)=$$
$$=\left(\sum_{i_N=1}^{m_N}\left(F_N'\left(S_{i_N}^{(N),k}\right)\cdot\sum_{i_{N-1}=1}^{m_{N-1}}w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\ldots\cdot\right.\right.$$

$$\cdot\ldots\cdot\sum_{i_n=1}^{m_n} w_{i_n i_{n+1}}^{(n+1)}\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}\Bigg)\times$$

$$\times\Bigg(F_N'\left(S_{i_N}^{(N),k}\right)\cdot\sum_{i_{N-1}=1}^{m_{N-1}} w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\ldots\cdot$$

$$\cdot\ldots\cdot\sum_{i_n=1}^{m_n} w_{i_n i_{n+1}}^{(n+1)}\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{l_{n-1}}^{(n-1),k}\cdot\delta_{l_n}^{i_n}\Bigg)+$$

$$+\sum_{i_N=1}^{m_N}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)\cdot F_N''\left(S_{i_N}^{(N),k}\right)\times$$

$$\times\Bigg(\sum_{i_{N-1}=1}^{m_{N-1}} w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\ldots$$

$$\ldots\cdot\sum_{i_n=1}^{m_n} w_{i_n i_{n+1}}^{(n+1)}\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}\Bigg)\times$$

$$\times\Bigg(\sum_{i_{N-1}=1}^{m_{N-1}} w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\ldots\cdot$$

$$\cdot\ldots\cdot\sum_{i_n=1}^{m_n} w_{i_n i_{n+1}}^{(n+1)}\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{l_{n-1}}^{(n-1),k}\cdot\delta_{l_n}^{i_n}\Bigg)+$$

$$+\sum_{i_N=1}^{m_N}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)\cdot F_N'\left(S_{i_N}^{(N),k}\right)\cdot\sum_{i_{N-1}=1}^{m_{N-1}} w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}''\left(S_{i_{N-1}}^{(N-1),k}\right)\times$$

$$\times\Bigg(\ldots\cdot\sum_{i_n=1}^{m_n} w_{i_n i_{n+1}}^{(n+1)}\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}\Bigg)\times$$

$$\times\Bigg(\ldots\cdot\sum_{i_n=1}^{m_n} w_{i_n i_{n+1}}^{(n+1)}\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{l_{n-1}}^{(n-1),k}\cdot\delta_{l_n}^{i_n}\Bigg)+$$

$$+\ldots+$$

$$+\sum_{i_N=1}^{m_N}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)\cdot F_N'\left(S_{i_N}^{(N),k}\right)\cdot\sum_{i_{N-1}=1}^{m_{N-1}} w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\ldots\cdot$$

$$\cdot\ldots\cdot F_{n+1}'\left(S_{i_{n+1}}^{(n+1),k}\right)\cdot\sum_{i_n=1}^{m_n} w_{i_n i_{n+1}}^{(n+1)}\cdot F_n''\left(S_{i_n}^{(n),k}\right)\cdot\left(y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}\right)\times$$

$$\times\left(y_{l_{n-1}}^{(n-1),k}\cdot\delta_{l_n}^{i_n}\right)\Big)=\left(M_{l_n l_{n-1}}^{(n)}\cdot Y^{(n-1),k}\right)^T\cdot U^{(n),k}\times$$

$$\times\left(M_{j_n j_{n-1}}^{(n)}\cdot Y^{(n-1),k}\right)=\left(K_{l_{n-1}l_n}^{(n),k}\right)^T\cdot U^{(n),k}\cdot\left(K_{j_{n-1}j_n}^{(n),k}\right)$$

где

$$K_{j_{n-1}j_n}^{(n),k}=M_{j_n j_{n-1}}^{(n)}\cdot Y^{(n-1),k},$$

а

$$U^{(n),k}=\left(W^{(n+1)}\cdot MF_n'\right)^T\cdot U^{(n+1),k}\cdot\left(W^{(n+1)}\cdot MF_n'\right)+W^{(n+1)}\cdot MF_n''$$

вычисляется рекуррентно, начиная с

$$U^{(N),k}=\left(MF_N'\right)^2+DE^{(N),k}\cdot MF_n'.$$

Аналогичным образом получаем, что

$$\frac{\partial^2 E_s^{(k)}}{\partial w_{j_{n-1}j_n}^{(n)}\partial T_{l_n}^{(n)}}=\left(K_{(m_{n-1}+1)l_n}^{(n),k}\right)^T\cdot U^{(n),k}\cdot\left(K_{j_{n-1}j_n}^{(n),k}\right),$$

$$\frac{\partial^2 E_s^{(k)}}{\partial T_{j_n}^{(n)}\partial T_{l_n}^{(n)}}=\left(K_{(m_{n-1}+1)l_n}^{(n),k}\right)^T\cdot U^{(n),k}\cdot\left(K_{(m_{n-1}+1)j_n}^{(n),k}\right).$$

