УДК 004.85

#### М.А. Сивак, В.С. Тимофеев

# Настройка робастных нейронных сетей для решения задачи классификации

Рассматривается задача построения и настройки устойчивых к выбросам нейронных сетей. Для построения моделей используются различные робастные функции потерь Коши, Мешалкина, Geman—McCluer, Charbonnier и биквадратная функция потерь Тьюки. Исследования точности работы сетей проводились при различных значениях доли засоряющих наблюдений, различном числе эпох и различном объеме набора данных. Для каждой функции определены значения параметров, при которых достигается наилучшая точность работы, а также рекомендуемые интервалы значений параметра в зависимости от числа эпох обучения нейронной сети. Для сравнения рассматривались также классическая ИНС (с квадратичной функцией потерь) и робастная ИНС с функцией потерь Хьюбера. Анализ результатов показал, что использование робастного подхода дает значительный выигрыш в точности работы и в скорости обучения нейронной сети, однако выбор значения параметра функции потерь за пределами рекомендуемых интервалов может привести к ухудшению точности работы робастной сети. Ключевые слова: искусственная нейронная сеть, алгоритм обратного распространения ошибки, выбросы, вычислительный эксперимент, робастный подход, функция потерь, машинное обучение, задача классификации. doi: 10.21293/1818-0442-2021-24-3-26-32

Искусственные нейронные сети (ИНС) представляют собой один из наиболее популярных инструментов машинного обучения с учителем. Они применяются для решения разного рода задач, таких как задачи управления, прогнозирования или классификации. Однако при решении практических задач зачастую оказывается так, что данные содержат в себе нетипичные наблюдения (выбросы). Как правило, это отрицательным образом сказывается на точности работы нейронной сети [1–3].

Существует несколько способов решения этой проблемы. Наиболее популярные из них сводятся или к усложнению архитектуры используемой ИНС, или к предварительной обработке данных с целью исключения нетипичных наблюдений. Однако первый подход может привести к возникновению дополнительных затрат на вычислительные ресурсы, а второй – к потере важной информации.

Кроме того, можно использовать робастный подход [4, 5], который позволит не исключать выбросы, но снизить их негативное влияние при обучении нейронной сети. В [6] авторами сделана первая попытка использовать такой подход, чтобы получить нейронные сети, обладающие новыми свойствами. Рассматривались функции потерь Хьюбера, Уэлша, Эндрюса, Рамсея и «Fair». Проведенные исследования показали, что робастные ИНС способны дать значительный выигрыш в точности работы по сравнению с классической сетью, построенной с использованием квадратичной функции потерь. В настоящей работе предлагаются к рассмотрению еще несколько робастных функций потерь, а также модели, полученные с использованием этих функций.

#### Постановка задачи

Рассмотрим задачу обучения искусственной нейронной сети с простой архитектурой. Такая сеть включает в себя входной слой, выходной слой и один или несколько скрытых слоев. Введем следующие обозначения: *N* – число слоев нейронной сети,

 $x_i,\ i=1,...,\ T$ — значения, которые подаются на вход сети,  $y_k,\ k=1,...,\ Y$ — значения на выходном слое нейронной сети,  $w_{ij}^{(n-1)},\ i=1,...,\ l^{(n-1)},\ j=1,...,\ l^{(n)}$ — вес между j-м нейроном слоя n и i-м нейроном слоя n-1 ( $l^{(n)}$  — количество нейронов на слое n). Кроме того, обозначим через  $o_j^{(n)}$ — выходное значение j-го нейрона на слое n, через  $s_j^{(n)}$  — входное значение j-го нейрона на слое n, через  $\phi=\phi(z)$  — функцию активации.

