входящих в комплект, с координатами границ зон, если хотя бы какие-то два контейнера текущего комплекта находятся в разных зонах или хотя бы один из контейнеров и назначенный вагон находятся в разных зонах, то комплект определяется как межзонный и исключается из зоны;

2.5) последовательно просматривается каждая зона и определяется количество комплектов принадлежащих текущей зоне (исключая межзонные комплекты), если в зоне комплектов нет, то количество зон уменьшается на 1 и выполняется переход к п.2.1 до тех пор, пока N>0;

3) решение задачи коммивояжера для получения маршрута передвижения кранов;

3.1) для каждой зоны, полученной на этапе 2, формируются данные для задачи коммивояжера;

3.2) решается задача коммивояжера методом включения ближайшего соседа. При этом определяется псевдооптимальный путь движения крана в зоне z для загрузки комплектов на вагоны, которые относятся к текущей зоне z (не считая межзонные комплекты), если комплект загружен на вагон, то помечается вагон как загруженный, а комплект как распределенный;

4) если все вагоны загружены, то задача решена. В противном случае, есть незагруженные межзонные комплекты. Полагается $N=N$-1 и этапы 1-3 повторяются.

5. РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ ПРОГРАММЫ

Для снижения затрат на тестирование был разработан модуль генерации тестов. Тесты генерируются случайным образом. Закон распределения случайной величины – равномерный. В генераторе тестов учитываются такие параметры, как средневзвешенная ширина комплекта, количество комплектов. А также учитываются следующие ограничения: количество комплектов равно количеству вагонов, количество вагонов не должно превышать максимально возможное количество на площадке, количество типов вагонов и типов комплектов должно совпадать.

Основным критерием, по которому оценивалось качество программы, являлось минимальное время счета задачи (не более 10 мин).

В результате генерации тестов и работы программы получены следующие результаты при решении задачи для 20 комплектов и 20 вагонов: среднее время решения равно 1.38 с.

Таким образом, предложенный алгоритм решения задачи показал достаточно высокое быстродействие.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Кузнецов А. В., Сакович В. А., Холод Н. И., Высшая математика. Математическое программирование, Мн.: Выш. шк., 1994.
2. Few L., The Shortest Path and the Shortest Road Through n Points, Mathematica, 2, 1955, 141]
3. Kaluga V. V., Muravjev S. A., Siridonov S. V., Telyatnikov R. V., Application of genetic algorithms for solutions of the task is frequent – territorial plannings group radio electronic equipment, International Conference of Neural Networks and Artificial Intelligence ICNNAI’99|Proceedings. Editted by Vladimir Golovko, - Brest: BPI, 1999, 224p.
4. Гимади Э. Х., Перепелица В. А., Асимптотический подход к решению задачи коммивояжера, Сб. «Управляемые системы», Новосибирск, вып. 12, 1974, 35-45.
5. Гимади Э. Х., Перепелица В. А., Статистически эффективный алгоритм выделения гамильтонова контура (цикла), Сб. «Дискретный анализ», Новосибирск, вып. 22, 1973, 15-28.
6. Меламед И. И., Сергеев С. И., Сигал И. Х., Задача коммивояжера. Вопросы теории // Автоматика и телемеханика. – М.: Наука, 1989. №9. с.3-33.
7. Меламед И. И., Сергеев С. И., Сигал И. Х., Задача коммивояжера. Точные методы // Автоматика и телемеханика. – М.: Наука, 1989. №10. с.3-29.
8. Меламед И. И., Сергеев С. И., Сигал И. Х., Задача коммивояжера. Приближенные алгоритмы // Автоматика и телемеханика. – М.: Наука, 1989. №11. с.3-26.
9. Новиков Ф. А., Дискретная математика для программистов, СПб.: Питер, 2000.–304 с.:ил.
10. Перепелица В. А., Гимади Э. Х., К задаче нахождения минимального гамильтонова контура на графе со взвешенными дугами, Сб. «Дискретный анализ», Новосибирск, вып. 15, 1969, 57-65.

УДК 681.324:519.711.7

Маньяков Н.В., Махнист Л.П.

ОБУЧЕНИЕ НЕЙРОННЫХ СЕТЕЙ С ИСПОЛЬЗОВАНИЕМ КОВАРИАЦИОННОЙ МАТРИЦЫ

1. ВВЕДЕНИЕ

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

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 строго монотонной функцией активации F (строгая монотонность необходима для существования обратной функции). На вход сети подаются входные образы – векторы $\overline{x^k}=\left(x_1^k,...,x_n^k\right)$, $\left(k=\overline{1,L}\right)$ или, что тоже самое, на вход сети подается вектор $\overline{\xi}=\left(\xi_1^T,...,\xi_n^T,\xi_{n+1}^T\right)$, где векторa $\xi_q=\left(x_q^1,...,x_q^L\right)$, $\left(q=\overline{1,n}\right)$ и $\xi_{n+1}=\left(-1,...,-1\right)$.

