УДК 004.021:032.26

ИСПОЛЬЗОВАНИЕ АЛГОРИТМОВ ОБУЧЕНИЯ ОДНОСЛОЙНОЙ СЕТИ ДЛЯ МНОГОСЛОЙНЫХ НЕЙРОННЫХ СЕТЕЙ ПРЯМОГО РАСПРОСТРАНЕНИЯ

Л. П. Махнист[1], В. А. Головко[2], И. И. Гладкий[3], Т. И. Каримова[4]

[1] *К. т. н., доцент, заведующий кафедрой высшей математики УО «Брестский государственный технический университет», Брест, Беларусь*
[2] *Д. т. н., профессор, заведующий кафедрой интеллектуальных информационных технологий УО «Брестский государственный технический университет», Брест, Беларусь*
[3] *Старший преподаватель кафедры высшей математики УО «Брестский государственный технический университет», Брест, Беларусь*
[4] *К. физ.-мат. н., доцент кафедры высшей математики УО «Брестский государственный технический университет», Брест, Беларусь*

Реферат

В работе рассматриваются различные подходы к выбору шага обучения нейронной сети прямого распространения, производится их сравнительный анализ с точки зрения сходимости алгоритма обучения с использованием метода наискорейшего спуска. Получены формулы для вычисления шага обучения и ограничения для их использования. Предложенная методика может быть использована в алгоритме обратного распространения ошибки обучения нейронной сети.

Ключевые слова: нейронная сеть прямого распространения, алгоритм обратного распространения ошибки, метод наискорейшего спуска.

ESTIMATION OF THE CONVERGENCE RATE AND THE CHOICE OF THE LEARNING STEP OF ARTIFICIAL FEED FORWARD NEURAL NETWORKS

L. P. Makhnist, V. A. Golovko, I. I. Hladki, T. I. Karimova

Abstract

The paper presents various approaches to the choice of the learning step for a feed forward neural network and compares then in terms of the learning algorithm convergence using the steepest descent approach. The equations for calculating the learning step and the restrictions for their use are obtained. The proposed technique can be used in the back propagation algorithm for training a multilayer perceptron.

Keywords: the feedforward neural network, the backpropagation algorithm, the steepest descent method.

Ведение

В работе рассматривается многослойная нейронная сеть прямого распространения. Предлагается после вычисления весовых коэффициентов и порогов последнего слоя вычислять ожидаемые выходы нейронов предыдущего слоя, а затем производить изменение весовых коэффициентов и порогов предыдущего слоя.

Задача обучения нейронной сети прямого распространения состоит в нахождении весовых коэффициентов w_{ij} и порогов T_j нейронной сети, которые минимизируют функцию ошибки сети

$$E(w_{11}, w_{21}, \ldots, w_{m1}, T_1, w_{12}, w_{22}, \ldots, w_{m2}, T_2, \ldots, w_{1n}, w_{2n}, w_{mn}, T_n) = \frac{1}{2}\sum_{j=1}^{n}\left(y_j - t_j\right)^2,$$

где $y_j = F\left(S_j\right)$ – значение функции активации j -го выходного нейрона сети, $S_j = \sum_{i=1}^{m} w_{ij}x_i - T_j$, x_i – выходное значение i -го нейрона предыдущего слоя, t_j – ожидаемый выход j -го выходного нейрона ($i = \overline{1, m}$, $j = \overline{1, n}$).

В работе рассматриваются различные подходы к выбору шага обучения нейронной сети с использованием метода наискорейшего спуска.

Выбор шага обучения весовых коэффициентов и ожидаемых выходов нейронов

Введем обозначения:

$$\overline{W} = \left(w_{11}, w_{21}, \ldots, w_{m1}, T_1, w_{12}, w_{22}, \ldots, w_{m2}, T_2, \ldots, w_{1n}, w_{2n}, \ldots, w_{mn}, T_n\right)^T$$

– вектор-столбец весовых коэффициентов w_{ij} и порогов T_j нейронной сети, а $\overline{W}_j = \left(w_{1j}, w_{2j}, \ldots, w_{mj}, T_j\right)^T$ – вектор-столбец весовых коэффициентов w_{ij} и порога T_j, связанных с j -ым выходным нейроном сети, $E\left(\overline{W}\right) = \frac{1}{2}\sum_{j=1}^{n}\left(y_j - t_j\right)^2$ – функция ошибки сети, $E\left(\overline{W}_j\right) = \frac{1}{2}\left(y_j - t_j\right)^2$ – функция ошибки j -го выходного нейрона сети.

Очевидно, что функция ошибки сети $E\left(\overline{W}\right)=\frac{1}{2}\sum_{j=1}^{n}\left(y_j-t_j\right)^2$ и функции ошибок выходных нейронов сети связаны соотношением $E\left(\overline{W}\right)=\sum_{j=1}^{n}E\left(\overline{W}_j\right)$.

Предполагается, что функция активации нейронной сети F является дважды дифференцируемой функцией.