Классическим алгоритмом обучения нейронных сетей является алгоритм обратного распространения ошибки [7]. Обучение нейронной сети с использованием этого алгоритма предполагает решение следующей задачи оптимизации:

$$E = \sum_{j=1}^{l^{(N)}} f(t_j, y_j) \to \min_{\substack{w_{ij}^{(1)}, \dots, w_{ij}^{(N-1)}}},$$
 (1)

где E — суммарная функция потерь,  $t_j$  — требуемое значение на j-м нейроне выходного слоя сети. Будем называть классической нейронной сетью такую сеть, в алгоритме обучения которой используется квадратичная функция потерь:

$$f(t_j, y_j) = \frac{1}{2} (y_j - t_j)^2$$
.

Недостаток классической нейронной сети заключается в том, что используемая функция потерь является чувствительной к наличию нетипичных наблюдений в данных. В [6] приводятся первые результаты исследования устойчивости построенных нейронных сетей к выбросам. В рамках текущей работы будут получены новые робастные сети, будет выполнена их настройка для решения задачи классификации. Кроме того, будет проведено сравнение моделей, полученных в рамках этой работы, с моделями, полученными в [6].

#### Робастная модификация алгоритма обучения

Для решения задачи (1) требуется вычислить производную суммарной функции потерь по весам нейронной сети, поэтому в алгоритме обратного распространения ошибки желательно использовать непрерывно дифференцируемую функцию потерь. В данной работе рассматриваются пять робастных функций потерь  $f_R(t_j,y_j)$  [7–9], которые удовлетворяют этому условию ( $\beta$  — параметр робастной функции потерь, принимающий неотрицательные значения):

1. Биквадратная функция потерь Тьюки:

$$f_{R}(t_{j}, y_{j}) = \begin{cases} \frac{(y_{j} - t_{j})^{6}}{6\beta^{4}} - \frac{(y_{j} - t_{j})^{4}}{2\beta^{2}} + \frac{(y_{j} - t_{j})^{2}}{2}, & |y_{j} - t_{j}| < \beta, \\ \frac{\beta^{2}}{6}, & |y_{j} - t_{j}| \ge \beta. \end{cases}$$

2. Функция потерь Коши:

$$f_R(t_j, y_j) = \ln(\frac{1}{2}(\frac{y_j - t_j}{\beta})^2 + 1).$$

3. Функция потерь Geman-McCluer:

$$f_R(t_j, y_j) = \frac{(y_j - t_j)^2}{\beta + (y_j - t_j)^2}.$$

4. Функция потерь Charbonnier:

$$f_R(t_j, y_j) = \sqrt{(\frac{y_j - t_j}{\beta})^2 + 1}.$$

5. Функция потерь Мешалкина:

$$f_R(t_j, y_j) = \beta^{-1} (1 - \exp(\frac{-\beta(y_j - t_j)^2}{2})).$$

Анализ выбранных функций был ранее представлен в [10] — были получены их производные и сделаны выводы о возможности применения этих функций при построении робастных нейронных сетей.

Опуская подробное описание всех шагов алгоритма, представленное в [11], приведем лишь конечное соотношение для вычисления частной производной суммарной функции потерь:

$$\frac{\partial E}{\partial w_{ii}^{(n-1)}} = \delta_j^{(n)} o_i^{(n-1)},$$

где  $\delta_j^{(n)}$  при замене квадратичной функции потерь на робастную будет вычисляться следующим образом:

$$\delta_j^{(n)} = \frac{\partial E}{\partial o_j^{(n)}} \frac{\partial o_j^{(n)}}{\partial s_j^{(n)}} = \begin{cases} \frac{\partial f_R(y_j, t_j)}{\partial y_j} \phi'(y_j), & n = N, \\ \sum\limits_{k=1}^{l^{(n+1)}} w_{jk}^{(n)} \delta_k^{(n+1)} \phi'(s_j^{(n)}), & \text{иначе.} \end{cases}$$

Искусственные нейронные сети, обучение которых проводилось с помощью робастной модификации алгоритма обратного распространения ошибки, представляют собой класс абсолютно новых

нейронных сетей, в связи с чем возникает необходимость в исследовании их свойств. Далее в работе приводятся наилучшие значения параметров робастных функций потерь, полученные в ходе настройки сетей, а также результаты исследования устойчивости построенных моделей к наличию выбросов.