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

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

где

$$S_j^k=\sum_{i=1}^{n}w_{ij}x_i^k-T_j,\quad j=\overline{1,m},\quad k=\overline{1,L}.$$

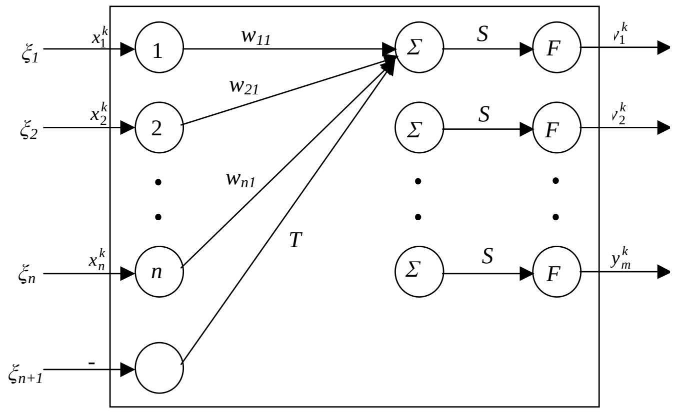

Рис.1. Схема функционирования нейронной сети

Задача обучения нейронной сети с фиксированной функцией активации $\boldsymbol{F}$ [1] состоит в нахождении весовых коэффициентов $\boldsymbol{w_{ij}}$ $\left(\boldsymbol{i=\overline{1,n},\ \ j=\overline{1,m}}\right)$ и порогов нейронных элементов $\boldsymbol{T_j}$ $\left(\boldsymbol{j=\overline{1,m}}\right)$, которые минимизируют некоторую ошибку сети $\boldsymbol{E_S}$, как отклонение выходных значений $\boldsymbol{y_j^k}$ от эталонных значений $\boldsymbol{t_j^k}$ - $\boldsymbol{j}$-ого нейрона сети для $\boldsymbol{k}$-ого образа. В качестве ошибки будем рассматривать усредненное по количеству образов «квадратичное отклонение» $\boldsymbol{E_S=\frac{1}{2L}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(y_j^k-t_j^k\right)^2}$.

Столбец $\boldsymbol{\overline{W}=\left(\overline{W_1}^T,\overline{W_2}^T,...,\overline{W_m}^T\right)^T}$, где $\boldsymbol{\overline{W_j}=\left(w_{1j},w_{2j},...,w_{nj},T_j\right)^T,\quad j=\overline{1,m}}$, будем называть приближенным решением или просто решением системы (по методу наименьших квадратов):

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

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

2. ОПИСАНИЕ МЕТОДА С ИСПОЛЬЗОВАНИЕМ КОВАРИАЦИОННОЙ МАТРИЦЫ

Для решения данной задачи рассмотрим взаимосвязанную с ней задачу с линейной нейронной сетью (рис.2).

Решение задачи с линейной нейронной сетью будет получено из системы:

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

для чего необходимо минимизировать усредненное по количеству образов «квадратичное отклонение» задаваемое выражением:

$$\boldsymbol{E_S=\frac{1}{2L}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(S_j^k-F^{-1}\left(t_j^k\right)\right)^2=}$$

$$\boldsymbol{=\frac{1}{2L}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(\sum_{i=1}^{n}w_{ij}x_i^k-T_j-F^{-1}\left(t_j^k\right)\right)^2}.$$

Данная взаимосвязанность справедлива в силу однозначности и монотонности функции активации $\boldsymbol{F}$.

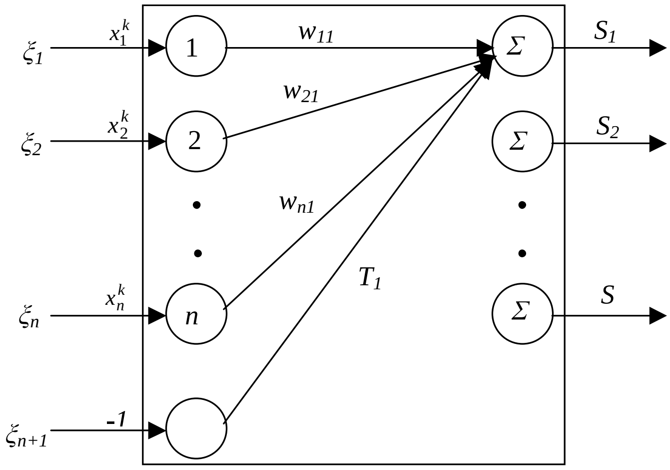

Рис.2. Нейронная сеть взаимосвязанной задачи.

