Procaccia Algorithm, Physical Review A 38, 1988. – PP. 3017-3026.
13. G. Benettin, L. Galgani, A. Giorgilli, J.-M. Strelcyn, Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems: A method for computing all of them. P. I: Theory. P. II: Numerical applications, Meccanica, Vol. 15, 1980. – PP. 9-30.
14. V. Golovko, Y. Savitsky, N. Maniakov. Neural Network for Signal Processing in Measurement Analysis and Industrial Applications: the Case of Chaotic Signal Processing - Chapter of NIMIA Book. – Amsterdam: IOS Press, 2003 – PP. 119-144.
15. X. Zeng, R. Eykholt and R.A. Pielke, Estimating the Lyapunov-Exponent Spectrum from shot Time Series of Low Precision, Physical Review Letter 66, 1991. – PP. 3229-3232.
16. J. Holzfuss and G. Mayer-Kress, An approach to error estimation in the applications of dimensional algorithms, in Dimensions and Entropies in Chaotic Systems, editor G. Mayer-Kress, Springer-Verlag, New York, 1986. – PP. 114-122.
17. M.T. Rosenstein, J.J. Colins, C.J. De Luca, Reconstruction expansion as a geometry-based framework for choosing proper delay time, Physica D 73, 1994. – PP. 82-98.
18. A.M. Fraser and H.L. Swinney, Independent coordinates for strange attractor from mutual information, Physical Review A 33, 1986. – PP. 1134-1140.
19. Вентцель Е.С. Теория вероятностей: Учеб. для вузов. - 7-е изд. стер. – М.: Высшая школа., 2001. – 575 с.
20. H.D.I. Abarbanel, R. Brown, J. Sidorovich and L. Tsimring, The analysis of observed chaotic data in physical systems, Reviews of Modern Physics, Vol. 65, №4, 1993. – PP. 1331-1392.
21. P. Grassberger and I. Procaccia, Measuring the strangeness of strange attractors, Physica D 9, 1983.
22. R. Castro, T. Sauer, Correlation dimension of attractor through interspike intervals, Physical Review E 55, 1997.
23. M.B. Kennel, R. Brown and H.D.I. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction, Physical Review A 45, 1992. – PP. 3403-3411.
24. K.-I. Funahashi. On the approximate realization of continuous mappings by neural networks, Neural Networks, vol. 2, 1989. – PP. 183-192.
25. 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. – СС. 60-64.
26. 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. – PP. 185-190.
27. V. Golovko, Y. Savitsky, N. Maniakov, V. Rubanov. Some Aspects of Chaotic Time Series Analysis // Proceedings of the 2nd International Conference on Neural Networks and Artificial Intelligence. (ICNNAI’2001), October 2-5, 2001 – Minsk: BSU, 2001. – PP. 66-69.

УДК 681.324:519.711.7

Гладкий И.И., Маньяков Н.В., Махнист Л.П.

ОБ ОДНОМ АЛГОРИТМЕ ОБУЧЕНИЯ МНОГОСЛОЙНЫХ НЕЙРОННЫХ СЕТЕЙ НА ОСНОВЕ УСЛОВНОЙ ОПТИМИЗАЦИИ ОШИБКИ

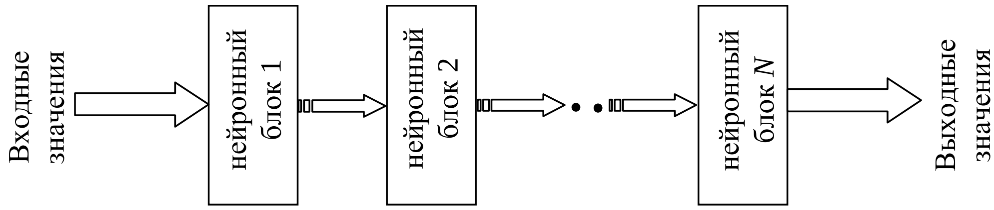

Рис. 1. Блочное представление многослойной нейронной сети.

1. ВВЕДЕНИЕ

Рассмотрим многослойную гетерогенную нейронную сеть, состоящую из N нейронных блоков (рис.1), каждый из которых имеет структуру, представленную на рис. 2.

Входными значениями для каждого нейронного блока являются выходы предыдущего; для первого – последовательность входных образов $\overline{x^k}=\left(x_1^k,\dots,x_{m_0}^k\right),\quad \left(k=\overline{1,L}\right)$.

Выходное значение i_n-ого нейрона n-ого блока сети для k-ого образа определяется рекуррентным соотношением:

$$y_{i_n}^{(n),k}=F_n\left(S_{i_n}^{(n),k}\right),$$

где

$$S_{i_n}^{(n),k}=\sum_{i_{n-1}=1}^{m_{n-1}} w_{i_{n-1}i_n}^{(n)} y_{i_{n-1}}^{(n-1),k}-T_{i_n}^{(n)},\quad i_n=\overline{1,m_n},\quad k=\overline{1,L}.$$

При этом формируется вектор

$$Y^{(n),k}=\begin{pmatrix} y_1^{(n),k} & y_2^{(n),k} & \dots & y_{m_n}^{(n),k} & -1\end{pmatrix}^T.$$

Задача обучения данной многослойной гетерогенной нейронной сети состоит в нахождении матриц весовых коэффициентов

Гладкий Иван Иванович, *ст. преподаватель каф. высшей математики Брестского государственного технического университета.*
Махнист Леонид Петрович, *к.т.н., доцент каф. высшей математики Брестского государственного технического университета.*
Беларусь, БГТУ, 224017, г. Брест, ул. Московская, 267.

