УДК 681.324

Гладкий И.И., Махнист Л.П.

ОЦЕНКИ КВАЗИОПТИМАЛЬНОГО ШАГА ОБУЧЕНИЯ НЕЙРОННЫХ СЕТЕЙ

1. ВВЕДЕНИЕ

В работе рассматривается нейронная сеть, состоящая из n нейронных элементов распределительного слоя и m - выходного слоя с использованием нелинейной функции активации. Получены выражения для определения квазиоптимального шага обучения нейронной сети в случае группового обучения, а также выражения для изменения весовых коэффициентов и порогов нейронных элементов. Показано, что эти соотношения можно получить, используя выражение квазиоптимальной величины шага для метода наискорейшего спуска. Получена оценка квазиоптимального шага обучения нейронной сети в случае группового обучения.

Рассмотрим однослойную нейронную сеть, состоящую из n нейронных элементов распределительного слоя и m - выходного слоя (рисунок 1).

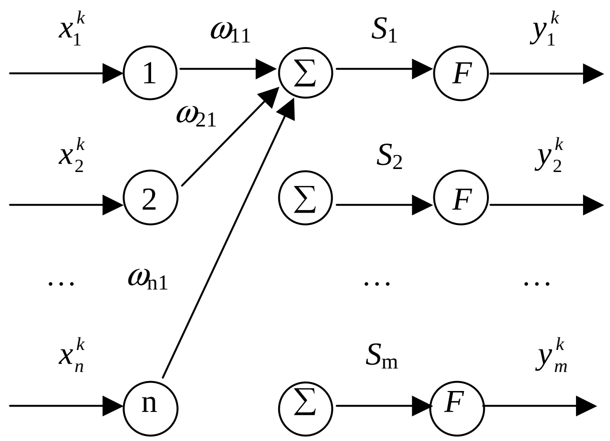

Рисунок 1 – Схема функционирования нейронной сети.

Для данной сети каждый нейрон распределительного слоя имеет синаптические связи ω_{ij}, $\left(i=\overline{1,n},\ j=\overline{1,m}\right)$ со всеми нейронами обрабатывающего слоя. В качестве нейронов выходного слоя используются элементы с некоторой функцией активации F [1]. На вход сети подаются несколько образов $\overline{x^k}=\left(x_1^k,\ldots,x_n^k\right)$ $\left(k=\overline{1,L}\right)$.

Выходное значение j-ого нейрона сети для k-ого образа в момент времени t определяется выражением:

$$y_j^k(t)=F\left(S_j^k(t)\right), \qquad (1)$$

где

$$S_j^k(t)=\sum_i \omega_{ij}(t)x_i^k-T_j(t),\ \ j=\overline{1,m},\ \ k=\overline{1,L}.$$

Задача обучения нейронной сети с фиксированной функцией активации F состоит в нахождении весовых коэффициентов ω_{ij} $\left(i=\overline{1,n},\ j=\overline{1,m}\right)$ и порогов нейронных элементов T_j $\left(j=\overline{1,m}\right)$, которые минимизируют некоторую ошибку сети E_S, как отклонение выходных значений $y_j^k(t)$ от эталонных значений t_j^k – j-ого нейрона сети для k-ого образа. Рассмотрим процедуру обучения нейронной сети с использованием метода наискорейшего спуска.

2. ВЫРАЖЕНИЕ КВАЗИОПТИМАЛЬНОЙ ВЕЛИЧИНЫ ШАГА ОБУЧЕНИЯ

Рассмотрим дважды непрерывно дифференцируемую функцию $E_S(t)$ – ошибку сети, как функцию нескольких переменных:

$$E_S(\omega_{11},\omega_{21},\ldots,\omega_{n1},T_1,\omega_{12},\omega_{22},\ldots,\omega_{n2},T_2,\ldots,\omega_{1m},\omega_{2m},\ldots,\omega_{nm},T_m).$$

Введем обозначения:

вектора переменных

$$\overline{W}=(\omega_{11},\omega_{21},\ldots,\omega_{n1},T_1,\omega_{12},\omega_{22},\ldots,$$
$$,\omega_{n2},T_2,\ldots,\omega_{1m},\omega_{2m},\ldots,\omega_{nm},T_m)^T,$$

вектора градиента функции E_S

$$\nabla E_S=\left(\frac{\partial E_S}{\partial\omega_{11}},\frac{\partial E_S}{\partial\omega_{21}},\ldots,\frac{\partial E_S}{\partial\omega_{n1}},\frac{\partial E_S}{\partial T_1},\frac{\partial E_S}{\partial\omega_{12}},\frac{\partial E_S}{\partial\omega_{22}},\right.$$
$$\left.,\ldots,\frac{\partial E_S}{\partial\omega_{n2}},\frac{\partial E_S}{\partial T_2},\ldots,\frac{\partial E_S}{\partial\omega_{1m}},\frac{\partial E_S}{\partial\omega_{2m}},\ldots,\frac{\partial E_S}{\partial\omega_{nm}},\frac{\partial E_S}{\partial T_m}\right)^T$$

и матрицы Гессе вторых производных $\nabla^2 E_S$.