В дальнейшем будем обозначать векторы эталонных значений выхода линейной сети как $\boldsymbol{\eta_j=\left(F^{-1}\left(t_j^1\right),...,F^{-1}\left(t_j^L\right)\right),\quad j=\overline{1,m}}$.

Функция ошибки $\boldsymbol{E_S}$, которую необходимо минимизировать, после преобразования, примет вид:

$$\boldsymbol{E_S=\frac{1}{2L}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(\sum_{i=1}^{n}w_{ij}x_i^k-T_j-F^{-1}\left(t_j^k\right)\right)^2=}$$

$$\boldsymbol{=\frac{1}{2L}\sum_{k=1}^{L}\sum_{j=1}^{m}\Bigg(\left(w_{1j}\ \ w_{2j}\ \ w_{nj}\ \ T\right)\times}$$

$$\boldsymbol{\times\begin{pmatrix}\left(x_1^k\right)^2 & x_1^kx_2^k & ... & x_1^kx_n^k & -x_1^k\\ x_2^kx_1^k & \left(x_2^k\right)^2 & ... & x_2^kx_n^k & -x_2^k\\ ... & ... & ... & ... & ...\\ x_n^kx_1^k & x_n^kx_2^k & ... & \left(x_n^k\right)^2 & -x_n^k\\ -x_1^k & -x_2^k & ... & -x_n^k & 1\end{pmatrix}\cdot\begin{pmatrix}w_{1j}\\ w_{2j}\\ ...\\ w_{nj}\\ T\end{pmatrix}-}$$

$$-2\cdot\begin{pmatrix}w_{1j} & w_{2j} & w_{nj} & T\end{pmatrix}\cdot\begin{pmatrix}x_1^k\\ x_2^k\\ \dots\\ x_n^k\\ -1\end{pmatrix}\cdot F^{-1}\left(t_j^k\right)+\left(F^{-1}\left(t_j^k\right)\right)^2\Bigg)=$$

$$=\frac{1}{2}\sum_{j=1}^{m}\Bigg(W_j^T\times$$

$$\times\begin{pmatrix}
\sum\limits_{k=1}^{L}\frac{\left(x_1^k\right)^2}{L} & \sum\limits_{k=1}^{L}\frac{x_1^k x_2^k}{L} & \dots & \sum\limits_{k=1}^{L}\frac{x_1^k x_n^k}{L} & \sum\limits_{k=1}^{L}\frac{-x_1^k}{L}\\
\sum\limits_{k=1}^{L}\frac{x_2^k x_1^k}{L} & \sum\limits_{k=1}^{L}\frac{\left(x_2^k\right)^2}{L} & \dots & \sum\limits_{k=1}^{L}\frac{x_2^k x_n^k}{L} & \sum\limits_{k=1}^{L}\frac{-x_2^k}{L}\\
\dots & \dots & \dots & \dots & \dots\\
\sum\limits_{k=1}^{L}\frac{x_n^k x_1^k}{L} & \sum\limits_{k=1}^{L}\frac{x_n^k x_2^k}{L} & \dots & \sum\limits_{k=1}^{L}\frac{\left(x_n^k\right)^2}{L} & \sum\limits_{k=1}^{L}\frac{-x_n^k}{L}\\
\sum\limits_{k=1}^{L}\frac{-x_1^k}{L} & \sum\limits_{k=1}^{L}\frac{-x_2^k}{L} & \dots & \sum\limits_{k=1}^{L}\frac{-x_n^k}{L} & 1
\end{pmatrix}\cdot W_j-$$

$$-2\cdot W_j^T\cdot\begin{pmatrix}
\sum\limits_{k=1}^{L}\frac{x_1^k F^{-1}\left(t_j^k\right)}{L}\\
\sum\limits_{k=1}^{L}\frac{x_2^k F^{-1}\left(t_j^k\right)}{L}\\
\dots\\
\sum\limits_{k=1}^{L}\frac{x_n^k F^{-1}\left(t_j^k\right)}{L}\\
\sum\limits_{k=1}^{L}\frac{-F^{-1}\left(t_j^k\right)}{L}
\end{pmatrix}+\sum_{k=1}^{L}\frac{\left(F^{-1}\left(t_j^k\right)\right)^2}{L}\Bigg)=$$