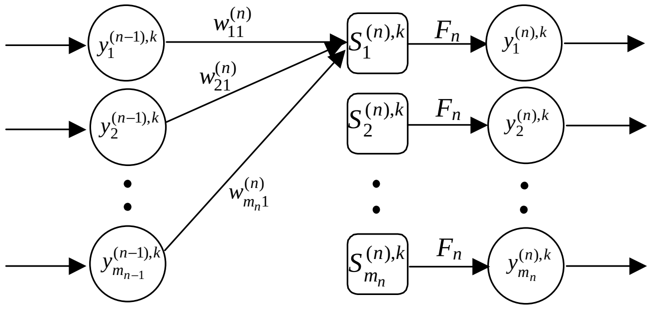

Рис. 2. Архитектура n-ого блока многослойной нейронной сети.

$$W^{(n)}=\begin{pmatrix} w_{11}^{(n)} & w_{21}^{(n)} & \dots & w_{m_{n-1}1}^{(n)} \\ w_{12}^{(n)} & w_{22}^{(n)} & \dots & w_{m_{n-1}2}^{(n)} \\ \dots & \dots & \dots & \dots \\ w_{1m_n}^{(n)} & w_{2m_n}^{(n)} & \dots & w_{m_{n-1}m_n}^{(n)} \end{pmatrix}_{m_n\times m_{n-1}}$$

и столбцов порогов $\overline{T^{(n)}}=\left(T_1^{(n)},\ T_2^{(n)},\dots,\ T_{m_n}^{(n)}\right)^T$, $n=\overline{1,N}$, которые минимизируют некоторую ошибку сети E_S, как отклонение выходных значений сети $y_{i_N}^{(N),k}$ от эталонных $t_{i_N}^k$ – i_N-ого нейрона сети для k-ого образа. В качестве ошибки рассматривается усредненное по количеству образов «квадратичное отклонение»

$$E_S=\frac{1}{2L}\sum_{k=1}^{L}\sum_{i_N=1}^{m_N}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)^2.$$

2. АЛГОРИТМ ОБУЧЕНИЯ

При программной реализации обучения данного класса сетей эффективен метод матричной алгоритмизации [1], который описывается следующим образом:

Модификация синаптических связей и порогов многослойной гетерогенной нейронной сети (рис.1) производится в соответствии с формулами:

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

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

,

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 $m_n\times\left(m_{n-1}+1\right)$ состоит из числа 1 на позиции j_nj_{n-1} и нулей в качестве остальных элементов матрицы.

Изменение синаптических связей и порогов сети при его использовании производится начиная с последнего N-ого до первого блока сети.

При таком типе настройки весовых коэффициентов шаг обучения $\alpha^{(n)}$ может быть как постоянным, так и адаптивным. В последнем случае, являющимся более эффективным, его настраивают на базе метода наискорейшего спуска. Для двухслойных гетерогенных нейронных сетей прямого распространения в данном случае применимы следующие методы: послойное обучение, двухпараметрическое обучение и обобщенный метод наискорейшего спуска [2]. Однако распространение данных методик на нейронные сети, с количеством слоев превышающем два, приводит к громоздким формулам.

В связи с этим разработан эвристический метод обучения сети с вычислением адаптивного шага на основе условной минимизации ошибок каждого слоя. При этом каждый слой нейронной сети рассматривается как однослойная нейронная сеть, обучение которой производится путем настройки градиентным методом выходных значений к полученным «эталонным». Таким образом, в ходе обучения пересчитываются эталонные значения для каждого слоя.

Алгоритм данного метода можно представить следующим образом:

Procedure Обучение сети
begin
 задаем точность обучения ε
repeat
 модифицируем синаптические связи N-ого слоя
 for $n=N-1$ **downto 1 do**
  begin
   for $k=1$ **to** L **do**
    begin
     находим «эталонные» выходы n-ого слоя для каждого образа
    end
   модифицируем синаптические связи n-ого слоя
  end
 вычисляем суммарную ошибку обучения E_S
until $E_S<\varepsilon$
end

3. ТЕОРИТИЧЕСКИЕ ОСНОВЫ АЛГОРИТМЫ ОБУЧЕНИЯ

В основе алгоритма обучения использованы следующие результаты, сформулированные в виде теоремы.

Теорема.