Обучение нейронной сети с использованием метода наискорейшего спуска состоит в изменении весовых коэффициентов w_{ij} и порогов T_j нейронной сети на каждом шаге обучения $(t+1)$ ($t=1,2,\dots$) в соответствии с формулами:

$$w_{ij}(t+1)=w_{ij}(t)-\alpha_j(t)\frac{\partial E\left(\overline{W}_j(t)\right)}{\partial w_{ij}}, \qquad (1)$$

$$T_j(t+1)=T_j(t)-\alpha_j(t)\frac{\partial E\left(\overline{W}_j(t)\right)}{\partial T_j}, \qquad (2)$$

если шаг обучения $\alpha_j(t)$ выбирается только для весовых коэффициентов w_{ij} и порога T_j, связанных только с j-ым выходным нейроном сети, для минимизации функции ошибки сети j-го выходного нейрона $E\left(\overline{W}_j\right)$;

$$w_{ij}(t+1)=w_{ij}(t)-\alpha(t)\frac{\partial E\left(\overline{W}(t)\right)}{\partial w_{ij}}, \qquad (3)$$

$$T_j(t+1)=T_j(t)-\alpha(t)\frac{\partial E\left(\overline{W}(t)\right)}{\partial T_j}, \qquad (4)$$

если шаг обучения $\alpha(t)$ выбирается для всех весовых коэффициентов w_{ij} и порогов T_j для минимизации функции ошибки сети $E\left(\overline{W}\right)$ (например, в [1–3]).

Частные производные первого порядка функции ошибки сети j-го выходного нейрона $E\left(\overline{W}_j\right)$ по соответствующим переменным w_{ij} и T_j определяются соотношениями:

$$\frac{\partial E\left(\overline{W}_j\right)}{\partial w_{ij}}=\left(y_j-t_j\right)F'\left(S_j\right)x_i,$$

$$\frac{\partial E\left(\overline{W}_j\right)}{\partial T_j}=-\left(y_j-t_j\right)F'\left(S_j\right) \ (i=\overline{1,m}).$$

Очевидно, что частные производные первого порядка функции ошибки сети $E\left(\overline{W}\right)$ по соответствующим переменным w_{ij} и T_j удовлетворяют соотношениям:

$$\frac{\partial E\left(\overline{W}\right)}{\partial w_{ij}}=\frac{\partial E\left(\overline{W}_j\right)}{\partial w_{ij}}=\left(y_j-t_j\right)F'\left(S_j\right)x_i,$$

$$\frac{\partial E\left(\overline{W}\right)}{\partial T_j}=\frac{\partial E\left(\overline{W}_j\right)}{\partial T_j}=-\left(y_j-t_j\right)F'\left(S_j\right),$$

($i=\overline{1,m}$, $j=\overline{1,n}$).

Тогда соотношения (1) и (2) можно записать в виде

$$w_{ij}(t+1)=w_{ij}(t)-\alpha_j(t)\left(y_j(t)-t_j\right)F'\left(S_j(t)\right)x_i, \qquad (5)$$

$$T_j(t+1)=T_j(t)+\alpha_j(t)\left(y_j(t)-t_j\right)F'\left(S_j(t)\right), \qquad (6)$$

а соотношения (3) и (4) в виде:

$$w_{ij}(t+1)=w_{ij}(t)-\alpha(t)\left(y_j(t)-t_j\right)F'\left(S_j(t)\right)x_i, \qquad (7)$$

$$T_j(t+1)=T_j(t)+\alpha(t)\left(y_j(t)-t_j\right)F'\left(S_j(t)\right). \qquad (8)$$

Введем обозначения:

$$\nabla E\left(\overline{W}_j\right)=\left(\frac{\partial E\left(\overline{W}_j\right)}{\partial w_{1j}},\frac{\partial E\left(\overline{W}_j\right)}{\partial w_{2j}},\dots,\frac{\partial E\left(\overline{W}_j\right)}{\partial w_{mj}},\frac{\partial E\left(\overline{W}_j\right)}{\partial T_j}\right)^T$$

– градиент функции ошибки сети j-го выходного нейрона $E\left(\overline{W}_j\right)$.

Заметим, что

$$\nabla E\left(\overline{W}_j\right)=\left(y_j-t_j\right)F'\left(S_j\right)\cdot\left(x_1,\ x_2,\ \dots\ x_m,\ -1\right)^T.$$

Тогда соотношения (1), (2) и (3), (4) можно, соответственно, записать в виде:

$$\overline{W}_j(t+1)=\overline{W}_j(t)-\alpha_j(t)\nabla E\left(\overline{W}_j(t)\right), \qquad (9)$$

$$\overline{W}_j(t+1)=\overline{W}_j(t)-\alpha(t)\nabla E\left(\overline{W}_j(t)\right). \qquad (10)$$

Рассмотрим случай, когда шаг обучения $\alpha_j(t)$ выбирается только для весовых коэффициентов w_{ij} и порога T_j, связанных только с j-ым выходным нейроном сети для минимизации функции ошибки сети j-го выходного нейрона $E\left(\overline{W}_j\right)$.

Введем обозначения $\Delta w_{ij}(t)=w_{ij}(t+1)-w_{ij}(t)$, $\Delta T_j(t)=T_j(t+1)-T_j(t)$ ($i=\overline{1,m}$, $j=\overline{1,n}$).

