АЛГОРИТМ ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ С ИСПОЛЬЗОВАНИЕМ МЕТОДА СОПРЯЖЕННЫХ ГРАДИЕНТОВ

Лахмицкий А.А., БГТУ, Брест

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

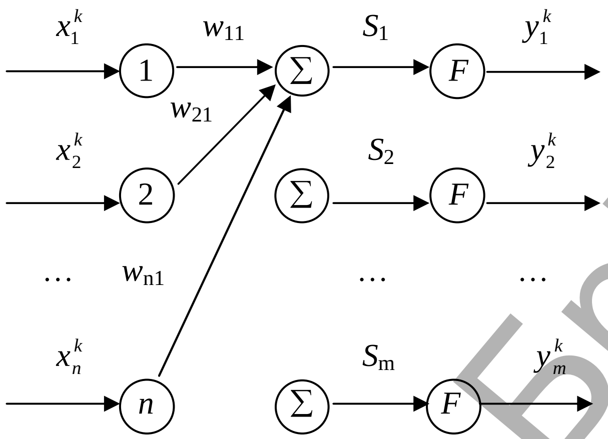

Рис. 1. Схема функционирования нейронной сети

Для данной сети каждый нейрон распределительного слоя имеет синаптические связи w_{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, 2]. На вход сети подаются входные образы – векторы $\overline{x^k}=\left(x_1^k,\dots,x_n^k\right)$ $\left(k=\overline{1,L}\right)$.

Выходное значение *j*-ого нейрона сети для *k*-ого образа определяется выражением:

$$y_j^k=F\left(S_j^k\right),\text{ где } S_j^k=\sum_{i=1}^{n}w_{ij}x_i^k-T_j,\quad j=\overline{1,m},\ \ k=\overline{1,L}. \qquad (1)$$

Задача обучения нейронной сети с фиксированной функцией активации *F* состоит в нахождении весовых коэффициентов w_{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 от эталонных значений t_j^k – *j*-ого нейрона сети для *k*-ого образа. В качестве ошибки сети можно рассмотреть “квадратичное отклонение” $E_S=\dfrac{1}{2}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(y_j^k-t_j^k\right)^2$, которое будем называть квадратичной ошибкой сети.

Столбец $\overline{W}=\left(w_{11},w_{21},\dots,w_{n1},T_1,w_{12},w_{22},\dots,w_{n2},T_2,\dots,w_{1m},w_{2m},\dots,w_{nm},T_m\right)^T$ будем называть приближенным решением или просто решением системы (по методу наименьших квадратов):

$$F\left(\sum_{i=1}^{n}w_{ij}x_i^k-T_j\right)=t_j^k,\quad j=\overline{1,m},\ \ k=\overline{1,L},$$

если “квадратичное отклонение” $E_S=\dfrac{1}{2}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(F\left(\sum_{i=1}^{n}w_{ij}x_i^k-T_j\right)-t_j^k\right)^2$ достигает своего наименьшего значения.

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

$$E_S\left(w_{11},w_{21},\dots,w_{n1},T_1,w_{12},w_{22},\dots,w_{n2},T_2,\dots,w_{1m},w_{2m},\dots,w_{nm},T_m\right).$$

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

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

вектор градиента функции $E_S(t)$, $\nabla^2 E_S(t)$ – матрица Гессе вторых производных функции $E_S(t)$ в момент времени *t*.