Разложим функцию в ряд Тейлора, ограничиваясь частными производными второго порядка включительно:

$$E_S(t+1)=E_S(t)+\left(\nabla E_S(t),\ \overline{W}(t+1)-\overline{W}(t)\right)+$$
$$+\frac{1}{2}\left(\nabla^2 E_S(t)\cdot\left(\overline{W}(t+1)-\overline{W}(t)\right),\ \overline{W}(t+1)-\overline{W}(t)\right)'$$

где $\left(\nabla E_S(t),\ \overline{W}(t+1)-\overline{W}(t)\right)$ – скалярное произведение векторов $\nabla E_S(t)$ и $\overline{W}(t+1)-\overline{W}(t)$, $\left(\nabla^2 E_S(t)\cdot\left(\overline{W}(t+1)-\overline{W}(t)\right),\ \overline{W}(t+1)-\overline{W}(t)\right)$ – скалярное произведение векторов $\nabla^2 E_S(t)\cdot\left(\overline{W}(t+1)-\overline{W}(t)\right)$ и $\left(\overline{W}(t+1)-\overline{W}(t)\right)$, а $\nabla^2 E_S(t)\cdot\left(\overline{W}(t+1)-\overline{W}(t)\right)$ – произведение матрицы $\nabla^2 E_S(t)$ на вектор $\left(\overline{W}(t+1)-\overline{W}(t)\right)$.

Учитывая, что в соответствии с идеей метода наискорейшего спуска

$$\overline{W}(t+1)=\overline{W}(t)-\alpha(t)\cdot\nabla E_S(t), \qquad (2)$$

получим

$$E_S(t+1)=E_S(t)+\left(\nabla E_S(t),\ -\alpha(t)\cdot\nabla E_S(t)\right)+$$
$$+\frac{1}{2}\left(\nabla^2 E_S(t)\cdot\left(-\alpha(t)\cdot\nabla E_S(t)\right),\ -\alpha(t)\cdot\nabla E_S(t)\right)$$

или

$$E_S(t+1)=E_S(t)-\alpha(t)\cdot\left(\nabla E_S(t),\ \nabla E_S(t)\right)+$$
$$+\frac{1}{2}\alpha^2(t)\cdot\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \nabla E_S(t)\right).$$

Тогда

$$\frac{\partial E_S}{\partial \alpha} = -\left(\nabla E_S(t),\ \nabla E_S(t)\right) +$$
$$+\alpha(t)\cdot\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \nabla E_S(t)\right) = 0,$$

если

$$\alpha(t) = \frac{\left(\nabla E_S(t),\ \nabla E_S(t)\right)}{\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \nabla E_S(t)\right)}$$

или

$$\alpha(t) = \frac{\left\|\nabla E_S(t)\right\|^2}{\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \nabla E_S(t)\right)}. \quad (3)$$

Таким образом, можно получить выражение квазиоптимальной величины шага для метода наискорейшего спуска с использованием частных производных первого и второго порядка: (4).

Пусть среднеквадратичная ошибка сети определяется соотношением

$$E_S(t) = \sum_k\sum_j E_j^k(t) = \frac{1}{2}\sum_k\sum_j\left(y_j^k(t) - t_j^k\right)^2, \quad (5)$$

где t_j^k - эталонное выходное значение j-ого нейрона сети для k-ого образа.

Используя соотношение (4), получим выражение квазиоптимальной величины шага для метода наискорейшего спуска.

Тогда, учитывая, что

$$\frac{\partial E_S}{\partial \omega_{ij}} = \sum_k\left(y_j^k - t_j^k\right)F'\left(S_j^k\right)x_i^k$$

и
$$\frac{\partial E_S}{\partial T_j} = -\sum_k\left(y_j^k - t_j^k\right)F'\left(S_j^k\right), \quad (6)$$

имеем:

$$\frac{\partial^2 E_S}{\partial \omega_{ij}\partial \omega_{qp}} = \sum_k\left(\left(F'\left(S_j^k\right)\right)^2 + \left(y_j^k - t_j^k\right)F''\left(S_j^k\right)\right)x_i^k x_q^k,$$

если $j = p$, и $\frac{\partial^2 E_S}{\partial \omega_{ij}\partial \omega_{qp}} = 0$ в противном случае;

$$\frac{\partial^2 E_S}{\partial \omega_{ij}\partial T_m} = -\sum_k\left(\left(F'\left(S_j^k\right)\right)^2 + \left(y_j^k - t_j^k\right)F''\left(S_j^k\right)\right)x_i^k,$$

если $j = m$, и $\frac{\partial^2 E_S}{\partial \omega_{ij}\partial T_m} = 0$ в противном случае;

$$\frac{\partial^2 E_S}{\partial T_j\partial T_m} = \sum_k\left(\left(F'\left(S_j^k\right)\right)^2 + \left(y_j^k - t_j^k\right)F''\left(S_j^k\right)\right),$$

если $j = m$, и $\frac{\partial^2 E_S}{\partial T_j\partial T_m} = 0$ в противном случае.

Поэтому (7).