Заметим, что разложение функции $E\left(\overline{W}_j(t+1)\right)$ в ряд Тейлора в точке $\overline{W}_j(t)$, с точностью до частных производных второго порядка, вычисляемых, как и частные производные первого порядка, на шаге t, может быть представлено в виде:

$$E\left(\overline{W}_j(t+1)\right)=E\left(\overline{W}_j(t)\right)+\left(\nabla E\left(\overline{W}_j(t)\right),\Delta\overline{W}_j(t)\right)+$$
$$+\frac{1}{2}\left(\nabla^2E\left(\overline{W}_j(t)\right)\cdot\Delta\overline{W}_j(t),\Delta\overline{W}_j(t)\right),$$

где $\nabla^2E\left(\overline{W}_j(t)\right)\cdot\Delta\overline{W}_j(t)$ – произведение матрицы Гессе $\nabla^2E\left(\overline{W}_j(t)\right)$ функции $E\left(\overline{W}_j(t)\right)$ и вектора $\Delta\overline{W}_j(t)$, $\left(\nabla E\left(\overline{W}_j(t)\right),\Delta\overline{W}_j(t)\right)$ – скалярное произведение вектора $\nabla E\left(\overline{W}_j(t)\right)$ и вектора $\Delta\overline{W}_j(t)=\overline{W}_j(t+1)-\overline{W}_j(t)$, а $\left(\nabla^2E\left(\overline{W}_j(t)\right)\cdot\Delta\overline{W}_j(t),\Delta\overline{W}_j(t)\right)$ – скалярное произведение вектора $\nabla^2E\left(\overline{W}_j(t)\right)\cdot\Delta\overline{W}_j(t)$ и вектора $\Delta\overline{W}_j(t)$.

Тогда, учитывая (9), получим

$$E\left(\overline{W}_j(t+1)\right)=E\left(\overline{W}_j(t)\right)-\alpha_j(t)\left(\nabla E\left(\overline{W}_j(t)\right),\nabla E\left(\overline{W}_j(t)\right)\right)+$$
$$+\frac{\alpha_j^2(t)}{2}\left(\nabla^2E\left(\overline{W}_j(t)\right)\cdot\nabla E\left(\overline{W}_j(t)\right),\nabla E\left(\overline{W}_j(t)\right)\right).$$

Рассмотрим функцию $E\left(\overline{W}_j(t+1)\right)=E\left(\alpha_j(t)\right)$ как функцию аргумента $\alpha_j(t)$.

Заметим, что производная этой функции равна

$$E'\left(\alpha_j(t)\right)=-\left(\nabla E\left(\overline{W}_j(t)\right),\nabla E\left(\overline{W}_j(t)\right)\right)+$$

$$+\alpha_j(t)\left(\nabla^2 E\left(\overline{W}_j(t)\right)\cdot\nabla E\left(\overline{W}_j(t)\right),\nabla E\left(\overline{W}_j(t)\right)\right).$$

Предположим, что вторая производная функции $E\left(\alpha_j(t)\right)$ удовлетворяет неравенству

$$E''\left(\alpha_j(t)\right)=\left(\nabla^2 E\left(\overline{W}_j(t)\right)\cdot\nabla E\left(\overline{W}_j(t)\right),\nabla E\left(\overline{W}_j(t)\right)\right)>0 \text{ [4].}$$

Тогда наименьшее значение функции $E\left(\alpha_j(t)\right)$ достигается, если для шага обучения $\alpha_j(t)$ выполняется соотношение

$$\alpha_j(t)=\frac{\left\|\nabla E\left(\overline{W}_j(t)\right)\right\|^2}{\left(\nabla^2 E\left(\overline{W}_j(t)\right)\cdot\nabla E\left(\overline{W}_j(t)\right),\nabla E\left(\overline{W}_j(t)\right)\right)},\quad(11)$$

где $\left\|\nabla E\left(\overline{W}_j(t)\right)\right\|$ – длина вектора градиента $\nabla E\left(\overline{W}_j(t)\right)$, связанная со скалярным произведением $\left(\nabla E\left(\overline{W}_j(t)\right),\nabla E\left(\overline{W}_j(t)\right)\right)$.

Так как

$$\nabla E\left(\overline{W}_j\right)=\left(y_j-t_j\right)F'\left(S_j\right)\cdot\left(x_1,\quad x_2,\quad\ldots\quad x_m,\quad -1\right)^T,$$

то $\left\|\nabla E\left(\overline{W}_j\right)\right\|^2=\left(E'\left(\overline{W}_j\right)\right)^2\cdot\left(\sum\limits_{i=1}^{m}x_i^2+1\right)$, где

$$E'\left(\overline{W}_j\right)=\left(y_j-t_j\right)F'\left(S_j\right).$$

Частные производные второго порядка функции ошибки сети j-го выходного нейрона $E\left(\overline{W}_j\right)$ по соответствующим переменным определяются соотношениями:

$$\frac{\partial^2 E\left(\overline{W}_j\right)}{\partial w_{ij}\partial w_{kj}}=\left(\left(F'\left(S_j\right)\right)^2+\left(y_j-t_j\right)F''\left(S_j\right)\right)\cdot x_i x_k,$$

$$\frac{\partial^2 E\left(\overline{W}_j\right)}{\partial w_{ij}\partial T_j}=-\left(\left(F'\left(S_j\right)\right)^2+\left(y_j-t_j\right)F''\left(S_j\right)\right)\cdot x_i,$$

$$\frac{\partial^2 E\left(\overline{W}_j\right)}{\partial T_j^2}=\left(F'\left(S_j\right)\right)^2+\left(y_j-t_j\right)F''\left(S_j\right),$$

($i=\overline{1,m}$, $j=\overline{1,n}$).

