УДК 519.67

С. И. Хашин

СРАВНЕНИЕ ЭФФЕКТИВНОСТИ ПЕРЕДАТОЧНЫХ ФУНКЦИЙ НЕЙРОСЕТИ

Рассматриваются несколько простых вариантов архитектуры нейросети с двумя и тремя входными параметрами и с различными активаторными функциями. Для каждого случая производится 40 000 попыток обучения нейросети с различными случайными начальными значениями. Сравнивается эффективность различных выборов активаторных функций и приводится оценка общего количества экстремумов целевой функции.

Ключевые слова: нейросеть, передаточная (активаторная) функция.

A few simple variants of the neural network architecture with two and three input parameters and with various activation functions are considered. For each case, 40 000 attempts are made to train a neural network with various random initial values. The effectiveness of different choices of activation functions is compared and an estimate of the total number of local extremums of the objective function is given.

Key words: neural network, activation function.

1. Введение

Использование нейронных сетей за последнее время привело к значительным успехам во многих разделах компьютерной графики [1, 2, 3]. К сожалению, большинство результатов в этой области носит экспериментальный характер, выбор той или иной архитектуры нейросети, выбор активаторных функций основывается на интуиции разбработчика. Обучение нейронной сети требует больших вычислительных ресурсов, и у разработчиков обычно не доходят руки до проверки и сравнения эффективности различных вариантов.

В настоящей работе мы пытаемся на простейших примерах из компьютерной графики сравнить эффективность нескольких различных архитектур и активаторных функций. Все рассмотренные алгоритмы были реализованы на C++ (Visual Studio, [4]).

2. Архитектура нейросети

В работе мы рассматриваем простейшую задачу NN-регрессии (Neural Network Regression). Под нейросетью будем понимать действительную функцию от двух аргументов $F(X,W)$, где $X \in \mathbb{R}^d$ — вектор входных значений и $W \in \mathbb{R}^k$ — вектор внутренних параметров нейросети.

В работе мы будем рассматривать нейросети с двумя и тремя входными параметрами, содержащие два или три нейрона.

Определение. а) Нейросетью с двумя входными параметрами (x_1, x_2) и с двумя нейронами будем называть функцию $y = F(x_1, x_2)$ от двух переменных, вычисляемую по формулам:

$$
\begin{aligned}
x_3 &= f(w_0 + w_1x_1 + w_2x_2),\\
y &= w_3 + w_4x_1 + w_5x_2 + w_6x_3,
\end{aligned}
$$

где $f\colon \mathbb{R} \to \mathbb{R}$ — передаточная функция и $(w_0, \ldots, w_6)$ — внутренние параметры нейрона.

б) Нейросетью с двумя входными параметрами (x_1, x_2) и с тремя нейронами будем называть функцию $y = F(x_1, x_2)$ от двух переменных, вычисляемую по формулам:

$$\begin{aligned} x_3 &= f(w_0 + w_1x_1 + w_2x_2),\\ x_4 &= f(w_0 + w_1x_1 + w_2x_2 + w_3x_3),\\ y\ &= w_4 + w_5x_1 + w_6x_2 + w_7x_3 + w_8x_4, \end{aligned}$$

где $(w_0, \ldots, w_8)$ — внутренние параметры нейрона.

в) Нейросетью с тремя входными параметрами (x_1, x_2, x_3) и с двумя нейронами будем называть функцию $y = F(x_1, x_2, x_3)$ от трёх переменных, вычисляемую по формулам:

$$\begin{aligned} x_4 &= f(w_0 + w_1x_1 + w_2x_2 + w_3x_3),\\ y\ &= w_4 + w_5x_1 + w_6x_2 + w_7x_3 + w_8x_4, \end{aligned}$$

где $(w_0, \ldots, w_8)$ — внутренние параметры нейрона.

г) Нейросетью с тремя входными параметрами (x_1, x_2, x_3) и с тремя нейронами будем называть функцию $y = F(x_1, x_2, x_3)$ от трёх переменных, вычисляемую по формулам:

$$\begin{aligned} x_4 &= f(w_0 + w_1x_1 + w_2x_2 + w_3x_3),\\ x_5 &= f(w_4 + w_5x_1 + w_6x_2 + w_7x_3 + w_8x_4),\\ y\ &= w_9 + w_{10}5x_1 + w_{11}x_2 + w_{12}x_3 + w_{13}x_4 + w_{14}x_4, \end{aligned}$$

где $(w_0, \ldots, w_{14})$ — внутренние параметры нейрона.