Вычислим отдельно слагаемые числителя:

$$\sum_i\sum_j\left(\frac{\partial E_S}{\partial \omega_{ij}}\right)^2 = \sum_j\sum_k\left(y_j^k - t_j^k\right)F'\left(S_j^k\right)\times$$
$$\times\left(\sum_p\left(y_j^p - t_j^p\right)F'\left(S_j^p\right)\sum_i x_i^k x_i^p\right)$$

и

$$\sum_j\left(\frac{\partial E_S}{\partial T_j}\right)^2 = \sum_j\sum_k\left(y_j^k - t_j^k\right)F'\left(S_j^k\right)\times$$
$$\times\left(\sum_p\left(y_j^p - t_j^p\right)F'\left(S_j^p\right)\right).$$

Тогда числитель равен

$$\sum_i\sum_j\left(\frac{\partial E_S}{\partial \omega_{ij}}\right)^2 + \sum_j\left(\frac{\partial E_S}{\partial T_j}\right)^2 = \sum_j\sum_k\left(y_j^k - t_j^k\right)F'\left(S_j^k\right)\times$$
$$\times\left(\sum_p\left(y_j^p - t_j^p\right)F'\left(S_j^p\right)\left(\sum_i x_i^p x_i^k + 1\right)\right) =$$
$$= \sum_j\sum_k\left(y_j^k - t_j^k\right)F'\left(S_j^k\right)a_j^k, \text{ где}$$

$$\alpha(t) = \frac{\sum_i\sum_j\left(\frac{\partial E_S}{\partial \omega_{ij}}\right)^2 + \sum_j\left(\frac{\partial E_S}{\partial T_j}\right)^2}{\sum_i\sum_j\sum_q\sum_p\frac{\partial^2 E_S}{\partial \omega_{ij}\partial \omega_{qp}}\frac{\partial E_S}{\partial \omega_{ij}}\frac{\partial E_S}{\partial \omega_{qp}} + 2\cdot\sum_i\sum_j\sum_m\frac{\partial^2 E_S}{\partial \omega_{ij}\partial T_m}\frac{\partial E_S}{\partial \omega_{ij}}\frac{\partial E_S}{\partial T_m} + \sum_j\sum_m\frac{\partial^2 E_S}{\partial T_j\partial T_m}\frac{\partial E_S}{\partial T_j}\frac{\partial E_S}{\partial T_m}}. \quad (4)$$

$$\alpha(t) = \frac{\sum_i\sum_j\left(\frac{\partial E_S}{\partial \omega_{ij}}\right)^2 + \sum_j\left(\frac{\partial E_S}{\partial T_j}\right)^2}{\sum_j\sum_i\sum_q\frac{\partial^2 E_S}{\partial \omega_{ij}\partial \omega_{qj}}\frac{\partial E_S}{\partial \omega_{ij}}\frac{\partial E_S}{\partial \omega_{qj}} + 2\sum_j\sum_i\frac{\partial^2 E_S}{\partial \omega_{ij}\partial T_j}\frac{\partial E_S}{\partial \omega_{ij}}\frac{\partial E_S}{\partial T_j} + \sum_j\frac{\partial^2 E_S}{\partial T_j^{\,2}}\left(\frac{\partial E_S}{\partial T_j}\right)^2} \quad (7)$$

$$\alpha(t) = \frac{\sum_{j=1}^{m}\left(F'\left(S_j\right)\right)^2\left(y_j - t_j\right)^2}{\left(\sum_{i=1}^{n}x_i^2 + 1\right)\sum_{j=1}^{m}\left(\left(F'\left(S_j\right)\right)^2 + \left(y_j - t_j\right)F''\left(S_j\right)\right)\left(F'\left(S_j\right)\right)^2\left(y_j - t_j\right)^2} \quad (12)$$

$$a_j^k = \sum_i \frac{\partial E_S}{\partial \omega_{ij}(t)} x_i^k - \frac{\partial E_S}{\partial T_j(t)} = \sum_p \left(y_j^p - t_j^p\right) F'\left(S_j^p\right) \times \left(\sum_i x_i^p x_i^k + 1\right). \quad (8)$$

Вычисляя слагаемые знаменателя, получим

$$\sum_j \sum_i \sum_q \frac{\partial^2 E_S}{\partial \omega_{ij} \partial \omega_{qj}} \frac{\partial E_S}{\partial \omega_{ij}} \frac{\partial E_S}{\partial \omega_{qj}} =$$

$$= \sum_j \sum_k \left(\left(F'\left(S_j^k\right)\right)^2 + \left(y_j^k - t_j^k\right) F''\left(S_j^k\right) \right) \times$$

$$\times \left(\sum_p \left(F'\left(S_j^p\right)\right)\left(y_j^p - t_j^p\right) \sum_i x_i^k x_i^p \right)^2$$

$$2 \cdot \sum_i \sum_j \frac{\partial^2 E_S}{\partial \omega_{ij} \partial T_j} \frac{\partial E_S}{\partial \omega_{ij}} \frac{\partial E_S}{\partial T_j} =$$

$$= 2 \sum_j \sum_k \left(\left(F'\left(S_j^k\right)\right)^2 + \left(y_j^k - t_j^k\right) F''\left(S_j^k\right) \right) \times$$

$$\times \left(\sum_p \left(F'\left(S_j^p\right)\right)\left(y_j^p - t_j^p\right) \sum_i x_i^k x_i^p \right) \left(\sum_m \left(F'\left(S_j^m\right)\right)\left(y_j^m - t_j^m\right) \right),$$

$$\sum_j \frac{\partial^2 E_S}{\partial T_j^2} \left(\frac{\partial E_S}{\partial T_j} \right)^2 =$$