Матрицу Гессе $\nabla^2 E\left(\overline{W}_j\right)$ можно записать в виде:

$$\nabla^2 E\left(\overline{W}_j\right)=E''\left(\overline{W}_j\right)\cdot\begin{pmatrix} x_1^2 & x_1x_2 & \ldots & x_1x_m & -x_1\\ x_2x_1 & x_2^2 & \ldots & x_2x_m & -x_2\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ x_mx_1 & x_mx_2 & \ldots & x_m^2 & -x_m\\ -x_1 & -x_2 & \ldots & -x_m & 1\end{pmatrix},$$

где $E''\left(\overline{W}_j\right)=\left(F'\left(S_j\right)\right)^2+\left(y_j-t_j\right)F''\left(S_j\right)$.

Следовательно, скалярное произведение вектора $\nabla^2 E\left(\overline{W}_j\right)\cdot\nabla E\left(\overline{W}_j\right)$ и вектора $\nabla E\left(\overline{W}_j\right)$ определяется соотношением:

$$\left(\nabla^2 E\left(\overline{W}_j\right)\cdot\nabla E\left(\overline{W}_j\right),\nabla E\left(\overline{W}_j\right)\right)=$$

$$=\left(E'\left(\overline{W}_j\right)\right)^2\cdot E''\left(\overline{W}_j\right)\cdot\left(\sum_{i=1}^{m}x_i^2+1\right)^2.$$

Тогда соотношение (11) можно записать в виде:

$$\alpha_j(t)=\frac{1}{\left(\left(F'\left(S_j(t)\right)\right)^2+\left(y_j(t)-t_j\right)F''\left(S_j(t)\right)\right)\cdot\left(\sum\limits_{i=1}^{m}x_i^2+1\right)}.\quad(12)$$

Таким образом, если шаг обучения $\alpha_j(t)$ выбирается только для весовых коэффициентов w_{ij} и порога T_j, связанных только с j-ым выходным нейроном сети, для минимизации функции ошибки сети j-го выходного нейрона $E\left(\overline{W}_j\right)$, то изменение весовых коэффициентов w_{ij} и порогов T_j нейронной сети на каждом шаге обучения производится в соответствии с формулами:

$$w_{ij}(t+1)=w_{ij}(t)-$$

$$-\frac{\left(y_j(t)-t_j\right)F'\left(S_j(t)\right)\cdot x_i}{\left(\sum\limits_{i=1}^{m}x_i^2+1\right)\cdot\left(\left(F'\left(S_j(t)\right)\right)^2+\left(y_j(t)-t_j\right)F''\left(S_j(t)\right)\right)},\quad(13)$$

$$T_j(t+1)=T_j(t)+$$

$$+\frac{\left(y_j(t)-t_j\right)F'\left(S_j(t)\right)}{\left(\sum\limits_{i=1}^{m}x_i^2+1\right)\cdot\left(\left(F'\left(S_j(t)\right)\right)^2+\left(y_j(t)-t_j\right)F''\left(S_j(t)\right)\right)}.\quad(14)$$

Теперь рассмотрим случай, когда шаг обучения $\alpha(t)$ выбирается для всех весовых коэффициентов w_{ij} и порогов T_j для минимизации функции ошибки сети $E\left(\overline{W}\right)$.

Разложение функции $E\left(\overline{W}(t+1)\right)$ в ряд Тейлора в точке $\overline{W}(t)$, с точностью до частных производных второго порядка, аналогично как и для функций $E\left(\overline{W}_j(t+1)\right)$ может быть представлено в виде:

$$E\left(\overline{W}(t+1)\right)=E\left(\overline{W}(t)\right)+\left(\nabla E\left(\overline{W}(t)\right),\Delta\overline{W}(t)\right)+$$

$$+\frac{1}{2}\left(\nabla^2 E\left(\overline{W}(t)\right)\cdot\Delta\overline{W}(t),\Delta\overline{W}(t)\right),$$

где $\nabla E\left(\overline{W}\right)=\left(\nabla E\left(\overline{W}_1\right)^T,\nabla E\left(\overline{W}_2\right)^T,\ldots,\nabla E\left(\overline{W}_n\right)^T\right)^T$ – градиент функция ошибки сети $E\left(\overline{W}\right)$,

$\Delta\overline{W}=\left(\Delta\overline{W}_1^{\,T},\Delta\overline{W}_2^{\,T},\ldots,\Delta\overline{W}_n^{\,T}\right)^T$, $\nabla^2 E\left(\overline{W}(t)\right)$ – матрица Гессе $\nabla^2 E\left(\overline{W}\right)$ функции $E\left(\overline{W}\right)$.