При использовании выше предложенного алгоритма, вычисление «эталонных» значений выходов, при некотором заданном параметре β, производится в соответствии с формулой:

$$t_{j_n}^{(n),k}=y_{j_n}^{(n),k}-\alpha\cdot C^{(n+1)}\cdot W^{(n+1)}\cdot\Delta_{j_n}^n,\ j_n=\overline{1,m_n},$$

с последующей коррекцией:

$$t_{i_n}^{(n),k}:=\begin{cases} a+\beta, & if\ t_{i_n}^{(n),k}<a+\beta \\ t_{i_n}^{(n),k}, & if\ t_{i_n}^{(n),k}\in[a+\beta,b-\beta] \\ b-\beta, & if\ t_{i_n}^{(n),k}>b-\beta \end{cases};$$

адаптивный шаг при этом вычисляется в соответствии с равенством:

$$\alpha = \frac{\sum_{j_n=1}^{m_n}\left(C^{(n+1)}\cdot P_{j_n}^n\right)^2}{\sum_{j_n=1}^{m_n}\sum_{l_n=1}^{m_n}\left(C^{(n+1)}\cdot P_{j_n}^n\right)\cdot\left(\left(P_{j_n}^n\right)^T\cdot U^{(n+1),k}\cdot\left(P_{l_n}^n\right)\right)\cdot\left(C^{(n+1)}\cdot P_{l_n}^n\right)},$$

где

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

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

$$P_{j_n}^n = W^{(n+1)}\cdot\Delta_{j_n}^n,$$

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

а $\Delta_{j_n}^n$ - вектор столбец длины n с нулями в качестве всех элементов, кроме стоящего на позиции j_n равного 1.

При этом настройка весов и порогов сети производится по формулам:

$$w_{j_{n-1}j_n}^{(n)}(t+1) = w_{j_{n-1}j_n}^{(n)}(t)-\alpha^{(n)}\cdot G_{j_{n-1}j_n}^{(n),layer},$$

$$j_{n-1}=\overline{1,m_{n-1}},\quad j_n=\overline{1,m_n}$$

$$T_{j_n}^{(n)}(t+1) = T_{j_n}^{(n)}(t)-\alpha^{(n)}\cdot G_{j_{n-1}j_n}^{(n),layer},\qquad j_n=\overline{1,m_n},$$

с адаптивным шагом, вычисляемым в соответствии с

$$\alpha^{(n)} = \frac{\sum_{j_{n-1}=1}^{m_{n-1}+1}\sum_{j_n=1}^{m_n}\left(G_{j_{n-1}j_n}^{(n),layer}\right)^2}{\sum_{j_{n-1},l_{n-1}=1}^{m_{n-1}+1}\sum_{j_n,l_n=1}^{m_n}G_{l_{n-1}l_n}^{(n),layer}\cdot\left(S_{(n)}\right)_{l_{n-1}l_l}^{j_{n-1}j_n}\cdot G_{l_{n-1}l_n}^{(n),layer}},$$

где $G_{j_{n-1}j_n}^{(n),layer} = \sum_{k=1}^{L}C_{layer}^{(n),k}\cdot K_{j_{n-1}j_n}^{(n),k}$, $\quad C_{layer}^{(n),k} = \varepsilon_n^k\cdot MF_n'$,

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

и

$$\left(S_{(n)}\right)_{l_{n-1}l_l}^{j_{n-1}j_n} = \sum_{k=1}^{L}\left(\left(K_{l_{n-1}l_n}^{(n),k}\right)^T\cdot\left(\left(MF_n'\right)^2+DE^{(n),k}\cdot MF_n''\right)\cdot K_{j_{n-1}j_n}^{(n),k}\right),$$

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

Доказательство.

Рассмотрим n-ый блок многослойной нейронной сети (рис.2). Будем считать его как однослойную нейронную сеть прямого распространения с входными значениями $Y^{(n-1),k} = \left(y_1^{(n-1),k}\quad y_2^{(n-1),k}\quad\ldots\quad y_{m_{n-1}}^{(n-1),k}\quad -1\right)^T$ и целевыми значениями на выходе, получаемыми нижеследующим образом.

Процесс нахождения «эталонные» значения выходов n-ого слоя сети $t_{i_n}^{(n),k}$, $i_n=\overline{1,m_n}$ в соответствии с градиентным методом примет вид

$$t_{j_n}^{(n),k} = y_{j_n}^{(n),k}-\alpha\cdot\frac{\partial E_s^k}{\partial y_{j_n}^{(n),k}},\ j_n=\overline{1,m_n}.$$

В выше приведенной формуле, получаемое значение выхода $y_{j_n}^{(n),k}(t+1)$ принимается за «эталонное» $t_{i_n}^{(n),k}$ при дальнейшей настройке синаптических связей n-ого слоя.

Найдем значение частной производной:

$$\frac{\partial E_s^k}{\partial y_{j_n}^{(n),k}} = \frac{\partial\left(\sum_{i_N=1}^{m_N}\frac{1}{2}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)^2\right)}{\partial y_{j_n}^{(n),k}} = \sum_{i_N=1}^{m_N}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)\cdot\frac{\partial y_{i_N}^{(N),k}}{\partial y_{j_n}^{(n),k}} =$$

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

$$= \sum_{i_N=1}^{m_N}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)\cdot F_N'\left(S_{i_N}^{(N),k}\right)\cdot\sum_{i_{N-1}=1}^{m_{N-1}}w_{i_{N-1}i_N}^{(N)}\cdot\frac{\partial y_{i_{N-1}}^{(N-1),k}}{\partial y_{j_n}^{(n),k}} =$$