и $= \sum_j \sum_k \left(\left(F'\left(S_j^k\right)\right)^2 + \left(y_j^k - t_j^k\right) F''\left(S_j^k\right) \right) \times$

$$\times \left(\sum_m \left(F'\left(S_j^m\right)\right)\left(y_j^m - t_j^m\right) \right)^2.$$

Тогда знаменатель равен

$$\sum_j \sum_i \sum_q \frac{\partial^2 E_S}{\partial \omega_{ij} \partial \omega_{qj}} \frac{\partial E_S}{\partial \omega_{ij}} \frac{\partial E_S}{\partial \omega_{qj}} +$$

$$+ 2 \cdot \sum_i \sum_j \frac{\partial^2 E_S}{\partial \omega_{ij} \partial T_j} \frac{\partial E_S}{\partial \omega_{ij}} \frac{\partial E_S}{\partial T_j} + \sum_j \frac{\partial^2 E_S}{\partial T_j^2} \left(\frac{\partial E_S}{\partial T_j} \right)^2 =$$

$$= \sum_j \sum_k \left(\left(F'\left(S_j^k\right)\right)^2 + \left(y_j^k - t_j^k\right) F''\left(S_j^k\right) \right)^2 \times$$

$$\times \left(\sum_p \left(F'\left(S_j^p\right)\right)\left(y_j^p - t_j^p\right) \left(\sum_i x_i^k x_i^p + 1 \right) \right) =$$

$$= \sum_j \sum_k \left(\left(F'\left(S_j^k\right)\right)^2 + \left(y_j^k - t_j^k\right) F''\left(S_j^k\right) \right) \left(a_j^k\right)^2,$$

учитывая (8).

Таким образом, в соответствии с (7), получим

$$\alpha(t) = \frac{\sum_{k=1}^{L} \sum_{j=1}^{m} F'\left(S_j^k\right)\left(y_j^k - t_j^k\right) a_j^k}{\sum_{k=1}^{L} \sum_{j=1}^{m} \left(\left(F'\left(S_j^k\right)\right)^2 + \left(y_j^k - t_j^k\right) F''\left(S_j^k\right) \right) \left(a_j^k\right)^2}, \quad (9)$$

где

$$a_j^k = \sum_{p=1}^{L} \left(y_j^p - t_j^p\right) F'\left(S_j^p\right) \left(\sum_{i=1}^{n} x_i^p x_i^k + 1 \right),$$

$$j = \overline{1,m}, \; k = \overline{1,L}.$$

Учитывая (1), (6) и (9), получим выражения для модификации синаптических связей с использованием квазиоптимального шага обучения:

$$\omega_{ij}(t+1) = \omega_{ij}(t) - \alpha(t) \frac{\partial E_S}{\partial \omega_{ij}(t)} =$$

$$= \omega_{ij}(t) - \alpha(t) \cdot \sum_{k=1}^{L} \left(y_j^k(t) - t_j^k \right) \cdot F'\left(S_j^k(t)\right) x_i^k, \quad (10)$$

$$i = \overline{1,n}, \; j = \overline{1,m},$$

$$T_j(t+1) = T_j(t) - \alpha(t) \frac{\partial E_S}{\partial T_j(t)} =$$

$$= T_j(t) + \alpha(t) \cdot \sum_{k=1}^{L} \left(y_j^k(t) - t_j^k \right) F'\left(S_j^k(t)\right), \; j = \overline{1,m}.$$

Таким образом, доказано следующее утверждение.

Теорема 1. Для обучения нейронной сети с использованием метода наискорейшего спуска для среднеквадратичной ошибки сети, определяемой выражением (5), величина квазиоптимального шага обучения $\alpha(t)$ в момент времени t определяется соотношением (9). Модификации весовых коэффициентов ω_{ij}, $\left(i = \overline{1,n}, \; j = \overline{1,m}\right)$ и порогов нейронных элементов T_j, $\left(j = \overline{1,m}\right)$, с использованием такого шага обучения определяются выражениями (10).

Получим частные случаи соотношения (9), сформулированные в виде следствий.

Следствие 1. В случае одного j-ого выходного нейронного элемента, соотношение (9), примет вид:

$$\alpha(t) = \frac{\sum_{k=1}^{L} F'\left(S_j^k\right)\left(y_j^k - t_j^k\right) a_j^k}{\sum_{k=1}^{L} \left(\left(F'\left(S_j^k\right)\right)^2 + \left(y_j^k - t_j^k\right) F''\left(S_j^k\right) \right) \left(a_j^k\right)^2}, \quad (11)$$

где $a_j^k = \sum_{p=1}^{L} \left(y_j^p - t_j^p\right) F'\left(S_j^p\right) \left(\sum_{i=1}^{n} x_i^p x_i^k + 1 \right), \; k = \overline{1,L}$.