doi.org/10.36773/1818-1212-2020-123-5-32-37

Так как $\dfrac{\partial^2 E\left(\overline{W}\right)}{\partial w_{kj}\partial w_{ij}} = \dfrac{\partial^2 E\left(\overline{W}\right)}{\partial w_{ij}\partial w_{kj}} = \dfrac{\partial^2 E\left(\overline{W}_j\right)}{\partial w_{ij}\partial w_{kj}} =$

$$= \left(\left(F'\left(S_j\right)\right)^2 + \left(y_j - t_j\right)F''\left(S_j\right)\right)\cdot x_i x_k,$$

$$\frac{\partial^2 E\left(\overline{W}\right)}{\partial T_j \partial w_{ij}} = \frac{\partial^2 E\left(\overline{W}\right)}{\partial w_{ij}\partial T_j} = \frac{\partial^2 E\left(\overline{W}_j\right)}{\partial w_{ij}\partial T_j} =$$

$$= -\left(\left(F'\left(S_j\right)\right)^2 + \left(y_j - t_j\right)F''\left(S_j\right)\right)\cdot x_i,$$

$$\frac{\partial^2 E\left(\overline{W}\right)}{\partial T_j^2} = \frac{\partial^2 E\left(\overline{W}_j\right)}{\partial T_j^2} = \left(F'\left(S_j\right)\right)^2 + \left(y_j - t_j\right)F''\left(S_j\right)$$

($i = \overline{1, m}$, $k = \overline{1, m}$, $j = \overline{1, n}$),

а другие частные производные второго порядка функции $E\left(\overline{W}\right)$ равны нулю, то матрица $\nabla^2 E\left(\overline{W}\right)$ может быть записана в виде блочно-диагональной (квазидиагональной) матрицы, на главной диагонали которой расположены матрицы $\nabla^2 E\left(\overline{W}_j\right)$ ($j = \overline{1, n}$).

Тогда, учитывая, что

$E\left(\overline{W}(t)\right) = \sum_{j=1}^{n} E\left(\overline{W}_j(t)\right)$ и

$\left(\nabla E\left(\overline{W}(t)\right), \Delta\overline{W}(t)\right) = \sum_{j=1}^{n}\left(\nabla E\left(\overline{W}_j(t)\right), \Delta\overline{W}_j(t)\right)$, получим соотношение:

$$E\left(\overline{W}(t+1)\right) = \sum_{j=1}^{n} E\left(\overline{W}_j(t)\right) + \sum_{j=1}^{n}\left(\nabla E\left(\overline{W}_j(t)\right), \Delta\overline{W}_j(t)\right) +$$

$$+\frac{1}{2}\sum_{j=1}^{n}\left(\nabla^2 E\left(\overline{W}_j(t)\right)\cdot \Delta\overline{W}_j(t), \Delta\overline{W}_j(t)\right).$$

Учитывая (10), получим

$$E\left(\overline{W}(t+1)\right) = \sum_{j=1}^{n} E\left(\overline{W}_j(t)\right) - \alpha(t)\sum_{j=1}^{n}\left(\nabla E\left(\overline{W}_j(t)\right), \nabla E\left(\overline{W}_j(t)\right)\right) +$$

$$+\frac{\alpha^2(t)}{2}\sum_{j=1}^{n}\left(\nabla^2 E\left(\overline{W}_j(t)\right)\cdot \nabla E\left(\overline{W}_j(t)\right), \nabla E\left(\overline{W}_j(t)\right)\right).$$

Рассмотрим функцию $E\left(\overline{W}(t+1)\right) = E\left(\alpha(t)\right)$ как функцию аргумента $\alpha(t)$.

Заметим, что производная этой функции равна

$$E'\left(\alpha_j(t)\right) = -\sum_{j=1}^{n}\left(\nabla E\left(\overline{W}_j(t)\right), \nabla E\left(\overline{W}_j(t)\right)\right) +$$

$$+\alpha(t)\sum_{j=1}^{n}\left(\nabla^2 E\left(\overline{W}_j(t)\right)\cdot \nabla E\left(\overline{W}_j(t)\right), \nabla E\left(\overline{W}_j(t)\right)\right).$$

Предположим, что вторая производная функции $E\left(\alpha(t)\right)$ удовлетворяет неравенству [4]

$$E''\left(\alpha(t)\right) = \sum_{j=1}^{n}\left(\nabla^2 E\left(\overline{W}_j(t)\right)\cdot \nabla E\left(\overline{W}_j(t)\right), \nabla E\left(\overline{W}_j(t)\right)\right) > 0.$$

Тогда наименьшее значение функции $E\left(\alpha(t)\right)$ достигается, если для шага обучения $\alpha(t)$ выполняется соотношение:

$$\alpha(t) = \frac{\sum_{j=1}^{n}\left\|\nabla E\left(\overline{W}_j(t)\right)\right\|^2}{\sum_{j=1}^{n}\left(\nabla^2 E\left(\overline{W}_j(t)\right)\cdot \nabla E\left(\overline{W}_j(t)\right), \nabla E\left(\overline{W}_j(t)\right)\right)}. \quad (15)$$

Так как то соотношение (15) можно записать в виде:

$\left\|\nabla E\left(\overline{W}_j\right)\right\|^2 = \left(y_j - t_j\right)^2\left(F'\left(S_j\right)\right)^2\cdot\left(\sum_{i=1}^{m} x_i^2 + 1\right)$ и

$$\left(\nabla^2 E\left(\overline{W}_j\right)\cdot \nabla E\left(\overline{W}_j\right), \nabla E\left(\overline{W}_j\right)\right) =$$

$$= \left(y_j - t_j\right)^2\left(F'\left(S_j\right)\right)^2\left(\left(F'\left(S_j\right)\right)^2 + \left(y_j - t_j\right)F''\left(S_j\right)\right)\cdot\left(\sum_{i=1}^{m} x_i^2 + 1\right)^2.$$

При выборе шага обучения использование соотношений (12) будет обеспечивать лучшую скорость сходимости обучения нейронной сети с использованием метода наискорейшего спуска по сравнению с использованием соотношения (16) [6–8].

$$\alpha(t) = \frac{\sum_{j=1}^{n}\left(y_j(t) - t_j\right)^2\left(F'\left(S_j(t)\right)\right)^2}{\left(\sum_{i=1}^{m} x_i^2 + 1\right)\cdot\sum_{j=1}^{n}\left(y_j(t) - t_j\right)^2\left(F'\left(S_j(t)\right)\right)^2\left(\left(F'\left(S_j(t)\right)\right)^2 + \left(y_j(t) - t_j\right)F''\left(S_j(t)\right)\right)}. \quad (16)$$

В работе предлагается после вычисления весовых коэффициентов w_{ij} и порогов T_j последнего слоя вычислять ожидаемые выходы нейронов предыдущего слоя $t_i = x_i$, которые минимизируют ошибку сети как функцию, зависящую от значений x_i:

$$E\left(\overline{X}\right) = E\left(x_1, x_2, \ldots, x_m\right) = \frac{1}{2}\sum_{j=1}^{n}\left(y_j - t_j\right)^2 \text{ [9].}$$

Введем обозначение: $\overline{X} = \left(x_1, x_2, \ldots, x_m\right)^T$ – вектор-столбец выходных значений нейронов предыдущего слоя.

Вычисление ожидаемых выходов нейронов предыдущего слоя $t_i = x_i$ будем производить с использованием метода наискорейшего спуска на каждом шаге $(t+1)$ ($t = 1, 2, \ldots$) в соответствии с формулой:

$$\overline{X}(t+1) = \overline{X}(t) - \beta(t)\nabla E\left(\overline{X}(t)\right), \quad (17)$$

где $\nabla E\left(\overline{X}(t)\right) = \left(\dfrac{\partial E\left(\overline{X}(t)\right)}{\partial x_1}, \dfrac{\partial E\left(\overline{X}(t)\right)}{\partial x_2}, \ldots, \dfrac{\partial E\left(\overline{X}(t)\right)}{\partial x_m}\right)^T$

– градиент функции ошибки сети $E\left(\overline{X}(t)\right)$, где

$$\frac{\partial E\left(\overline{X}(t)\right)}{\partial x_i} = \frac{\partial\left(\frac{1}{2}\sum_{j=1}^{n}\left(y_j - t_j\right)^2\right)}{\partial x_i} =$$

$$= \sum_{j=1}^{n}\left(y_j - t_j\right)\frac{\partial\left(F\left(S_j\right) - t_j\right)}{\partial x_i} =$$

$$= \sum_{j=1}^{n} \left(y_j - t_j\right) F'\left(S_j\right) \frac{\partial S_j}{\partial x_i} = \sum_{j=1}^{n} \left(y_j - t_j\right) F'\left(S_j\right) w_{ij}$$

($i = \overline{1, m}$).

Разложение функции $E\left(\overline{X}(t+1)\right)$ в ряд Тейлора в точке $\overline{X}(t)$, с точностью до частных производных второго порядка, может быть представлено в виде:

$$E\left(\overline{X}(t+1)\right) = E\left(\overline{X}(t)\right) + \left(\nabla E\left(\overline{X}(t)\right), \Delta\overline{X}(t)\right) + \\ + \frac{1}{2}\left(\nabla^2 E\left(\overline{X}(t)\right) \cdot \Delta\overline{X}(t), \Delta\overline{X}(t)\right),$$

где $\nabla^2 E\left(\overline{X}(t)\right) \cdot \Delta\overline{X}(t)$ – произведение матрицы Гессе $\nabla^2 E\left(\overline{X}(t)\right)$ функции $E\left(\overline{X}(t)\right)$ и вектора $\Delta\overline{X}(t) = \overline{X}(t+1) - \overline{X}(t)$, а $\left(\nabla E\left(\overline{X}(t)\right), \Delta\overline{X}(t)\right)$ и $\left(\nabla^2 E\left(\overline{X}(t)\right) \cdot \Delta\overline{X}(t), \Delta\overline{X}(t)\right)$ – скалярные произведения соответствующих векторов.

Заметим, что элементы матрицы $\nabla^2 E\left(\overline{X}(t)\right)$ определяются соотношениями:

$$\frac{\partial^2 E\left(\overline{X}(t)\right)}{\partial x_k \partial x_i} = \frac{\partial^2 E\left(\overline{X}(t)\right)}{\partial x_i \partial x_k} = \\ = \sum_{j=1}^{n} \left(\left(F'\left(S_j\right)\right)^2 + \left(y_j - t_j\right) F''\left(S_j\right)\right) w_{ij} w_{kj}$$

($i = \overline{1, m}$, $k = \overline{1, m}$, $j = \overline{1, n}$).