$$=\frac{1}{2}\sum_{j=1}^{m}\Bigg(W_j^T\times$$

$$\times\begin{pmatrix}
K[\xi_1,\xi_1] & K[\xi_1,\xi_2] & \dots & K[\xi_1,\xi_n] & K[\xi_1,\xi_{n+1}]\\
K[\xi_2,\xi_1] & K[\xi_2,\xi_2] & \dots & K[\xi_2,\xi_n] & K[\xi_2,\xi_{n+1}]\\
\dots & \dots & \dots & \dots & \dots\\
K[\xi_n,\xi_1] & K[\xi_n,\xi_2] & \dots & K[\xi_n,\xi_n] & K[\xi_n,\xi_{n+1}]\\
K[\xi_{n+1},\xi_1] & K[\xi_{n+1},\xi_2] & \dots & K[\xi_{n+1},\xi_n] & K[\xi_{n+1},\xi_{n+1}]
\end{pmatrix}\cdot W_j-$$

$$-2\cdot W_j^T\cdot\begin{pmatrix}
K\left[\xi_1,\eta_j\right]\\
K\left[\xi_2,\eta_j\right]\\
\dots\\
K\left[\xi_n,\eta_j\right]\\
K\left[\xi_{n+1},\eta_j\right]
\end{pmatrix}+D\left[\eta_j\right]\Bigg)=$$

$$=\frac{1}{2}\sum_{j=1}^{m}\left(W_j^T\cdot K\left[\bar{\xi},\bar{\xi}\right]\cdot W_j-2\cdot W_j^T\cdot K\left[\bar{\xi},\eta_j\right]+D\left[\eta_j\right]\right) \tag{1}$$

где $K[\xi_i,\xi_j]$ - ковариация случайных векторов ξ_i,ξ_j, являющихся последовательностями входных образов i-ого и j-ого нейронов распределительного слоя, $K\left[\bar{\xi},\bar{\xi}\right]$ - ковариационная матрица случайного вектора $\bar{\xi}$, $D\left[\eta_j\right]$ – дисперсия случайного вектора η_j, являющегося выходной последовательностью j-ого нейрона взаимосвязанной задачи [2].

Теорема 1. Если входные образы $\xi_1,\xi_2,\dots,\xi_n,\xi_{n+1}$ линейно независимы, то весовые коэффициенты однослойной нейронной сети (рис.2) с монотонной функцией активации определяются однозначно из соотношения $W_j=K^{-1}\left[\bar{\xi},\bar{\xi}\right]\cdot K\left[\bar{\xi},\eta_j\right]$.

Доказательство. Будем рассматривать эквивалентную нейронную сеть (рис.2). Необходимо найти такой вектор весовых коэффициентов $\overline{W}$, который минимизирует функцию ошибки E_S. Так как она строилась в соответствии с методом наименьших квадратов, то такой вектор существует. Докажем его единственность.

Данный вектор будет являться стационарной точкой функции E_S. Найдем его из соотношения (1) взяв частные производные по элементам вектора $\overline{W}$ и приравняв их к нулю. Имеем:

$$\sum_{j=1}^{m}\left(K\left[\bar{\xi},\bar{\xi}\right]\cdot W_j-K\left[\bar{\xi},\eta_j\right]\right)=0, \tag{2}$$

что эквивалентно соотношению:

$$\begin{pmatrix}
K\left[\bar{\xi},\bar{\xi}\right] & 0 & \dots & 0\\
0 & K\left[\bar{\xi},\bar{\xi}\right] & \dots & 0\\
\dots & \dots & \dots & \dots\\
0 & 0 & \dots & K\left[\bar{\xi},\bar{\xi}\right]
\end{pmatrix}\cdot\begin{pmatrix}W_1\\ W_2\\ \dots\\ W_m\end{pmatrix}=\begin{pmatrix}K\left[\bar{\xi},\eta_1\right]\\ K\left[\bar{\xi},\eta_2\right]\\ \dots\\ K\left[\bar{\xi},\eta_m\right]\end{pmatrix}$$

Если ковариационная матрица $K\left[\bar{\xi},\bar{\xi}\right]$ невырождена, то

$$\overline{W}=\begin{pmatrix}W_1\\ W_2\\ \dots\\ W_m\end{pmatrix}=\begin{pmatrix}
K^{-1}\left[\bar{\xi},\bar{\xi}\right] & 0 & \dots & 0\\
0 & K^{-1}\left[\bar{\xi},\bar{\xi}\right] & \dots & 0\\
\dots & \dots & \dots & \dots\\
0 & 0 & \dots & K^{-1}\left[\bar{\xi},\bar{\xi}\right]
\end{pmatrix}\times$$

$$\times\begin{pmatrix}K\left[\bar{\xi},\eta_1\right]\\ K\left[\bar{\xi},\eta_2\right]\\ \dots\\ K\left[\bar{\xi},\eta_m\right]\end{pmatrix}$$

что совпадает с соотношением $W_j = K^{-1}\left[\bar{\xi},\bar{\xi}\right]\cdot K\left[\bar{\xi},\eta_j\right]$.