Следствие 2. В случае одного образа, т. е. при $L = 1$, соотношение (9), принимает вид:

$$\alpha(t) = \frac{\sum_{j=1}^{m} F'\left(S_j\right)\left(y_j - t_j\right) a_j}{\sum_{j=1}^{m} \left(\left(F'\left(S_j\right)\right)^2 + \left(y_j - t_j\right) F''\left(S_j\right) \right) \left(a_j\right)^2},$$

где $a_j = (y_j - t_j)F'(S_j^p)\left(\sum_{i=1}^{n} x_i^2 + 1\right)$, $j = \overline{1,m}$,

или (12).

Заметим, что соотношения (3), (4) могут быть использованы для получения выражений величины квазиоптимального шага обучения и для других функций E_S, например, $E_S = \sum_k \sum_j \left|y_j^k - t_j^k\right|$.

3. ОЦЕНКИ ШАГА ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ

Получим оценку величины квазиоптимального шага обучения для среднеквадратичной функции ошибки, определяемой соотношением (5).

Рассмотрим матрицу Гессе функции E_S. Учитывая выражения для частных производных второго порядка, полученные выше, имеем

$$\nabla^2 E_S = \begin{pmatrix} \nabla^2 E_1 & 0 & \cdots & 0 \\ 0 & \nabla^2 E_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \nabla^2 E_m \end{pmatrix}_{m(n+1)\times m(n+1)}, \quad (13)$$

где

$$\nabla^2 E_j = \begin{pmatrix} \frac{\partial^2 E}{\partial \omega_{1j}^2} & \frac{\partial^2 E}{\partial \omega_{1j}\partial \omega_{2j}} & \cdots & \frac{\partial^2 E}{\partial \omega_{1j}\partial \omega_{nj}} & \frac{\partial^2 E}{\partial \omega_{1j}\partial T_j} \\ \frac{\partial^2 E}{\partial \omega_{2j}\partial \omega_{1j}} & \frac{\partial^2 E}{\partial \omega_{2j}^2} & \cdots & \frac{\partial^2 E}{\partial \omega_{2j}\partial \omega_{nj}} & \frac{\partial^2 E}{\partial \omega_{2j}\partial T_j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 E}{\partial \omega_{nj}\partial \omega_{1j}} & \frac{\partial^2 E}{\partial \omega_{nj}\partial \omega_{2j}} & \cdots & \frac{\partial^2 E}{\partial \omega_{nj}^2} & \frac{\partial^2 E}{\partial \omega_{nj}\partial T_j} \\ \frac{\partial^2 E}{\partial T_j\partial \omega_{1j}} & \frac{\partial^2 E}{\partial T_j\partial \omega_{2j}} & \cdots & \frac{\partial^2 E}{\partial T_j\partial \omega_{nj}} & \frac{\partial^2 E}{\partial T_j^2} \end{pmatrix}_{(n+1)\times(n+1)}.$$

Тогда в случае одного образа

$$\nabla^2 E_j(t) = \left(\left(F'(S_j)\right)^2 + (y_j - t_j)F''(S_j)\right) \times \begin{pmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_n & -x_1 \\ x_2x_1 & x_2^2 & \cdots & x_2x_n & -x_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_nx_1 & x_nx_2 & \cdots & x_n^2 & -x_n \\ -x_1 & -x_2 & \cdots & -x_n & 1 \end{pmatrix}$$

Заметим, что строки и столбцы матрицы $\nabla^2 E_j(t)$ пропорциональны. Поэтому матрица $\nabla^2 E_j(t)$ является неотрицательно определенной, так как ее главные миноры Δ_p $(p = \overline{1,n+1})$ неотрицательны:

$$\Delta_1(i) = \left(\left(F'(S_j)\right)^2 + (y_j - t_j)F''(S_j)\right)x_i^2 \geq 0, \quad (i = \overline{2,n+1}),$$

$$\Delta_1(n+1) = \left(\left(F'(S_j)\right)^2 + (y_j - t_j)F''(S_j)\right) \geq 0,$$

$$\Delta_p = 0, \quad (p = \overline{2,n+1}).$$

Неравенство $\Delta_1(n+1) \geq 0$ гарантируется выбором соответствующего начального приближения весовых коэффициентов $\omega_{ij}(0)$ $(i = \overline{1,n},\ j = \overline{1,m})$ и порогов нейронных элементов сети $T_j(0)$ $(j = \overline{1,m})$.