Учитывая, что в соответствии с идеей метода сопряженных градиентов

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

и введя обозначение $\overline{\Delta W}(t)=\overline{W}(t)-\overline{W}(t-1)$, получим

$$E_S(t+1)=E_S(t)+\left(\nabla E_S(t),\ -\alpha(t)\cdot\nabla E_S(t)+\beta(t)\cdot\overline{\Delta W}(t)\right)+$$

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

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

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

Найдем критические точки функции $E_S(t+1)=E_S(\alpha;\beta)$, как функции двух переменных. Для этого найдем частные производные функции $E_S(\alpha;\beta)$ и решим систему уравнений:

$$\begin{cases}\dfrac{\partial E_S}{\partial\alpha}=-\left(\nabla E_S,\ \nabla E_S\right)+\alpha\cdot\left(\nabla^2 E_S\cdot\nabla E_S,\ \nabla E_S\right)-\beta\cdot\left(\nabla^2 E_S\cdot\nabla E_S,\ \overline{\Delta W}\right)=0\\[2mm] \dfrac{\partial E_S}{\partial\beta}=\left(\nabla E_S,\ \overline{\Delta W}\right)-\alpha\cdot\left(\nabla^2 E_S\cdot\nabla E_S,\ \overline{\Delta W}\right)+\beta\cdot\left(\nabla^2 E_S\cdot\overline{\Delta W},\ \overline{\Delta W}\right)=0\end{cases}$$

Тогда

$$\begin{cases}\alpha(t)=\dfrac{\left\|\nabla E_S(t)\right\|^2\cdot\left(\nabla^2 E_S(t)\cdot\overline{\Delta W}(t),\ \overline{\Delta W}(t)\right)-\left(\nabla E_S(t),\ \overline{\Delta W}(t)\right)\cdot\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \overline{\Delta W}(t)\right)}{\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \nabla E_S(t)\right)\cdot\left(\nabla^2 E_S(t)\cdot\overline{\Delta W}(t),\ \overline{\Delta W}(t)\right)-\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \overline{\Delta W}(t)\right)^2}\\[3mm] \beta(t)=\dfrac{\left\|\nabla E_S(t)\right\|^2\cdot\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \overline{\Delta W}(t)\right)-\left(\nabla E_S(t),\ \overline{\Delta W}(t)\right)\cdot\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \nabla E_S(t)\right)}{\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \nabla E_S(t)\right)\cdot\left(\nabla^2 E_S(t)\cdot\overline{\Delta W}(t),\ \overline{\Delta W}(t)\right)-\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \overline{\Delta W}(t)\right)^2}\end{cases} \qquad (3)$$

Вычислим частные производные второго порядка функции $E_S(\alpha;\beta)$:

$\dfrac{\partial^2 E_S}{\partial\alpha^2}=\left(\nabla^2 E_S\cdot\nabla E_S,\ \nabla E_S\right)$, $\dfrac{\partial^2 E_S}{\partial\alpha\partial\beta}=\dfrac{\partial^2 E_S}{\partial\beta\partial\alpha}=-\left(\nabla^2 E_S\cdot\nabla E_S,\ \overline{\Delta W}\right)$ и

$\dfrac{\partial^2 E_S}{\partial\beta^2}=\left(\nabla^2 E_S(t)\cdot\overline{\Delta W},\ \overline{\Delta W}\right)$.

Тогда гессиан $\left|\nabla^2 E_S(\alpha;\beta)\right|$ функции $E_S(\alpha;\beta)$ равен

$$\left|\nabla^2 E_S(\alpha;\beta)\right|=\begin{vmatrix}\left(\nabla^2 E_S\cdot\nabla E_S,\ \nabla E_S\right) & -\left(\nabla^2 E_S\cdot\nabla E_S,\ \overline{\Delta W}\right)\\ -\left(\nabla^2 E_S\cdot\nabla E_S,\ \overline{\Delta W}\right) & \left(\nabla^2 E_S\cdot\overline{\Delta W},\ \overline{\Delta W}\right)\end{vmatrix}=$$

$$=\left(\nabla^2 E_S\cdot\nabla E_S,\ \nabla E_S\right)\cdot\left(\nabla^2 E_S\cdot\overline{\Delta W},\ \overline{\Delta W}\right)-\left(\nabla^2 E_S\cdot\nabla E_S,\ \overline{\Delta W}\right)^2.$$

В случае если $\left|\nabla^2 E_S(\alpha;\beta)\right| > 0$ и $\left(\nabla^2 E_S \cdot \nabla E_S,\ \nabla E_S\right) > 0$, то функция $E_S(\alpha;\beta)$ достигает минимального значения при $\alpha(t)$ и $\beta(t)$, определяемыми соотношениями (3).