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

$$\times\sum_{i_{N-1}=1}^{m_{N-1}}w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\frac{\partial S_{i_{N-1}}^{(N-1),k}}{\partial y_{j_n}^{(n),k}} =$$

$$= \ldots =$$

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

$$\times\sum_{i_{N-1}=1}^{m_{N-1}}w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\ldots\cdot\sum_{i_{n+1}=1}^{m_{n+1}}w_{i_ni_{n+1}}^{(n+1)}\cdot\delta_{j_n}^{i_n} =$$

$$= \varepsilon_N^k\cdot MF_N'\cdot W^{(N)}\cdot MF_{N-1}'\cdot\ldots\cdot W^{(n+1)}\cdot\Delta_{j_n}^n =$$

$$= C^{(n+1)}\cdot W^{(n+1)}\cdot\Delta_{j_n}^n,$$

где

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

а $\Delta_{j_n}^n$ - вектор столбец длины n с нулями в качестве всех элементов, кроме стоящего на позиции j_n равного 1.

Таким образом, формулы для настройки «эталонных» значений выходов n-ого слоя сети $t_{i_n}^{(n),k}$, $i_n=\overline{1,m_n}$ в матричной форме примут вид

$$t_{j_n}^{(n),k} = y_{j_n}^{(n),k}-\alpha\cdot C^{(n+1)}\cdot W^{(n+1)}\cdot\Delta_{j_n}^n,\ j_n=\overline{1,m_n},$$

где в шаг обучения α может быть как постоянным, так и адаптивным.

Однако для большинства активационных функций областью значений является интервал $(a;b)$. Поэтому, задавшись некоторым небольшим значением β, необходимо проследить попадание выходного значения $t_{i_n}^{(n),k}$ нейрона в отрезок $[a+\beta;b+\beta]$. В противном случае, необходимо принять его за граничное предложенного интервала. Тем самым получаем следующее соотношение:

$$t_{i_n}^{(n),k} := \begin{cases} a+\beta, & if\ t_{i_n}^{(n),k} < a+\beta \\ t_{i_n}^{(n),k}, & if\ t_{i_n}^{(n),k}\in[a+\beta,b-\beta] \\ b-\beta, & if\ t_{i_n}^{(n),k} > a-\beta \end{cases}.$$

Для нахождения адаптивного шага обучения, найдем вторую производную функции ошибки по выходу n-ого слоя сети $t_{i_n}^{(n),k}$, $i_n=\overline{1,m_n}$:

$$\frac{\partial^2 E_s^k}{\partial y_{j_n}^{(n),k}\partial y_{l_n}^{(n),k}} = \frac{\partial}{\partial y_{l_n}^{(n),k}}\left(\sum_{i_N=1}^{m_N}\left(y_{i_N}^{(N),k}-t_{i_N}^k\right)\cdot F_N'\left(S_{i_N}^{(N),k}\right)\times\right.$$

$$\left.\times\sum_{i_{N-1}=1}^{m_{N-1}} w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\ldots\cdot\sum_{i_{n+1}=1}^{m_{n+1}} w_{i_n i_{n+1}}^{(n+1)}\cdot\delta_{j_n}^{i_n}\right)=$$

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

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

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

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

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

$$\times\sum_{i_{N-1}=1}^{m_{N-1}} w_{i_{N-1}i_N}^{(N)}\cdot\frac{\partial F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)}{\partial y_{l_n}^{(n),k}}\cdot\ldots\cdot\sum_{i_{n+1}=1}^{m_{n+1}} w_{i_n i_{n+1}}^{(n+1)}\cdot\delta_{j_n}^{i_n}+$$

$$+\ldots+$$

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

$$\left.\times\sum_{i_{N-1}=1}^{m_{N-1}} w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\ldots\cdot\frac{\partial F_{n+1}'\left(S_{i_{n+1}}^{(n+1),k}\right)}{\partial y_{l_n}^{(n),k}}\cdot\sum_{i_{n+1}=1}^{m_{n+1}} w_{i_n i_{n+1}}^{(n+1)}\cdot\delta_{j_n}^{i_n}\right)=$$

$$=\left(\sum_{i_N=1}^{m_N}\left(F_N'\left(S_{i_N}^{(N),k}\right)\cdot\sum_{i_{N-1}=1}^{m_{N-1}} w_{i_{N-1}i_N}^{(N)}\cdot F_{N-1}'\left(S_{i_{N-1}}^{(N-1),k}\right)\cdot\ldots\cdot\sum_{i_{n+1}=1}^{m_{n+1}} w_{i_n i_{n+1}}^{(n+1)}\cdot\delta_{j_n}^{i_n}\right)\times\right.$$

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

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

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

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

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

$$\left(\sum_{i_{N-2}}^{m_{N-2}} w_{i_{N-2}i_{N-1}}^{(N-1)}\cdots\sum_{i_{n+1}=1}^{m_{n+1}} w_{i_n i_{n+1}}^{(n+1)}\cdot\delta_{j_n}^{i_n}\right)\times$$

$$\times\left(\sum_{i_{N-2}}^{m_{N-2}} w_{i_{N-2}i_{N-1}}^{(N-1)}\cdots\sum_{i_{n+1}=1}^{m_{n+1}} w_{i_n i_{n+1}}^{(n+1)}\cdot\delta_{l_n}^{i_n}\right)+$$

$$+\ldots+$$

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

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

$$\left.\times\left(\sum_{i_{n+1}=1}^{m_{n+1}} w_{i_n i_{n+1}}^{(n+1)}\cdot\delta_{j_n}^{i_n}\right)\cdot\left(\sum_{i_{n+1}=1}^{m_{n+1}} w_{i_n i_{n+1}}^{(n+1)}\cdot\delta_{l_n}^{i_n}\right)\right)=$$

$$=\left(W^{(n+1)}\cdot\Delta_{l_n}^{n}\right)^T\cdot U^{(n+1),k}\cdot\left(W^{(n+1)}\cdot\Delta_{j_n}^{n}\right),$$