Покажем, что определитель ковариационной матрицы $K\left[\bar{\xi},\bar{\xi}\right]$ отличен от нуля.

Предположим, что компоненты вектора $\bar{\xi}$ линейно зависимы, т.е. существуют такие числа $\lambda_1,\lambda_2,...,\lambda_{n+1}$ не все равные нулю, что выполняется соотношение

$$\lambda_1\xi_1+\lambda_2\xi_2+...+\lambda_{n+1}\xi_{n+1}=\bar{0},$$

где $\bar{0}$ - нуль-вектор.

Тогда

$$D\left[\sum_{q=1}^{n+1}\lambda_q\xi_q\right]=\sum_{i=1}^{n+1}\sum_{j=1}^{n+1}\lambda_i\lambda_j K\left[\xi_i,\xi_j\right]=\bar{\lambda}^T K\left[\bar{\xi},\bar{\xi}\right]\bar{\lambda}=0$$

где $\bar{\lambda}=\left(\lambda_1 \quad \lambda_2 \quad ... \quad \lambda_{n+1}\right)^T$.

Т.е. неотрицательная в силу определения функция $D\left[\sum_{q=1}^{n+1}\lambda_q\xi_q\right]$ достигает в ненулевой точке своего наименьшего значения. Поэтому однородная система (необходимое условие экстремума)

$$\frac{\partial D}{\partial\bar{\lambda}}=K\left[\bar{\xi},\bar{\xi}\right]\bar{\lambda}=0 \quad (*)$$

имеет ненулевое решение, откуда следует вырожденность ковариационной матрицы $K\left[\bar{\xi},\bar{\xi}\right]$.

Пусть теперь матрица $K\left[\bar{\xi},\bar{\xi}\right]$ вырождена. Тогда система (*) имеет ненулевое решение. Умножая эту систему справа на $\bar{\lambda}^T$ получаем

$$\bar{\lambda}^T K\left[\bar{\xi},\bar{\xi}\right]\bar{\lambda}=0,$$

откуда следует $D\left[\sum_{q=1}^{n+1}\lambda_q\xi_q\right]=0$ и, значит, компоненты вектора $\bar{\xi}$ линейно зависимы.

Но так как по условию теоремы входные образы $\xi_1,\xi_2,...,\xi_n,\xi_{n+1}$ линейно независимы, то матрица ковариации невырождена и, значит, справедлива формула $W_j = K^{-1}\left[\bar{\xi},\bar{\xi}\right]\cdot K\left[\bar{\xi},\eta_j\right]$.

Теорема доказана.

Сформулируем в виде следствий утверждения, которые могут быть использованы при выборе архитектуры нейронной сети, применяемой, например, для прогнозирования временных рядов.

Следствие 1. Ковариационная матрица $K\left[\overline{\vartheta_k},\overline{\vartheta_k}\right]$, где $\overline{\vartheta_k}=\left(\xi_1 \quad \xi_2 \quad ... \quad \xi_k\right)$ – случайный вектор с элементами являющимися последовательностями входных образов k нейронов распределительного слоя, невырождена.

Следствие 2. Для оптимального прогнозирования функции методом скользящего окна необходимо выбирать количество входных нейронов таким образом, чтобы ковариационная матрица входных образов была невырождена.

Полученные в теореме 1 результаты, дают возможность получения приближенного решения задачи обучения нейронной сети с нелинейной функцией активации, используя следующее утверждение.

Предложение 1. Решение исходной задачи обучения нейронной сети (рис.1) существует с ошибкой не превосходящей величину $\varepsilon_m=\left(F'(0)\right)^2\cdot\varepsilon^*$.

Доказательство.

Пусть вектор

$$\overline{W}^*=\left(w_{11}^*,w_{21}^*,...,w_{n1}^*,T_1^*,w_{12}^*,w_{22}^*,...,w_{n2}^*,T_2^*,...,\right.$$
$$\left.,w_{1m}^*,w_{2m}^*,...,w_{nm}^*,T_m^*\right)^T$$

получен с использованием теоремы 1 и «квадратичное отклонение»

$$E_S^*=\frac{1}{2L}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(\sum_{i=1}^{n}w_{ij}^*x_i^k-T_j^*-F^{-1}\left(t_j^k\right)\right)^2=\varepsilon^*.$$

Предположим, что существует нехудшее решение

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

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

по методу наименьших квадратов.