Учитывая (17), получим

$$E\left(\overline{X}(t+1)\right) = E\left(\overline{X}(t)\right) - \beta(t)\left(\nabla E\left(\overline{X}(t)\right), \nabla\overline{X}(t)\right) + \\ + \frac{\beta^2(t)}{2}\left(\nabla^2 E\left(\overline{X}(t)\right) \cdot \nabla\overline{X}(t), \nabla\overline{X}(t)\right).$$

Рассмотрим функцию $E\left(\overline{X}(t+1)\right) = E\left(\beta(t)\right)$ как функцию аргумента $\beta(t)$.

Заметим, что производная этой функции равна

$$E'\left(\beta(t)\right) = -\left(\nabla E\left(\overline{X}(t)\right), \nabla\overline{X}(t)\right) + \\ + \beta(t)\left(\nabla^2 E\left(\overline{X}(t)\right) \cdot \nabla\overline{X}(t), \nabla\overline{X}(t)\right).$$

Предположим, что вторая производная функции $E\left(\beta(t)\right)$ удовлетворяет неравенству

$$E''\left(\beta(t)\right) = \left(\nabla^2 E\left(\overline{X}(t)\right) \cdot \nabla\overline{X}(t), \nabla\overline{X}(t)\right) > 0$$ [4].

Тогда наименьшее значение функции $E\left(\beta(t)\right)$ достигается, если для шага обучения $\beta(t)$ выполняется соотношение:

$$\beta(t) = \frac{\left(\nabla E\left(\overline{X}(t)\right), \nabla E\left(\overline{X}(t)\right)\right)}{\left(\nabla^2 E\left(\overline{X}(t)\right) \cdot \nabla E\left(\overline{X}(t)\right), \nabla E\left(\overline{X}(t)\right)\right)}. \quad (18)$$

Полученные по формулам (17) и (18) с заданной точностью значения $x_i = t_i$ – ожидаемые выходы нейронов предыдущего слоя. Далее в соответствии с (9) или (10) предлагается производить изменение весовых коэффициентов и порогов предыдущего слоя.

Заключение

В статье приведены формулы вычисления шага обучения для изменения весовых коэффициентов и порогов сети и ограничения для их использования.

Предлагается после вычисления весовых коэффициентов и порогов последнего слоя вычислять ожидаемые выходы нейронов предыдущего слоя. Приведены формулы вычисления шага и изменения ожидаемых выходов нейронов этого предыдущего слоя.

Предложенная методика может быть использована в алгоритме обратного распространения ошибки обучения нейронной сети.

Список цитированных источников

1. Golovko, M. Multilayer neural networks training methodic / M. Golovko, L. Makhnist, N. Maniakov // Second IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2003) : Proceedings, Lviv, Ukraine, 8–10 Sept. 2003. – Lviv, 2003. – P. 185–190.
2. Maniakov, N. Traing algorithm for forecasting multilayer neural network / N. Maniakov, L. Makhnist, V. Rubanov // Pattern Recognition and Information Processing : Proceedings of The Seventh International Conferences (PRIP'2003), Minsk, Republic of Belarus, 21–23 May 2003 : in 2 vol. – Minsk, 2003. – Vol. 1. – P. 26–30.
3. Makhnist, L. Some Methods of Adaptive Multilayer Neural Network Training / L. Makhnist, N. Maniakov // International Journal of Computing. – 2004. – Vol. 3. – P. 99–106.
4. Maxnist, L. Convergence Analysis of Neural Networks Training Based on Steepest Descent Method / L. Maxnist, A. Doudkin, V. Golovko // Pattern Recognition and Information Processing (PRIP'2007) : Proceedings of the Ninth International Conference, Minsk, Republic of Belarus, 22–24 May 2007 : in 2 vol. – Minsk, 2007. – Vol. 1. – P. 285–289.
5. Гладкий, И. И. О выборе шага обучения искусственных нейронных сетей прямого распространения / И. И. Гладкий, Л. П. Махнист, В. С. Рубанов, Т. Ю. Юхимук // Вычислительные методы, модели и образовательные технологии : сб. материалов VIII Междунар. науч.-практ. конф., Брест, 18 окт. 2019 г. / Брест. гос. ун-т им. А. С. Пушкина ; под общ. ред. А. А. Козинского. – Брест : БрГУ, 2019. – С. 121–122.
6. Махнист, Л. П. О сходимости алгоритмов обучения искусственных нейронных сетей / Л. П. Махнист, А. В. Санюкевич, В. П. Черненко, М. М. Юхимук // Вычислительные методы, модели и образовательные технологии : сб. материалов VIII Междунар. науч.-практ. конф., Брест, 18 окт. 2019 г. / Брест. гос. ун-т им. А. С. Пушкина ; под общ. ред. А. А. Козинского. – Брест : БрГУ, 2019. – С. 133–135.
7. Гладкий, И. И. О скорости сходимости и выборе шага обучения искусственных нейронных сетей прямого распространения / И. И. Гладкий, В. В. Будяков, Б. И. Чикалов, Т. Ю. Юхимук // Современные проблемы математики и вычислительной техники : сб. материалов XI Респ. науч. конф. молодых ученых и студентов, Брест, 21–22 нояб.