где

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

вычисляется рекуррентно, начиная с

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

Разложим функцию ошибки в ряд Тейлора:

$$E_s^{(n),k}(t+1)=E_s^{(n),k}(t)+\sum_{j_n=1}^{m_n}\frac{\partial E_s^k}{\partial y_{j_n}^{(n),k}}\cdot\left(t_{j_n}^{(n),k}-y_{j_n}^{(n),k}\right)+$$

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

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

$$+\alpha^2\cdot\sum_{j_n=1}^{m_n}\sum_{l_n=1}^{m_n}\frac{\partial^2 E_s^k}{\partial y_{j_n}^{(n),k}\partial y_{l_n}^{(n),k}}\cdot\frac{\partial E_s^k}{\partial y_{j_n}^{(n),k}}\cdot\frac{\partial E_s^k}{\partial y_{l_n}^{(n),k}}=$$

$$=E_s^{(n),k}(t)-\alpha\cdot\sum_{j_n=1}^{m_n}\left(C^{(n+1)}\cdot W^{(n+1)}\cdot\Delta_{j_n}^{n}\right)^2+$$

$$+\frac{\alpha^2}{2}\cdot\sum_{j_n=1}^{m_n}\sum_{l_n=1}^{m_n}\left(C^{(n+1)}\cdot W^{(n+1)}\cdot\Delta_{j_n}^{n}\right)\times$$

$$\times\left(\left(W^{(n+1)}\cdot\Delta_{l_n}^{n}\right)^T\cdot U^{(n+1),k}\cdot\left(W^{(n+1)}\cdot\Delta_{j_n}^{n}\right)\right)\cdot\left(C^{(n+1)}\cdot W^{(n+1)}\cdot\Delta_{l_n}^{n}\right)=$$

$$=E_s^{(n),k}(t)-\alpha\cdot\sum_{j_n=1}^{m_n}\left(C^{(n+1)}\cdot P_{j_n}^{n}\right)^2+$$

$$+\frac{\alpha^2}{2}\cdot\sum_{j_n=1}^{m_n}\sum_{l_n=1}^{m_n}\left(C^{(n+1)}\cdot P_{j_n}^{n}\right)\cdot\left(\left(P_{j_n}^{n}\right)^T\cdot U^{(n+1),k}\cdot\left(P_{l_n}^{n}\right)\right)\cdot\left(C^{(n+1)}\cdot P_{l_n}^{n}\right),$$

где $P_{j_n}^{n}=W^{(n+1)}\cdot\Delta_{j_n}^{n}$.

Для нахождения минимального значения функции ошибки от шага α, необходимо приравнять ее производную

$$\frac{\partial E_s^{(n),k}(t+1)}{\partial\alpha}=-\sum_{j_n=1}^{m_n}\left(C^{(n+1)}\cdot P_{j_n}^{n}\right)^2+$$

$$+\alpha\cdot\sum_{j_n=1}^{m_n}\sum_{l_n=1}^{m_n}\left(C^{(n+1)}\cdot P_{j_n}^{n}\right)\cdot\left(\left(P_{j_n}^{n}\right)^T\cdot U^{(n+1),k}\cdot\left(P_{l_n}^{n}\right)\right)\cdot\left(C^{(n+1)}\cdot P_{l_n}^{n}\right)$$

к нулю. Тогда получим

$$\alpha=\frac{\sum_{j_n=1}^{m_n}\left(C^{(n+1)}\cdot P_{j_n}^{n}\right)^2}{\sum_{j_n=1}^{m_n}\sum_{l_n=1}^{m_n}\left(C^{(n+1)}\cdot P_{j_n}^{n}\right)\cdot\left(\left(P_{j_n}^{n}\right)^T\cdot U^{(n+1),k}\cdot\left(P_{l_n}^{n}\right)\right)\cdot\left(C^{(n+1)}\cdot P_{l_n}^{n}\right)}.$$

Принимая полученное значение $t_{i_n}^{(n),k}$ за «эталонное», найдем формулы, для модификации весов и порогов n-ого слоя сети.