Тогда

$$E_S=\frac{1}{2L}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(F\left(\sum_{i=1}^{n}w_{ij}x_i^k-T_j\right)-F\left(F^{-1}\left(t_j^k\right)\right)\right)^2\le$$
$$\le\frac{1}{2L}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(F\left(\sum_{i=1}^{n}w_{ij}^*x_i^k-T_j^*\right)-F\left(F^{-1}\left(t_j^k\right)\right)\right)^2$$

С другой стороны, используя формулу конечных приращений Лагранжа, имеем, что

$$\frac{1}{2L}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(F\left(\sum_{i=1}^{n}w_{ij}^*x_i^k-T_j\right)-F\left(F^{-1}\left(t_j^k\right)\right)\right)^2\le$$
$$\le\left(\max_S F'(S)\right)^2\cdot\frac{1}{2L}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(\sum_{i=1}^{n}w_{ij}^*x_i^k-T_j^*-F^{-1}\left(t_j^k\right)\right)^2$$

Следовательно,

$$E_S=\frac{1}{2L}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(F\left(\sum_{i=1}^{n}w_{ij}x_i^k-T_j\right)-t_j^k\right)^2\le$$
$$\le\left(\max_S F'(S)\right)^2\cdot\varepsilon^*=\left(F'(0)\right)^2\cdot\varepsilon^*$$

т. к. $\max_S F'(S)=F'(0)$, для функций активации F [1].

Таким образом, полученное неравенство может служить верхней оценкой для минимальной квадратичной ошибки сети ε_m, которой необходимо достичь в процессе обучения, так как при $\varepsilon_m=\left(F'(0)\right)^2\cdot\varepsilon^*$ решение исходной системы существует.

Предложение доказано.

Используя теорему 1 можно получить необходимое условие единственности решения задачи обучения нейронной сети (1).

Предложение 2. Если задача обучения линейной нейронной сети имеет единственное решение, то выполняется неравенство: $L \geq n+1$, где L – количество образов, n – количество нейронов в распределительном слое.

Доказательство. Покажем что в этом случае вектора $\xi_q = \left(x_q^1, ..., x_q^L\right)$, $\left(q = \overline{1,n}\right)$ и $\xi_{n+1} = (-1, ..., -1)$, подаваемые на нейроны распределительного слоя, где количество образов $L \leq n$, являются линейно зависимыми.

Составим их линейную комбинацию и приравняем её к нулю:

$$\lambda_1 \xi_1 + \lambda_2 \xi_2 + ... + \lambda_{n+1} \xi_{n+1} = \overline{0},$$

что равносильно системе:

$$\begin{cases} x_1^1 \lambda_1 + x_2^1 \lambda_2 + ... + x_n^1 \lambda_n - \lambda_{n+1} = 0 \\ x_1^2 \lambda_1 + x_2^2 \lambda_2 + ... + x_n^2 \lambda_n - \lambda_{n+1} = 0 \\ ... \\ x_1^L \lambda_1 + x_2^L \lambda_2 + ... + x_n^L \lambda_n - \lambda_{n+1} = 0 \end{cases},$$

где $L \leq n$, с матрицей системы

$$A = \begin{pmatrix} x_1^1 & x_2^1 & ... & x_n^1 & -1 \\ x_1^2 & x_2^2 & ... & x_n^2 & -1 \\ ... & ... & ... & ... & ... \\ x_1^L & x_2^L & ... & x_n^L & -1 \end{pmatrix},$$

размерности $L \times (n+1)$.

Т.к. $rang(A) \leq \min(L, n+1) = L < n+1$, то рассматриваемая система имеет ненулевое решение $\overline{\lambda} = \begin{pmatrix} \lambda_1 & \lambda_2 & ... & \lambda_{n+1} \end{pmatrix}$, а значит $\xi_q = \left(x_q^1, ..., x_q^L\right)$, $\left(q = \overline{1,n}\right)$ и $\xi_{n+1} = (-1, ..., -1)$ линейно зависимы, что в соответствии с теоремой 1 означает вырожденность ковариационной матрицы $K\left[\overline{\xi}, \overline{\xi}\right]$. Поэтому вектор $\overline{W}$ определяется не единственным образом.

Предложение доказано.

Таким образом, задачи нахождения минимума функции (1) и решения системы (2) эквивалентны. Для численного нахождения матрицы, обратной к ковариационной, можно использовать любой из известных методов [3].

3. К ВОПРОСУ О СХОДИМОСТИ МЕТОДА ОБУЧЕНИЯ

Полученные в данном разделе результаты дают возможность оценить скорость сходимости алгоритмов различных численных методов обучения нейронных сетей как, например, в [4].

Теорема 2. Функция E_S эквивалентной задачи, определяемая соотношением (1) является сильно выпуклой функцией.