#### Результаты исследований

В рамках данной работы исследования проводились в три этапа. На первых двух этапах использовался набор данных «Ирисы Фишера» [12], состоящий из 150 объектов  $X_m$ ,  $m=1,\ldots,150$ , описывающихся четырьмя признаками  $x_{mi}$ ,  $i=1,\ldots,4$ . Все имеющиеся объекты делятся на три сбалансированных класса  $y_k$ ,  $k=1,\ldots,3$ . Выбранный набор данных был разделен на обучающую и тестовую выборки, причем обучающая выборка  $L=\{X_I,\ldots,X_{|L|}\}$  включала в себя 80% объектов (|L|=120 объектов — объем обучающей выборки), а тестовая выборка  $D=\{X_{|L|+1},\ldots,X_{|X|}\}$  включала в себя оставшиеся 20% объектов (|D|=30 объектов — объем тестовой выборки).

Исходя из конфигурации анализируемого набора данных, для исследования была выбрана однослойная нейронная сеть, имеющая следующую архитектуру. Скрытый слой сети состоял из 4 нейронов, входной слой – также из 4 нейронов (по количеству признаков объектов), выходной – из 3 нейронов (по количеству классов). В качестве функции активации  $\varphi = \varphi(z)$  использовалась сигмоида [11]

$$\varphi(z) = \frac{1}{1 + e^z}.$$

Поскольку в анализируемом наборе данных классы сбалансированы по объему, оценка точности работы построенных моделей проводилась только на основе метрики  $\alpha$  [13]:

$$\alpha = \frac{D_{\text{corr}}}{|D|} 100\% ,$$

где  $D_{\text{corr}}$  – количество объектов, отнесенных к правильному классу.

Набор данных «Ирисы Фишера» не является синтетическим – это означает, что данные представляют собой результат реальных измерений и могут содержать погрешности (фоновый шум). В связи с этим главным образом исследовалось влияние на работу нейронных сетей доли засоряющих наблюдений в выборке. Зашумление обучающей выборки выполнялось по третьему и четвертому признаку:

$$\tilde{x}_{mi} = x_{mi} + \varepsilon_{mi}, i = 3, 4, \qquad (2)$$

где  $\varepsilon_{mi}$  — случайные, независимые, одинаково распределенные ошибки. Данные ошибки имеют следующую функцию распределения:

$$F_i(x) = (1 - \lambda)F_1(x, 0, \sigma_{i1}) + \lambda F_2(x, 0, \sigma_{i2}), i=3,4,$$
(3)

где  $F_j(x, 0, \sigma_{ij}), j = 1, 2$  — функция нормального распределения с нулевым математическим ожиданием и дисперсией  $\sigma_{ij}^2$ ,  $\lambda \in [0, 1]$  — параметр смеси, играющий роль доли засоряющих наблюдений.

В данном исследовании задавались не сами значения дисперсий  $\sigma_{i1}^2$  и  $\sigma_{i2}^2$ , а соответствующие им значения уровня шума [14], причем полагалось, что  $\sigma_{i1}^2 < \sigma_{i2}^2$ :

$$\rho_{ij} = \frac{\sigma_{ij}}{c} 100\%,$$

где  $c^2$  — дисперсия незашумленной выборки. Для третьего признака дисперсия  $\sigma_{31}^2$  соответствовала уровню шума  $\,\rho_{31} = 30\%$  , дисперсия  $\sigma_{32}^2 - \,$  уровню шума  $\,\rho_{32} = 120\%$  ; для четвертого признака  $\sigma_{41}^2$  соответствовала уровню шума  $\rho_{41} = 40\%$ , а  $\sigma_{42}^2$  – уровню шума  $\rho_{42} = 150\%$ .