Разложим суммарную среднеквадратичную ошибку n-ого слоя следующим образом:

$$E_S^{(n)} = \frac{1}{2L}\sum_{k=1}^{L}\sum_{i_n=1}^{m_n}\left(F_n\left(\sum_{i_{n-1}=1}^{m_{n-1}} w_{i_{n-1}i_n}^{(n)} y_{i_{n-1}}^{(n-1),k} - T_{i_n}^{(n)}\right) - t_{i_n}^{(n),k}\right)^2 =$$

$$= \frac{1}{L}\sum_{k=1}^{L} E_s^{(n),k}$$

Тогда,

$$\frac{\partial E_s^{(n),k}}{\partial w_{j_{n-1}j_n}^{(n)}} = \frac{\partial\left(\sum_{i_n=1}^{m_n}\frac{1}{2}\left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)^2\right)}{\partial w_{j_{n-1}j_n}^{(n)}} =$$

$$= \sum_{i_n=1}^{m_n}\left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)\cdot\frac{\partial y_{i_n}^{(n),k}}{\partial w_{j_{n-1}j_n}^{(n)}} =$$

$$= \sum_{i_n=1}^{m_n}\left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot\frac{\partial S_{i_n}^{(n),k}}{\partial w_{j_{n-1}j_n}^{(n)}} =$$

$$= \sum_{i_n=1}^{m_n}\left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n},$$

где

$$\delta_{j_n}^{i_n} = \begin{cases} 1, & i_n = j_n \\ 0, & i_n \neq j_n \end{cases} \quad \text{– символ Кронекера.}$$

В матричной форме данная формула примет вид

$$\frac{\partial E_s^{(n),k}}{\partial w_{j_{n-1}j_n}^{(n)}} = \sum_{i_n=1}^{m_n}\left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n} =$$

$$= \varepsilon_n^k\cdot MF_n'\cdot M_{j_nj_{n-1}}^{(n)}\cdot Y^{(n-1),k} = C_{layer}^{(n),k}\cdot K_{j_{n-1}j_n}^{(n),k}.$$

Аналогично

$$\frac{\partial E_s^{(n),k}}{\partial T_{j_n}^{(n)}} = \frac{\partial\left(\sum_{i_n=1}^{m_n}\frac{1}{2}\left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)^2\right)}{\partial T_{j_n}^{(n)}} = \sum_{i_n=1}^{m_n}\left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)\cdot\frac{\partial y_{i_n}^{(n),k}}{\partial T_{j_n}^{(n)}} =$$

$$= \sum_{i_n=1}^{m_n}\left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot\frac{\partial S_{i_n}^{(n),k}}{\partial T_{j_n}^{(n)}} =$$

$$= \sum_{i_n=1}^{m_n}\left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot(-1)\cdot\delta_{j_n}^{i_n} =$$

$$= \varepsilon_n^k\cdot MF_n'\cdot M_{j_n(m_{n-1}+1)}^{(n)}\cdot Y^{(n-1),k} = C_{layer}^{(n),k}\cdot K_{(m_{n-1}+1)j_n}^{(n),k},$$

где $C_{layer}^{(n),k} = \varepsilon_n^k\cdot MF_n'$, $K_{ij}^{(n),k} = M_{ji}^{(n)}\cdot Y^{(n-1),k}$.

Тогда формулы для изменения весов и порогов n-ого слоя сети в данной матричной форме примут вид:

$$w_{j_{n-1}j_n}^{(n)}(t+1) = w_{j_{n-1}j_n}^{(n)}(t) - \alpha\cdot\frac{1}{L}\cdot\sum_{k=1}^{L}\varepsilon_n^k\cdot MF_n'\cdot M_{j_nj_{n-1}}^{(n)}\cdot Y^{(n-1),k},$$

$$j_{n-1} = \overline{1,m_{n-1}},\quad j_n = \overline{1,m_n}$$

$$T_{j_n}^{(n)}(t+1) = T_{j_n}^{(n)}(t) - \alpha\cdot\frac{1}{L}\cdot\sum_{k=1}^{L}\varepsilon_n^k\cdot MF_n'\cdot M_{j_n(m_{n-1}+1)}^{(n)}\cdot Y^{(n-1),k},$$

$$j_n = \overline{1,m_n}$$

Или, что тоже самое

$$w_{j_{n-1}j_n}^{(n)}(t+1) = w_{j_{n-1}j_n}^{(n)}(t) - \alpha\cdot\frac{1}{L}\cdot G_{j_{n-1}j_n}^{(n),layer},$$

$$j_{n-1} = \overline{1,m_{n-1}},\quad j_n = \overline{1,m_n}$$

$$T_{j_n}^{(n)}(t+1) = T_{j_n}^{(n)}(t) - \alpha\cdot\frac{1}{L}\cdot G_{j_{n-1}j_n}^{(n),layer},$$

$$j_n = \overline{1,m_n}$$

где $G_{j_{n-1}j_n}^{(n),layer} = \sum_{k=1}^{L} C_{layer}^{(n),k}\cdot K_{j_{n-1}j_n}^{(n),k}$.