doi.org/10.36773/1818-1212-2020-123-5-32-37

2019 г. / Брест. гос. техн. ун-т ; редкол.: В. А. Головко (гл. ред.) [и др.]. – Брест : БрГТУ, 2019. – С. 11–14.
8. Махнист, Л. П. Оценки скорости сходимости и выбор шага обучения искусственных нейронных сетей прямого распространения / Л. П. Махнист, В. А. Головко, И. И. Гладкий, Т. И. Каримова // Вестник Брестского государственного технического университета. – 2019. – № 5 : Физика, математика, информатика.– С. 27–35.
9. Махнист, Л. П. Об одном подходе к обучению искусственных нейронных сетей прямого распространения / Л. П. Махнист, И. И. Гладкий, Т. И. Каримова, В. С. Рубанов // Математическое моделирование и новые образовательные технологии в математике : сб. материалов Респ. науч.-практ. конф., Брест, 23–24 апр. 2020 г. / Брест. гос. ун-т им. А. С. Пушкина ; под общ. ред. А. И. Басика. – Брест : БрГУ, 2020. – С. 41–44.

References
1. Golovko, M. Multilayer neural networks training methodic / M. Golovko, L. Makhnist, N. Maniakov // Second IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2003) : Proceedings, Lviv, Ukraine, 8–10 Sept. 2003. – Lviv, 2003. – P. 185–190.
2. Maniakov, N. Traing algorithm for forecasting multilayer neural network / N. Maniakov, L. Makhnist, V. Rubanov // Pattern Recognition and Information Processing : Proceedings of The Seventh International Conferences (PRIP'2003), Minsk, Republic of Belarus, 21–23 May 2003 : in 2 vol. – Minsk, 2003. – Vol. 1. – P. 26–30.
3. Makhnist, L. Some Methods of Adaptive Multilayer Neural Network Training / L. Makhnist, N. Maniakov // International Journal of Computing. – 2004. – Vol. 3. – P. 99–106.
4. Maxnist, L. Convergence Analysis of Neural Networks Training Based on Steepest Descent Method / L. Maxnist, A. Doudkin, V. Golovko // Pattern Recognition and Information Processing (PRIP'2007) : Proceedings of the Ninth International Conference, Minsk, Republic of Belarus, 22–24 May 2007 : in 2 vol. – Minsk, 2007. – Vol. 1. – P. 285–289.
5. Gladkij, I. I. O vybore shaga obucheniya iskusstvennyh nejronnyh setej pryamogo rasprostraneniya / I. I. Gladkij, L. P. Mahnist, V. S. Rubanov, T. YU. YUhimuk // Vychislitel'nye metody, modeli i obrazovatel'nye tekhnologii : sb. materialov VIII Mezhdunar. nauch.-prakt. konf., Brest, 18 okt. 2019 g. / Brest. gos. un-t im. A. S. Pushkina ; pod obshch. red. A. A. Kozinskogo. – Brest : BrGU, 2019. – S. 121–122.
6. Mahnist, L. P. O skhodimosti algoritmov obucheniya iskusstvennyh nejronnyh setej / L. P. Mahnist, A. V. Sanyukevich, V. P. CHernenko, M. M. YUhimuk // Vychislitel'nye metody, modeli i obrazovatel'nye tekhnologii : sb. materialov VIII Mezhdunar. nauch.-prakt. konf., Brest, 18 okt. 2019 g. / Brest. gos. un-t im. A. S. Pushkina ; pod obshch. red. A. A. Kozinskogo. – Brest : BrGU, 2019. – S. 133–135.
7. Gladkij, I. I. O skorosti skhodimosti i vybore shaga obucheniya iskusstvennyh nejronnyh setej pryamogo rasprostraneniya / I. I. Gladkij, V. V. Budyakov, B. I. CHikalov, T. YU. YUhimuk // Sovremennye problemy matematiki i vychislitel'noj tekhniki : sb. materialov XI Resp. nauch. konf. molodyh uchenyh i studentov, Brest, 21–22 noyab. 2019 g. / Brest. gos. tekhn. un-t ; redkol.: V. A. Golovko (gl. red.) [i dr.]. – Brest : BrGTU, 2019. – S. 11–14.
8. Mahnist, L. P. Ocenki skorosti skhodimosti i vybor shaga obucheniya iskusstvennyh nejronnyh setej pryamogo rasprostraneniya / L. P. Mahnist, V. A. Golovko, I. I. Gladkij, T. I. Karimova // Vestnik Brestskogo gosudarstvennogo tekhnicheskogo universiteta. – 2019. – № 5 : Fizika, matematika, informatika. – S. 27–35.
9. Mahnist, L. P. Ob odnom podhode k obucheniyu iskusstvennyh nejronnyh setej pryamogo rasprostraneniya / L. P. Mahnist, I. I. Gladkij, T. I. Karimova, V. S. Rubanov // Matematicheskoe modelirovanie i novye obrazovatel'nye tekhnologii v matematike : sb. materialov Resp. nauch.-prakt. konf., Brest, 23–24 apr. 2020 g. / Brest. gos. un-t im. A. S. Pushkina ; pod obshch. red. A. I. Basika. – Brest : BrGU, 2020. – S. 41–44.

Материал поступил в редакцию 05.01.2021