Разложим функцию ошибки в ряд Тейлора:

$$E_S(t+1)=\frac{1}{L}\cdot\sum_{k=1}^{L}E_s^k(t+1)=\frac{1}{L}\sum_{k=1}^{L}E_s^k(t)+$$

$$+\frac{1}{L}\cdot\left(\sum_{j_{n-1}=1}^{m_{n-1}}\sum_{j_n=1}^{m_n}\left(\sum_{k=1}^{L}\frac{\partial E_s^k}{\partial w_{j_{n-1}j_n}^{(n)}}\right)\cdot\left(w_{j_{n-1}j_n}^{(n)}(t+1)-w_{j_{n-1}j_n}^{(n)}(t)\right)+\right.$$

$$\left.+\sum_{j_n=1}^{m_n}\left(\sum_{k=1}^{L}\frac{\partial E_s^k}{\partial T_{j_n}^{(n)}}\right)\cdot\left(T_{j_n}^{(n)}(t+1)-T_{j_n}^{(n)}(t)\right)\right)=$$

$$+\frac{1}{2L}\cdot\sum_{k=1}^{L}\Bigg(\sum_{j_{n-1}=1}^{m_{n-1}}\sum_{j_n=1}^{m_n}\sum_{l_{n-1}=1}^{m_{n-1}}\sum_{l_n=1}^{m_n}\frac{\partial^2 E_s^k}{\partial w_{j_{n-1}j_n}^{(n)}\partial w_{l_{n-1}l_n}^{(n)}}\times$$

$$\times\left(w_{j_{n-1}j_n}^{(n)}(t+1)-w_{j_{n-1}j_n}^{(n)}(t)\right)\cdot\left(w_{l_{n-1}l_n}^{(n)}(t+1)-w_{l_{n-1}l_n}^{(n)}(t)\right)+$$

$$+\sum_{j_{n-1}=1}^{m_{n-1}}\sum_{j_n=1}^{m_n}\sum_{l_n=1}^{m_n}\frac{\partial^2 E_s^k}{\partial w_{j_{n-1}j_n}^{(n)}\partial T_{l_n}^{(n)}}\cdot\left(w_{j_{n-1}j_n}^{(n)}(t+1)-w_{j_{n-1}j_n}^{(n)}(t)\right)\times$$

$$\times\left(T_{l_n}^{(n)}(t+1)-T_{l_n}^{(n)}(t)\right)+\sum_{j_n=1}^{m_n}\sum_{l_{n-1}=1}^{m_{n-1}}\sum_{l_n=1}^{m_n}\frac{\partial^2 E_s^k}{\partial T_{j_n}^{(n)}\partial w_{l_{n-1}l_n}^{(n)}}\times$$

$$\times\left(T_{j_n}^{(n)}(t+1)-T_{j_n}^{(n)}(t)\right)\cdot\left(w_{l_{n-1}l_n}^{(n)}(t+1)-w_{l_{n-1}l_n}^{(n)}(t)\right)+$$

$$+\sum_{j_n=1}^{m_n}\sum_{l_n=1}^{m_n}\frac{\partial^2 E_s^k}{\partial T_{j_n}^{(n)}\partial T_{l_n}^{(n)}}\cdot\left(T_{j_n}^{(n)}(t+1)-T_{j_n}^{(n)}(t)\right)\times$$

$$\times\left(T_{l_n}^{(n)}(t+1)-T_{l_n}^{(n)}(t)\right)\Bigg)=$$

$$=E_S(t)-\alpha^{(n)}\cdot\frac{1}{L^2}\cdot\sum_{j_{n-1}=1}^{m_{n-1}}\sum_{j_n=1}^{m_n}\left(\sum_{k=1}^{L}C^{(n)}\cdot M_{j_n j_{n-1}}^{(n)}\cdot Y^{(n-1),k}\right)+$$

$$+\left(\alpha^{(n)}\right)^2\cdot\frac{1}{2L^3}\cdot\sum_{j_{n-1}=1}^{m_{n-1}}\sum_{j_n=1}^{m_n}\sum_{l_{n-1}=1}^{m_{n-1}}\sum_{l_n=1}^{m_n}\left(\sum_{k=1}^{L}\left(\left(K_{l_{n-1}l_n}^{(n),k}\right)^T\cdot U^{(n),k}\cdot\left(K_{j_{n-1}j_n}^{(n),k}\right)\right)\right)$$

Взяв производную функции ошибки по $\alpha^{(n)}$ и прировняв ее, найдем точку минимума:

$$\alpha^{(n)}=\frac{L\cdot\sum_{j_{n-1}=1}^{m_{n-1}}\sum_{j_n=1}^{m_n}\left(\sum_{k=1}^{L}C^{(n)}\cdot M_{j_n j_{n-1}}^{(n)}\cdot Y^{(n-1),k}\right)}{\sum_{j_{n-1}=1}^{m_{n-1}}\sum_{j_n=1}^{m_n}\sum_{l_{n-1}=1}^{m_{n-1}}\sum_{l_n=1}^{m_n}\left(\sum_{k=1}^{L}\left(\left(K_{l_{n-1}l_n}^{(n),k}\right)^T\cdot U^{(n),k}\cdot\left(K_{j_{n-1}j_n}^{(n),k}\right)\right)\right)}.$$

Теорема доказана.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Rumelhart D.E., Hinton G.E., Williams R.J., Learning Internal Representations by Error Propagation, In Parallel Distributed Processing, part 1, chapter 8; Rumelhart D.E. and McClelland J.L., Eds., Cambridge, MA, M.I.T. Press, 1986.
2. Маньяков Н.В., Махнист Л.П. Матричная алгоритмизация обучения многослойных нейронных сетей с использованием градиентных методов // Вестник Брестского государственного технического университета. – Брест: БГТУ, 2002. - №5 (17): Физика, математика, химия. – С. 60-64.
3. V. Golovko, N. Maniakov, L. Makhnist. Multilayer Neural Networks Training Methodic // Proceedings of the Second IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2003). – Lviv, Ukraine, September 8-10, 2003. - P. 185-190.