Доказательство. Как было показано ранее в условии теорем 1 при линейной независимости входных образов $\xi_1, \xi_2, ..., \xi_n, \xi_{n+1}$ у функции (1) существует единственная точка минимума. А это значит, что матрица Гессе этой функции в точке минимума является неотрицательно определенной.

Функцию (1) можно записать как следующую квадратичную функцию:

$$E_S = \frac{1}{2}\begin{pmatrix} W_1^T & W_2^T & ... & W_m^T \end{pmatrix} \times$$

$$\times \begin{pmatrix} K\left[\overline{\xi}, \overline{\xi}\right] & 0 & ... & 0 \\ 0 & K\left[\overline{\xi}, \overline{\xi}\right] & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & K\left[\overline{\xi}, \overline{\xi}\right] \end{pmatrix} \cdot \begin{pmatrix} W_1 \\ W_2 \\ ... \\ W_m \end{pmatrix} +$$

$$+ \begin{pmatrix} W_1^T & W_2^T & ... & W_m^T \end{pmatrix} \cdot \begin{pmatrix} -K\left[\overline{\xi}, \eta_1\right] \\ -K\left[\overline{\xi}, \eta_2\right] \\ ... \\ -K\left[\overline{\xi}, \eta_m\right] \end{pmatrix} + \frac{1}{2}\sum_{j=1}^{m} D\left[\eta_j\right] =$$

$$= \frac{1}{2}\left(Q\overline{W}, \overline{W}\right) + \left(c, \overline{W}\right) + b, \text{ где } b \geq 0 \qquad (3)$$

Но матрица Гессе этой квадратичной функции имеет вид

$$Q = \begin{pmatrix} K\left[\overline{\xi}, \overline{\xi}\right] & 0 & ... & 0 \\ 0 & K\left[\overline{\xi}, \overline{\xi}\right] & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & K\left[\overline{\xi}, \overline{\xi}\right] \end{pmatrix},$$

который не зависит от точки $\overline{W}$. Таким образом, в любой точке симметричная матрица Гессе неотрицательная. А это значит, что все её диагональные (главные) миноры не отрицательны, в частности, и угловые миноры [4,5].

Рассмотрим сначала угловые миноры порядка k не большего, чем $n+1$. Эти миноры являются минорами ковариационной матрицы $K\left[\overline{\vartheta_k}, \overline{\vartheta_k}\right]$, где $\overline{\vartheta_k} = \begin{pmatrix} \xi_1^T & \xi_2^T & ... & \xi_k^T \end{pmatrix}$. А как показано в следствии 1 теоремы 1, они отличны от нуля, а значит, строго положительны.

Рассмотрим миноры порядка k большего $n+1$. Они имеют вид:

$$\begin{vmatrix} K\left[\overline{\xi}, \overline{\xi}\right] & 0 & ... & 0 \\ 0 & K\left[\overline{\xi}, \overline{\xi}\right] & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & K\left[\overline{\vartheta_r}, \overline{\vartheta_r}\right] \end{vmatrix},$$

где $r = k \mod (n+1)$, и, соответственно, равны как определители блочной матрицы произведению определителей:

$$\left|K\left[\overline{\xi}, \overline{\xi}\right]\right| \cdot \left|K\left[\overline{\xi}, \overline{\xi}\right]\right| \cdot ... \cdot \left|K\left[\overline{\vartheta_r}, \overline{\vartheta_r}\right]\right|,$$

каждый из которых, в соответствии с выше доказанным, отличен от нуля. Т. е. является положительным числом.

Таким образом, мы показали, что все угловые миноры матрицы Q, строго положительны. В соответствии с критерием Сильвестра это означает положительную определенность матрицы $\boldsymbol{Q}$ квадратичной функции (3).

Учитывая тождества $\alpha^2 = \alpha - \alpha(1-\alpha)$ и $(1-\alpha)^2 = (1-\alpha) - \alpha(1-\alpha)$, получим:

$$\left(Q\left(\alpha\overline{W^1} + (1-\alpha)\overline{W^2}\right), \alpha\overline{W^1} + (1-\alpha)\overline{W^2}\right) =$$

$$= \alpha^2\left(Q\overline{W^1}, \overline{W^1}\right) + \alpha(1-\alpha)\left(\left(Q\overline{W^1}, \overline{W^2}\right) + \left(Q\overline{W^2}, \overline{W^1}\right)\right) +$$

$$+ (1-\alpha)^2\left(Q\overline{W^2}, \overline{W^2}\right) =$$

$$= \alpha\left(Q\overline{W^1}, \overline{W^1}\right) + (1-\alpha)\left(Q\overline{W^2}, \overline{W^2}\right) -$$

$$-\alpha(1-\alpha)\left(Q\left(\overline{W^1} - \overline{W^2}\right), \overline{W^1} - \overline{W^2}\right)$$