Найдем собственные значения матрицы $\nabla^2 E_j(t)$:

$$\left|\nabla^2 E_j(t) - \lambda I\right| = \begin{vmatrix} cx_1^2 - \lambda & cx_1x_2 & \cdots & cx_1x_n & -cx_1 \\ cx_2x_1 & cx_2^2 - \lambda & \cdots & cx_2x_n & -cx_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ cx_nx_1 & cx_nx_2 & \cdots & cx_n^2 - \lambda & -cx_n \\ -cx_1 & -cx_2 & \cdots & -cx_n & c - \lambda \end{vmatrix} =$$

$$= \begin{vmatrix} cx_1^2 - \lambda & cx_1x_2 & \cdots & cx_1x_n & -cx_1 \\ \frac{x_2}{x_1}\lambda & -\lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{x_n}{x_1}\lambda & 0 & \cdots & -\lambda & 0 \\ -\frac{1}{x_1}\lambda & 0 & \cdots & 0 & -\lambda \end{vmatrix} =$$

$$= (-1)^n \lambda^n \begin{vmatrix} cx_1^2 - \lambda & cx_1x_2 & \cdots & cx_1x_n & -cx_1 \\ -\frac{x_2}{x_1} & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\frac{x_n}{x_1} & 0 & \cdots & 1 & 0 \\ \frac{1}{x_1} & 0 & \cdots & 0 & 1 \end{vmatrix} =$$

$$= (-1)^n \lambda^n \begin{vmatrix} c\left(\sum_i x_i^2 + 1\right) - \lambda & 0 & \cdots & 0 & 0 \\ -\frac{x_2}{x_1} & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\frac{x_n}{x_1} & 0 & \cdots & 1 & 0 \\ \frac{1}{x_1} & 0 & \cdots & 0 & 1 \end{vmatrix} =$$

$$= (-1)^n \lambda^n \left(c\left(\sum_i x_i^2 + 1\right) - \lambda\right)$$

где

$$c = \left(F'(S_j)\right)^2 + (y_j - t_j)F''(S_j).$$

Тогда

$$\lambda_j = c\left(\sum_i x_i^2 + 1\right) = \left(\left(F'(S_j)\right)^2 + \left(y_j - t_j\right)F''(S_j)\right) \times$$
$$\times\left(\sum_i x_i^2 + 1\right)$$

является наибольшим собственным значением $\nabla^2 E_j(t)$.

Поэтому, норма матрицы $\nabla^2 E_j(t)$, связанная с метрикой $\left\|\bar{x}\right\| = \sqrt{(\bar{x}, \bar{x})}$ определяется соотношением

$$\left\|\nabla^2 E_j(t)\right\| = \left(\left(F'(S_j)\right)^2 + \left(y_j - t_j\right)F''(S_j)\right)\left(\sum_i x_i^2 + 1\right).$$

Следовательно, для такой же нормы матрицы $\nabla^2 E_S(t)$, учитывая (13), выполняется

$$\left\|\nabla^2 E_S(t)\right\| = \max_j \left\|\nabla^2 E_j(t)\right\| = \left(\sum_i x_i^2 + 1\right) \times$$
$$\times \max_j \left(\left(F'(S_j)\right)^2 + \left(y_j - t_j\right)F''(S_j)\right).$$

Учитывая, что для скалярного произведения верно неравенство

$$\left(\nabla^2 E_S(t) \cdot \nabla E_S(t),\ \nabla E_S(t)\right) \le \left\|\nabla^2 E_S(t)\right\| \cdot \left\|\nabla E_S(t)\right\|^2,$$

имеем

$$\alpha(t) = \frac{\left\|\nabla E_S(t)\right\|^2}{\left(\nabla^2 E_S(t) \cdot \nabla E_S(t),\ \nabla E_S(t)\right)} \ge$$
$$\ge \frac{\left\|\nabla E_S(t)\right\|^2}{\left\|\nabla^2 E_S(t)\right\| \cdot \left\|\nabla E_S(t)\right\|^2} = \frac{1}{\left\|\nabla^2 E_S(t)\right\|}.$$

Таким образом, имеет место следующая оценка адаптивного шага обучения:

$$\alpha(t) \ge \frac{1}{\left(\sum_i x_i^2 + 1\right) \cdot \max_j \left(\left(F'(S_j)\right)^2 + \left(y_j - t_j\right)F''(S_j)\right)} \tag{14}$$

Рассмотрим случай группового обучения. Введем обозначения

$$\nabla^2 E_j^k(t) = \left(\left(F'(S_j^k)\right)^2 + \left(y_j^k - t_j^k\right)F''(S_j^k)\right) \times$$
$$\times \begin{pmatrix} \left(x_1^k\right)^2 & x_1^k x_2^k & \cdots & x_1^k x_n^k & -x_1^k \\ x_2^k x_1^k & \left(x_2^k\right)^2 & \cdots & x_2^k x_n^k & -x_2^k \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^k x_1^k & x_n^k x_2^k & \cdots & \left(x_n^k\right)^2 & -x_n^k \\ -x_1^k & -x_2^k & \cdots & -x_n^k & 1 \end{pmatrix}.$$

Тогда, учитывая, что

$$\frac{\partial^2 E_S}{\partial \omega_{ij} \partial \omega_{qj}} = \sum_k \left(\left(F'(S_j^k)\right)^2 + \left(y_j^k - t_j^k\right)F''(S_j^k)\right) x_i^k x_q^k,$$

$$\frac{\partial^2 E_S}{\partial \omega_{ij} \partial T_j} = -\sum_k \left(\left(F'(S_j^k)\right)^2 + \left(y_j^k - t_j^k\right)F''(S_j^k)\right) x_i^k,$$

$$\frac{\partial^2 E_S}{\partial T_j^2} = \sum_k \left(\left(F'(S_j^k)\right)^2 + \left(y_j^k - t_j^k\right)F''(S_j^k)\right),$$