В качестве передаточной (активаторной) функции $f(x)$ мы рассматриваем одну из следующих:

(1) $f(x) = \mathrm{atg}(x)/\pi + 1/2$,
(2) $f(x) = 1/(1 + \exp(-x))$,
(3) $f(x) = (x < 0\ ?\ 0 : x)$ (ReLU),
(4) $f(x) = (x < 0\ ?\ 0 : x < 1\ ?\ x : 1)$,
(5) $f(x) = (x < 0\ ?\ 0 : 1)$ (sign),
(6) $f(x) = (1 + \mathrm{th}(x))/2$.

Задача обучения нейросети сводится к подбору внутренних параметров $W = W_0$ так, чтобы функция $X \to F(X, W_0)$ максимально точно удовлетворяла заданным условиям. Точнее, дано некоторое обучающее множество точек $(X_1, \ldots, X_n)$ и ожидаемые значения функции в них $(y_1, \ldots, y_n)$. Рассмотрим целевую функцию

$$S(W) = \sum_{i=1}^{n} (F(X_i, W) - y_i)^2/n$$

и будем искать вектор W_0, её минимизирующий.

3. Обучающее множество

Для примера рассмотрим задачу из компьютерной графики: аппроксимацию значения яркости в точке через её значения в соседних точках. Для удобства в качестве значения функции возьмем отклонения яркости в точке от её билинейной аппроксимации. Поделив яркости на 255, будем считать, что входные параметры x_i лежат в интервале от 0 до 1.

В случае двух входных параметров было взято следующее обучающее множество из 18 точек:

y	x_1	x_2
-0.0375000	-0.0039216	-0.1450980
-0.0375000	0.0117647	-0.0039216
-0.1004902	-0.0039216	-0.0196078
-0.0004902	-0.0039216	-0.0431373
0.0004902	0.0039216	-0.0352941
-0.0129902	0.0392157	-0.0039216
-0.0129902	0.0862745	-0.0039216
0.0134804	-0.0039216	-0.0352941
-0.0134804	0.0039216	0.0039216
-0.0014706	0.0509804	0.0039216
0.0144608	-0.0039216	-0.0235294
-0.0269608	0.0039216	-0.0509804
-0.0144608	0.0078431	0.0039216
-0.0024510	0.0431373	-0.0039216
-0.0154412	0.0274510	0.0039216
-0.0279412	0.0627451	0.0039216
0.0159314	0.0039216	-0.0196078
0.0034314	0.0313725	0.0039216

В случае трёх входных параметров было взято следующее обучающее множество опять из 18 точек:

y	x_1	x_2	x_3
0.0125000	-0.0039216	-0.0156863	-0.0156863
0.0250000	-0.0039216	-0.0352941	-0.0313725
-0.1004902	-0.0039216	-0.0196078	-0.0156863
-0.0269608	0.0274510	0.0039216	0.0235294
0.0024510	-0.0039216	-0.0078431	-0.0039216
-0.0149510	-0.0039216	-0.0156863	-0.0156863
0.0024510	0.0196078	0.0039216	0.0156863
0.0024510	0.0235294	-0.0039216	0.0274510
-0.0024510	0.0431373	-0.0039216	0.0470588
0.0154412	-0.0039216	-0.0117647	-0.0117647
0.0154412	-0.0039216	-0.0117647	-0.0117647
0.0174020	0.0117647	-0.0039216	0.0117647
-0.0424020	0.0156863	-0.0039216	0.0156863
-0.0174020	0.0196078	0.0039216	0.0196078
0.0299020	0.0235294	-0.0039216	0.0274510
0.0174020	0.0274510	-0.0039216	0.0274510
-0.0299020	0.0274510	-0.0039216	0.0313725
-0.0053922	-0.0039216	-0.0078431	-0.0039216

4. Полученные результаты

Алгоритмы поиска минимума целевой функции, применяемые в нейросетях, в основном являются вариантами градиентных методов и начинаются со случайного выбора начального вектора W.