Поэтому

$$E_S\left(\alpha\overline{W^1} + (1-\alpha)\overline{W^2}\right) =$$

$$= \frac{1}{2}\left(Q\left(\alpha\overline{W^1} + (1-\alpha)\overline{W^2}\right), \alpha\overline{W^1} + (1-\alpha)\overline{W^2}\right) +$$

$$+ \left(c, \alpha\overline{W^1} + (1-\alpha)\overline{W^2}\right) + b =$$

$$= \alpha\frac{1}{2}\left(Q\overline{W^1}, \overline{W^1}\right) + (1-\alpha)\frac{1}{2}\left(Q\overline{W^2}, \overline{W^2}\right) +$$

$$+ \alpha\left(c, \overline{W^1}\right) + (1-\alpha)\left(c, \overline{W^2}\right) + b + b -$$

$$-b - \frac{1}{2}\alpha(1-\alpha)\left(Q\left(\overline{W^1} - \overline{W^2}\right), \overline{W^1} - \overline{W^2}\right) =$$

$$= \alpha E_S\left(\overline{W^1}\right) + (1-\alpha)E_S\left(\overline{W^2}\right) - b -$$

$$-\frac{1}{2}\alpha(1-\alpha)\left(Q\left(\overline{W^1} - \overline{W^2}\right), \overline{W^1} - \overline{W^2}\right) \le$$

$$\le \alpha E_S\left(\overline{W^1}\right) + (1-\alpha)E_S\left(\overline{W^2}\right) -$$

$$-\frac{1}{2}\alpha(1-\alpha)\left(Q\left(\overline{W^1} - \overline{W^2}\right), \overline{W^1} - \overline{W^2}\right) \le$$

$$\le \alpha E_S\left(\overline{W^1}\right) + (1-\alpha)E_S\left(\overline{W^2}\right) -$$

$$-\frac{\lambda}{2}\alpha(1-\alpha)\left\|\overline{W^1} - \overline{W^2}\right\|^2$$

т.к., для положительно определенной матрицы $\boldsymbol{Q}$, справедливо неравенство

$\left(Q\left(\overline{W^1} - \overline{W^2}\right), \overline{W^1} - \overline{W^2}\right) \ge \lambda\left\|\overline{W^1} - \overline{W^2}\right\|^2$, где λ - её наименьшее собственное значение. Таким образом, мы пришли к выводу о сильной выпуклости функции ошибки E_S эквивалентной задачи.

Теорема доказана.

Эта теорема ещё раз подтверждает факт, что функция $\boldsymbol{E_S}$ имеет единственную точку минимума, т. к. она сильно выпукла.

4. АЛГОРИТМ МЕТОДА ИСПОЛЬЗУЮЩЕГО МАТРИЦУ КОВАРИАЦИИ

На основании вышеизложенного приведем алгоритм решения вспомогательной задачи на основании метода с использованием ковариационной матрица.

1. Вычисляется ковариационная матрица $K\left[\bar{\xi}, \bar{\xi}\right]$ последовательностей входных образов на нейроны распределительного слоя.
2. Вычисляются матрицы ковариации $K\left[\bar{\xi}, \eta_j\right]$, $j = \overline{1, m}$ вектора входных образов и последовательностей $\left\{\eta_j\right\}_k$, $k = \overline{1, L}$ выходных данных вспомогательной сети.
3. Обращается полученная ковариационная матрица $K\left[\bar{\xi}, \bar{\xi}\right]$.
4. Вычисляются весовые коэффициенты сети с использованием формул $W_j = K^{-1}\left[\bar{\xi}, \bar{\xi}\right] \cdot K\left[\bar{\xi}, \eta_j\right]$, где $j = \overline{1, m}$.

Данный алгоритм может быть использован для приближенного решения задачи обучения нейронной сети с нелинейной функцией активации. Кроме того, полученное с его помощью решение может быть использовано как начальное для решения задачи с использованием различных численных методов обучения таких как, например, градиентные методы [4].

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. 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. 56-61.
2. Боровков А.А., Теория вероятностей. – М.: Наука, 1976.
3. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. – М.: Наука. Главная редакция физ.-мат. литературы, 1984.
4. Аттетков А.В., Галкин С.В., Зарубин В.С. Методы оптимизации. – М.: Издательство МГТУ им. Н.Э. Баумана, 2001.
5. Милованов М.В., Толкачев М.М., Тышкевич Р.И., Феденко А.С. Алгебра и аналитическая геометрия. Часть 2. – Мн.: Амалфея, 2001.