имеем:

$$\nabla^2 E_j(t) = \sum_k \nabla^2 E_j^k(t).$$

Так как для любого k выполняется $\nabla^2 E_j^k(t) \ge 0$, то

$$\max_k \left\|\nabla^2 E_j^k(t)\right\| \le \left\|\nabla^2 E_j(t)\right\| \le \sum_k \left\|\nabla^2 E_j^k(t)\right\|.$$

Следовательно,

$$\left\|\nabla^2 E_j(t)\right\| \le \sum_k \left(\left(F'(S_j^k)\right)^2 + \left(y_j^k - t_j^k\right)F''(S_j^k)\right)\left(\sum_i \left(x_i^k\right)^2 + 1\right)$$

$$\left\|\nabla^2 E_j(t)\right\| \ge \max_k \left(\left(F'(S_j^k)\right)^2 + \left(y_j^k - t_j^k\right)F''(S_j^k)\right)\left(\sum_i \left(x_i^k\right)^2 + 1\right)$$

Тогда, учитывая (13), имеем

$$\left\|\nabla^2 E_S(t)\right\| = \max_j \left\|\nabla^2 E_j(t)\right\| \le$$
$$\le \max_j \sum_k \left(\left(F'(S_j^k)\right)^2 + \left(y_j^k - t_j^k\right)F''(S_j^k)\right)\left(\sum_i \left(x_i^k\right)^2 + 1\right)$$

и

$$\left\|\nabla^2 E_S(t)\right\| = \max_j \left\|\nabla^2 E_j(t)\right\| \ge$$
$$\ge \max_j \max_k \left(\left(F'(S_j^k)\right)^2 + \left(y_j^k - t_j^k\right)F''(S_j^k)\right)\left(\sum_i \left(x_i^k\right)^2 + 1\right).$$

Учитывая, что

$$\alpha(t) = \frac{\left\|\nabla E_S(t)\right\|^2}{\left(\nabla^2 E_S(t) \cdot \nabla E_S(t),\ \nabla E_S(t)\right)} \ge$$
$$\ge \frac{\left\|\nabla E_S(t)\right\|^2}{\left\|\nabla^2 E_S(t)\right\| \cdot \left\|\nabla E_S(t)\right\|^2} = \frac{1}{\left\|\nabla^2 E_S(t)\right\|}$$

имеем

$$\alpha(t) \ge \frac{1}{\left\|\nabla^2 E_S(t)\right\|} \ge$$
$$\ge \frac{1}{\max_j \sum_k \left(\left(F'(S_j^k)\right)^2 + \left(y_j^k - t_j^k\right)F''(S_j^k)\right)\left(\sum_i {x_i^k}^2 + 1\right)}$$

Теорема 2. Для квазиоптимального шага обучения, определяемого соотношением (9), верна следующая оценка снизу:

$$\alpha(t) \ge$$
$$\ge \frac{1}{\max_j \sum_k \left(\left(F'(S_j^k)\right)^2 + \left(y_j^k - t_j^k\right)F''(S_j^k)\right)\left(\sum_i \left(x_i^k\right)^2 + 1\right)} \tag{15}$$

Полученные оценки (14) и (15) квазиоптимального шага обучения нейронной сети с использованием метода наискорейшего спуска могут быть использованы для вычисления приближенного значения шага обучения, а также дают возможность исследовать сходимость процедуры обучения нейронной сети [2].

СПИСОК ИПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Головко В.А. Нейроинтеллект: теория и применение. Книга 1: Организация и обучение нейронных сетей с прямыми и обратными связями. – Брест: Изд. БПИ, 1999. – 264 с.
2. Головко В.А., Махнист Л.П. Модификации алгоритмов обучения линейных нейронных сетей // Вестник Брестского государственного технического университета. – Брест: БГТУ, 2000. – № 4 (4): Машиностроение, автоматизация, ЭВМ.

УДК 513.82

Андреев А.С.

ПОСТРОЕНИЕ КАНОНИЧЕСКОГО ЛИФТА ПОДМНОГООБРАЗИЯ ОДНОРОДНОГО ПРОСТРАНСТВА В СТРУКТУРНУЮ ГРУППУ ЛИ И В ЕЕ АЛГЕБРУ ЛИ

Приведём основные факты теории вычислительного аппарата метода построения канонического репера [1], [2].