Найдем частные производные второго порядка функции ошибки:

$$\frac{\partial^2 E_s^{(n),k}}{\partial w_{j_{n-1}j_n}^{(n)}\partial w_{l_{n-1}l_n}^{(n)}} =$$

$$= \frac{\partial}{\partial w_{l_{n-1}l_n}}\left(\sum_{i_n=1}^{m_n}\left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}\right) =$$

$$= \sum_{i_n=1}^{m_n}\left(\frac{\partial\left(\left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)\right)}{\partial w_{l_{n-1}l_n}}\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n} + \left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)\cdot\frac{\partial\left(F_n'\left(S_{i_n}^{(n),k}\right)\right)}{\partial w_{l_{n-1}l_n}}\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}\right) =$$

$$= \sum_{i_n=1}^{m_n}\left(\left(F_n'\left(S_{i_n}^{(n),k}\right)\right)^2\cdot\left(y_{l_{n-1}}^{(n-1),k}\cdot\delta_{l_n}^{i_n}\right)\cdot\left(y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}\right) + \left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)\cdot F_n''\left(S_{i_n}^{(n),k}\right)\cdot\left(y_{l_{n-1}}^{(n-1),k}\cdot\delta_{l_n}^{i_n}\right)\cdot\left(y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}\right)\right) =$$

$$= \left(M_{l_nl_{n-1}}^{(n)}\cdot Y^{(n-1),k}\right)^T\cdot\left(\left(MF_n'\right)^2 + DE^{(n),k}\cdot MF_n''\right)\cdot\left(M_{j_nj_{n-1}}^{(n)}\cdot Y^{(n-1),k}\right) =$$

$$= \left(K_{l_{n-1}l_n}^{(n),k}\right)^T\cdot\left(\left(MF_n'\right)^2 + DE^{(n),k}\cdot MF_n''\right)\cdot K_{j_{n-1}j_n}^{(n),k}$$

$$\frac{\partial^2 E_s^{(n),k}}{\partial T_{j_n}^{(n)}\partial T_{l_n}^{(n)}} = \frac{\partial}{\partial T_{l_n}^{(n)}}\left(\sum_{i_n=1}^{m_n}\left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot(-1)\cdot\delta_{j_n}^{i_n}\right) =$$

$$= \sum_{i_n=1}^{m_n}\left(\frac{\partial\left(\left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)\right)}{\partial T_{l_n}^{(n)}}\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot(-1)\cdot\delta_{j_n}^{i_n} + \left(y_{i_n}^{(n),k} - t_{i_n}^{(n),k}\right)\cdot\frac{\partial\left(F_n'\left(S_{i_n}^{(n),k}\right)\right)}{\partial T_{l_n}^{(n)}}\cdot(-1)\cdot\delta_{j_n}^{i_n}\right) =$$

$$= \sum_{i_n=1}^{m_n}\left(\left(F_n'\left(S_{i_n}^{(n),k}\right)\right)^2\cdot\left((-1)\cdot\delta_{l_n}^{i_n}\right)\cdot\left((-1)\cdot\delta_{j_n}^{i_n}\right) +\right.$$

$$+\left(y_{i_n}^{(n),k}-t_{i_n}^{(n),k}\right)\cdot F_n''\left(S_{i_n}^{(n),k}\right)\cdot\left((-1)\cdot\delta_{l_n}^{i_n}\right)\cdot\left((-1)\cdot\delta_{j_n}^{i_n}\right)\Big)=$$

$$=\left(M_{l_n(m_{n-1}+1)}^{(n)}\cdot Y^{(n-1),k}\right)^T\cdot\left(\left(MF_n'\right)^2+DE^{(n),k}\cdot MF_n''\right)\times$$

$$\times\left(M_{j_n(m_{n-1}+1)}^{(n)}\cdot Y^{(n-1),k}\right)=$$

$$=\left(K_{(m_{n-1}+1)l_n}^{(n),k}\right)^T\cdot\left(\left(MF_n'\right)^2+DE^{(n),k}\cdot MF_n''\right)\cdot K_{(m_{n-1}+1)j_n}^{(n),k}$$

$$\frac{\partial^2E_s^{(n),k}}{\partial w_{j_{n-1}j_n}^{(n)}\partial T_{l_n}^{(n)}}=\frac{\partial}{\partial T_{l_n}^{(n)}}\left(\sum_{i_n=1}^{m_n}\left(y_{i_n}^{(n),k}-t_{i_n}^{(n),k}\right)\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}\right)=$$

$$=\sum_{i_n=1}^{m_n}\left(\frac{\partial\left(\left(y_{i_n}^{(n),k}-t_{i_n}^{(n),k}\right)\right)}{\partial T_{l_n}^{(n)}}\cdot F_n'\left(S_{i_n}^{(n),k}\right)\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}+\right.$$

$$\left.+\left(y_{i_n}^{(n),k}-t_{i_n}^{(n),k}\right)\cdot\frac{\partial\left(F_n'\left(S_{i_n}^{(n),k}\right)\right)}{\partial T_{l_n}^{(n)}}\cdot y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}\right)=$$

$$=\sum_{i_n=1}^{m_n}\left(\left(F_n'\left(S_{i_n}^{(n),k}\right)\right)^2\cdot\left((-1)\cdot\delta_{l_n}^{i_n}\right)\cdot\left(y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}\right)+\right.$$

$$\left.+\left(y_{i_n}^{(n),k}-t_{i_n}^{(n),k}\right)\cdot F_n''\left(S_{i_n}^{(n),k}\right)\cdot\left((-1)\cdot\delta_{l_n}^{i_n}\right)\cdot\left(y_{j_{n-1}}^{(n-1),k}\cdot\delta_{j_n}^{i_n}\right)\right)=$$

$$=\left(M_{l_n(m_{n-1}+1)}^{(n)}\cdot Y^{(n-1),k}\right)^T\cdot\left(\left(MF_n'\right)^2+DE^{(n),k}\cdot MF_n''\right)\times$$

$$\times\left(M_{j_nj_{n-1}}^{(n)}\cdot Y^{(n-1),k}\right)=$$

$$=\left(K_{(m_{n-1}+1)l_n}^{(n),k}\right)^T\cdot\left(\left(MF_n'\right)^2+DE^{(n),k}\cdot MF_n''\right)\cdot K_{j_{n-1}j_n}^{(n),k}$$