Таким образом, доказано следующее утверждение.

Теорема. Величины квазиоптимальных параметров $\alpha(t)$ и $\beta(t)$ шага обучения нейронной сети с использованием метода сопряженных градиентов в момент времени *t* определяется соотношениями (3):

Приведем алгоритм обучения нейронной сети с использованием метода сопряженных градиентов, использующий соотношения (3):

1. Задается минимальная квадратичная ошибка сети ε_m, которой необходимо достичь в процессе обучения.

2. Записывается число *t*=0 в счетчик числа итераций алгоритма.

3. Случайным образом инициализируются весовые коэффициенты сети $w_{ij}(t)$ $\left(i=\overline{1,n},\ \ j=\overline{1,m}\right)$, и пороговые значения нейронных элементов $T_j(t)$ $\left(j=\overline{1,m}\right)$.

4. Подаются входные образы $\overline{x^k} = \left(x_1^k,\ldots,x_n^k\right)$ $\left(k=\overline{1,L}\right)$ на нейронную сеть и вычисляются векторы $\overline{y^k(t)} = \left(y_1^k(t),\ldots,y_m^k(t)\right)$ $\left(k=\overline{1,L}\right)$ выходной активности сети, определяемые соотношениями (1).

5. Если *t*≠0, то величины квазиоптимальных параметров $\alpha(t)$, $\beta(t)$ шага обучения с использованием метода сопряженных градиентов вычисляются в соответствии с соотношениями (3), в противном случае параметр $\alpha(t)$ определяется выражением $\alpha(t) = \dfrac{\left(\nabla E_S(t),\ \nabla E_S(t)\right)}{\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \nabla E_S(t)\right)}$ [2], а $\beta(t)$ полагается равным нулю.

6. Производится изменение весовых коэффициентов $w_{ij}(t+1)$ $\left(i=\overline{1,n},\ \ j=\overline{1,m}\right)$ и порогов нейронной сети $T_j(t+1)$ $\left(j=\overline{1,m}\right)$ согласно выражению (2).

7. Полагается *t*=*t*+1.

8. Алгоритм завершает свою работу, если суммарная квадратичная ошибка сети $E_S(t) = \dfrac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{L}\left(y_j^k(t) - t_j^k\right)^2$ или норма вектора $\overline{\Delta W}(t) = \overline{W}(t) - \overline{W}(t-1)$ не превосходят заданной величины ε_m, т. е. $E_S(t) \le \varepsilon_m$ или $\left\|\overline{\Delta W}(t)\right\| < \varepsilon_m$, в противном случае выполняется п. 4.

Литература

1. Головко В.А. Нейронные сети: обучение, организация и применение. Кн. 4: Учебное пособие для вузов / Общая ред. А.И. Галушкина. – М.: ИПРЖР, 2001. – 256 с.: ил. (Нейрокомпьютеры и их применение).

2. Гладкий И.И., Головко В.А., Махнист Л.П. Обучение нейронных сетей с использованием метода наискорейшего спуска // Вестник Брестского государственного технического университета. – Брест: БГТУ, 2001. – № 5: Физика, математика, химия.– С. 47-55.

ПРИМЕНЕНИЕ НЕЙРОПОДОБНЫХ СЕТЕЙ В РЕШЕНИИ ЗАДАЧ РЕКУРСИВНОГО ФОРМООБРАЗОВАНИЯ

Полозков Ю. В., ВГТУ, Витебск

Искусственные нейронные сети позволяют эффективно решать широкий круг аналитических информационных задач. В компьютерной технологии изготовления пространственно сложных (нерегулярных) объектов [1] - рекурсивном формообразовании – к таким задачам относятся обработка изображений; организация баз данных, содержащих