Рассмотрим множество Q_1 всех n-мерных подпространств, касательных к M в точке $\pi(e)$. Наряду с множеством Q_1 рассмотрим множество $Z_1 = \{d\pi_e^{-1}(😐) | 😐 \in Q_1\}$. Множество Q_1 является H-пространством. Множество Z_1 также является H-пространством, причем действие группы H в Z_1 индуцируется присоединенным представлением Ad. H-пространства Q_1 и Z_1 изоморфны [6]. Отсюда, в частности, следует, что

$$H_1 = \{a \in H | Ad\, a(😐'_1) = 😐'_1\}. \qquad (1)$$

Пусть $\{\omega^1, \omega^2, ..., \omega^r\}$ – базис пространства ☺*, дуального к алгебре Ли ☺ группы Ли G, $\{\omega^1, \omega^2, ..., \omega^t\}$ – базис пространства $😐'^{*}_1$, дуального к $😐'_1$, $\{\omega^1, \omega^2, ..., \omega^s\}$ – базис пространства ☟*, дуального к ☟. При этом $t = s + n$. Тогда система Пфаффа, определяющая пространство $😐'_1$, будет иметь вид:

$$\omega^{t+1} = 0,\ \omega^{t+2} = 0, ...,\ \omega^r = 0. \qquad (2)$$

Найдем внешние дифференциалы форм системы (2):

$$d\omega^{t+1} = 0,\ d\omega^{t+2} = 0, ...,\ d\omega^r = 0. \qquad (3)$$

Введем индексы суммирования:

$i, j = 1,2,...,r$; $\sigma, \tau = s+1, s+2,...,r$;

$\varepsilon, \mu = t+1, t+2,...,r$;

$a, b, c = 1,2,..., s_1$; $\alpha, \beta, \gamma = 1,2,...,t$;

$p, q, l = s_1+1, s_1+2,...,s$; $\rho, \delta = s+1, s+2,...,t$.

Разложим внешние дифференциалы (3) по базису:

$$d\omega^{t+1} = \Lambda^{t+1}_{ij}\omega^i \wedge \omega^j, ..., d\omega^r = \Lambda^r_{ij}\omega^i \wedge \omega^j. \qquad (4)$$

Предположим, что подмногообразие $(☟_0, f)$ продолжается в пространство $M_1 = G/H_1$ и $f_1 : ☟_0 \to M_1$ – соответствующее продолжение, $😐_1 = T_{f(0)}(✋Ɒ)$, $😐'_1 = d\pi_e^{-1}(😐_1)$, $😐_2 = T_{f_1(0)}(✋Ɒ)$, $😐'_2 = d\pi^{-1}_{1|e}(😐_2)$.

Лемма 1 [1], [2]. В формулах (4) равны нулю коэффициенты $\Lambda^{t+1}_{a\alpha}, ..., \Lambda^r_{a\alpha}$; $\Lambda^{t+1}_{p,q}, ..., \Lambda^r_{p,q}$.

Следствие 1 [1], [2]. Система форм

$$\begin{cases} \omega^{t+1} = 0, ..., \omega^r = 0; \\ d\omega^{t+1} = 0, ..., d\omega^r = 0 \end{cases} \qquad (5)$$

эквивалентна системе

$$\omega^{t+1} = 0, ..., \omega^r = 0; \quad \text{(а)}$$
$$\omega^\rho \wedge \theta^{t+1}_\rho = 0, ..., \omega^\rho \wedge \theta^r_\rho = 0. \text{(б)} \qquad (6)$$

Пусть ☟$_1$ – алгебра Ли группы H_1, тогда

$$☟_1 = \{v \in ☟ | [v, 😐'_1] \subset 😐'_1\} \qquad (7)$$

Пусть $😐'_1 = ☟ \oplus N$, $😐'_2 = ☟_1 \oplus N$.

Теорема 1 [1], [2]. Если выполняется условие

$$[N, N] \subset 😐'_1, \qquad (8)$$

то внешние дифференциалы $d\omega^{t+1}$,..., $d\omega^r$ обращаются в нуль в пространстве $😐'_2$.

Заметим, что условие (8) всегда выполняется для одномерных подмногообразий ($n=1$), а также для подмногообразий любой размерности в случае, когда группа Ли G является полупрямым произведением группы стационарности точки пространства M и абелевой группы, в частности, для всех евклидовых и псевдоевклидовых пространств.

Используя лемму Картана, систему (6,б) на пространстве $😐'_2$ можно переписать в виде [1], [2]: $\theta^{t+1}_\rho = A^{t+1}_{\rho\delta}\omega^\delta$,..., $\theta^r_\rho = A^r_{\rho\delta}\omega^\delta$,

а систему (5) в виде:

$$\omega^{t+1} = 0, ...,\ \omega^r = 0;$$
$$\Omega^{t+1}_\rho \equiv \theta^{t+1}_\rho - A^{t+1}_{\rho\delta}\omega^\delta = 0, ...,\ \Omega^r_\rho \equiv \theta^r_\rho - A^r_{\rho\delta}\omega^\delta = 0. \qquad (9)$$

Теорема 2 [1], [2]. Система 1-форм

$$\omega^{t+1} = 0, ...,\ \omega^r = 0,\ \Omega^{t+1}_\rho, ..., \Omega^r_\rho \qquad (10)$$

есть система форм Пфаффа, определяющая подпространство $😐'_2$.

Андреев Алексей Сергеевич. *Студент V курса математического факультета Брестского государственного технического университета.*
Беларусь, 224665, г. Брест, бульвар Космонавтов, 21.