Разложим каждое слагаемое функции ошибки однослойной нейронной сети по формуле Тейлора:

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

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

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

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

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

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

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

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

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

$$+\sum_{j_n=1}^{m_n}\sum_{l_n=1}^{m_n}\frac{\partial^2E_s^{(n),k}}{\partial T_{j_n}^{(n)}\partial T_{l_n}^{(n)}}\times$$

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

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

$$+\alpha^2\cdot\frac{1}{2L^3}\cdot\left(\sum_{j_{n-1}=1}^{m_{n-1}+1}\sum_{j_n=1}^{m_n}\sum_{l_{n-1}=1}^{m_{n-1}+1}\sum_{l_n=1}^{m_n}\left(G_{l_{n-1}l_n}^{(n),layer}\times\right.\right.$$

$$\left.\left.\times\sum_{k=1}^{L}\left(\left(K_{l_{n-1}l_n}^{(n),k}\right)^T\cdot\left(\left(MF_n'\right)^2+DE^{(n),k}\cdot MF_n''\right)\cdot K_{j_{n-1}j_n}^{(n),k}\right)\cdot G_{j_{n-1}j_n}^{(n),layer}\right)\right)$$

Найдем, при каком значении шага α полученное значение ошибки сети будет минимальным. Для этого решим уравнение:

$$\frac{\partial E_S^{(n)}(t+1)}{\partial\alpha}=\frac{1}{L^2}\cdot\sum_{j_{n-1}=1}^{m_{n-1}+1}\sum_{j_n=1}^{m_n}\left(G_{j_{n-1}j_n}^{(n),layer}\right)^2+$$

$$+\alpha\cdot\frac{1}{L^3}\cdot\left(\sum_{j_{n-1}=1}^{m_{n-1}+1}\sum_{j_n=1}^{m_n}\sum_{l_{n-1}=1}^{m_{n-1}+1}\sum_{l_n=1}^{m_n}\left(G_{l_{n-1}l_n}^{(n),layer}\times\right.\right.$$

$$\left.\left.\times\sum_{k=1}^{L}\left(\left(K_{l_{n-1}l_n}^{(n),k}\right)^T\cdot\left(\left(MF_n'\right)^2+DE^{(n),k}\cdot MF_n''\right)\cdot K_{j_{n-1}j_n}^{(n),k}\right)\cdot G_{j_{n-1}j_n}^{(n),layer}\right)\right)$$

Полученное значение α равно:

$$\alpha=\frac{L\cdot\sum_{j_{n-1}=1}^{m_{n-1}+1}\sum_{j_n=1}^{m_n}\left(G_{j_{n-1}j_n}^{(n),layer}\right)^2}{\sum_{j_{n-1},l_{n-1}=1}^{m_{n-1}+1}\sum_{j_n,l_n=1}^{m_n}G_{l_{n-1}l_n}^{(n),layer}\cdot\left(S_{(n)}\right)_{l_{n-1}l_l}^{j_{n-1}j_n}\cdot G_{l_{n-1}l_n}^{(n),layer}},$$

где

$$\left(S_{(n)}\right)_{l_{n-1}l_l}^{j_{n-1}j_n}=\sum_{k=1}^{L}\left(\left(K_{l_{n-1}l_n}^{(n),k}\right)^T\cdot\left(\left(MF_n'\right)^2+DE^{(n),k}\cdot MF_n''\right)\cdot K_{j_{n-1}j_n}^{(n),k}\right)$$

Таким образом

$$\alpha^{(n)}=\alpha\cdot\frac{1}{L}=\frac{\sum_{j_{n-1}=1}^{m_{n-1}+1}\sum_{j_n=1}^{m_n}\left(G_{j_{n-1}j_n}^{(n),layer}\right)^2}{\sum_{j_{n-1},l_{n-1}=1}^{m_{n-1}+1}\sum_{j_n,l_n=1}^{m_n}G_{l_{n-1}l_n}^{(n),layer}\cdot\left(S_{(n)}\right)_{l_{n-1}l_l}^{j_{n-1}j_n}\cdot G_{l_{n-1}l_n}^{(n),layer}}.$$

Теорема доказана.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Маньяков Н.В., Махнист Л.П. Матричная алгоритмизация обучения многослойных нейронных сетей с использованием градиентных методов // Вестник Брестского государственного технического университета. Физика, математика, химия. №5 (17) – Брест: БГТУ, 2002.– С. 60-64.
2. Маньяков Н.В., Махнист Л.П., Рубанов В.С. Нейросетевые алгоритмы решения одного класса разностно-дифференциальных уравнений // Тезисы докладов международной конференции «Аналитические методы анализа и дифференциальных уравнений» AMADE’2003, Минск, 4-9 сентября 2003 г. – Минск, 2003. – С. 118-119.