Для каждой архитектуры нейросети:

- два входных параметра, два нейрона;
- два входных параметра, три нейрона;
- три входных параметра, два нейрона;
- три входных параметра, три нейрона;

и для каждой активаторной функции (1) — (6) было проведено по 40 000 циклов обучения с различными случайными векторами начальных значений W. В следующих таблицах достигнутые значения целевой функции умножаются на 10^5 и округляются до целых. В строках таблиц указано, какое наименьшее значение целевой функции было достигнуто за $K = 10\,000$, 1000, 100 и 10 случайных попыток и какое наименьшее значение было получено за все 40 000 испытаний (строка «min»).

Два входных аргумента

В случае двух входных аргументов наименьшее найденное значение целевой функции равно $1,8 \cdot 10^{-4}$. Так как полученные значения целевой функции умножаются на 10^5 и округляются до целых, то наименьшее значение в таблице равно 18.

K	21	22	23	24	25	26	31	32	33	34	35	36
10000	62	62	23	24	34	61	59	60	20	21	39	58
1000	62	62	53	48	46	62	62	62	25	30	64	60
100	63	63	65	65	67	62	62	62	64	62	64	62
10	65	65	68	68	67	64	65	63	67	67	67	63
min	61	62	21	23	25	26	58	59	18	19	34	57

Оценим также, на основе полученных данных, количество локальных экстремумов целевой функции:

K	21	22	26	31	32	36
N_{extr}	$120K$	$105K$	$31K$	$> 50M$	$> 50M$	$10M$

Замечание. Если активаторная функция — кусочно-линейная (номера 3, 4, 5), то и целевая функция $S(W)$ будет кусочно-линейной. На каждой области, где она линейна, минимум будет достигаться в бесконечном множестве точек, поэтому в данном случае вопрос о количестве локальных минимумов нельзя ставить.

Три входных аргумента

В этом случае наименьшее найденное значение целевой функции равно $2,7 \cdot 10^{-4}$. В дальнейшем полученные значения целевой функции умножаются на 10^5 и округляются до целых, то есть наименьшее значение в таблице равно 27.

K	21	22	23	24	25	26	31	32	33	34	35	36
10000	83	84	45	47	52	79	67	72	33	31	52	48
1000	83	87	69	68	78	79	74	77	50	47	76	59
100	85	88	81	78	90	81	81	82	75	75	90	72
10	89	91	92	90	90	88	87	86	92	90	90	81
min	82	84	41	42	52	78	62	70	27	28	51	45

Как и выше, оценим на основе полученных данных количество локальных экстремумов целевой функции:

K	21	22	26	31	32	36
N_{extr}	$640K$	$200K$	$61K$	$\approx 50M$	$> 50M$	$\approx 40M$

5. Заключение

На основе проведенных экспериментов можно сделать следующие выводы.

1. Достаточно очевидный и ожидаемый результат: с помощью трёх нейронов можно достичь лучшей аппроксимации, чем с помощью двух.

2. Почти во всех случаях наилучших результатов можно добиться с помощью активаторной функции ReLU: $f(x) = max(0, x)$, с ней может соперничать функция RuLU1: $f(x) = min(max(0, x), 1)$.

Библиографический список

1. *Галушкин А. И.* Нейрокомпьютеры : учеб. пособие // М. : Альянс, 2014. 528 с.
2. *Николенко С., Кадурин А., Архангельская Е.* Глубокое обучение // СПб. : Питер, 2018. 480 с.
3. Основы нейрокибернетики. М. : Горячая линия – Телеком, 2015. 372 с.
4. Центр загрузки Microsoft. URL: https://www.microsoft.com/ru-ru/download (дата обращения: 20.01.2019).

УДК 512.714

Ю. А. Хашина

ПРЕДСТАВЛЕНИЕ N-КВАДРАТИЧНОЙ ФУНКЦИИ В ВИДЕ СУММЫ КВАДРАТОВ

n-квадратичная функция, представимая в виде суммы квадратов n-линейных функций, может быть представлена в виде такой суммы, состоящей не более чем из $2^n - 1$ слагаемых и неотрицательной константы.

Ключевые слова: n-квадратичная функция, экстремум.

An n-quadratic function represented as a sum of squares of n-linear functions can be represented as such a sum consisting of no more than $2^n - 1$ terms and a non-negative constants.

Key words: n-quadratic function, extremum.