Первый этап исследований заключался в поиске таких значений параметра в робастных функций потерь, при которых точность работы нейронной сети была бы наилучшей. Значения параметра рассматривались на различных интервалах ( $\beta_{min}$ ,  $\beta_{max}$ ), которые определялись в ходе предварительного исследования. Для всех рассматриваемых функций потерь предварительно фиксировались левая и правая границы интервалов ( $\beta_{min} = 0.00$ ,  $\beta_{max} = 20.00$ ), шаг разбиения 0,10. При этом было зафиксировано значение доли засоряющих наблюдений  $\lambda = 0.25$ . Затем для каждого значения в из этого интервала проводилось по 10 вычислительных экспериментов, по результатам которых фиксировались, а затем усреднялись значения метрики α на 500 эпох обучения.

Предварительное исследование позволило скорректировать правую границу интервала для каждой функции потерь - она определялась в соответствии с тем, начиная с какого значения в точность работы сети переставала значимо изменяться. Помимо этого, был изменен шаг разбиения для функции потерь Geman-McCluer: выбранный изначально шаг оказался слишком маленьким - с каждым новым значением В при таком разбиении точность работы сети изменялась незначительно (не более чем на 0.5%). Полученные в результате предварительного исследования границы и шаг разбиения интервалов приводятся в табл. 1.

Таблица 1 Интервалы значений параметра

рооастных функции потерь							
Функция потерь	$\beta_{min}$	$\beta_{max}$	Шаг				
Биквадратная Тьюки	0,00	5,00	0,10				
Коши	0,00	7,00	0,10				
Geman-McCluer	0,00	10,00	0,50				
Charbonnier	0,00	7,00	0,10				
Мешалкина	0,00	5,00	0,10				

Следует отметить, что для функции Geman-McCluer, помимо значений из интервала, приведенного в таблице, рассматривались также значения  $\beta = \{20, 50, 100\}$ , так как было отмечено улучшение точности работы нейронной сети с ростом значения параметра.

После того как были определены интервалы, на которых рассматривались значения параметра В,

исследования проводились следующим образом. Значение доли засоряющих наблюдений в обучающей выборке изменялось от 0,05 до 0,40 с шагом в 0,05. Для каждого λ при всех β из указанных интервалов вычислялось значение метрики α и фиксировалось при различном числе эпох, в течение которых проходило обучение нейронной сети (50, 100 и далее до 1000 с шагом в 100 эпох). Исходя из соотношения числа эпох обучения и значения метрики α, можно судить о скорости обучения нейронной сети - чем больше значение метрики и чем меньше число эпох, тем быстрее обучается ИНС.

В ходе исследований для каждой робастной сети проводилось по 100 вычислительных экспериментов, результаты которых после усреднялись. На основании полученных результатов для каждой нейронной сети были зафиксированы наилучшие значения параметра  $\beta$  – такие, при которых точность классификации была максимальной. Эти значения приводятся в табл. 2 для каждого  $\lambda$  на 100, 300, 500 и 1000 эпох обучения.

Таблица 2 Наилучшие значения параметра робастных

функций потерь								
λ	h	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$		
0,05	100	0,7	1,7	2,0	1,4	1,2		
	300	2,9	1,6	10,0	1,8	2,4		
	500	4,7	1,5	20,0	2,2	3,1		
	1000	5,2	1,2	20,0	3,6	3,8		
	100	0,8	1.9	3,0	1,4	1,3		
0.10	300	3,0	1,3	10,0	3,2	2,6		
0,10	500	5,0	1,3	20,0	4,0	3,2		
	1000	5,6	1,2	20,0	4,0	3,4		
	100	0,7	1,6	2,0	2,0	1,2		
0.15	300	2,8	1,3	10,0	3,2	2,5		
0,15	500	5,1	1,2	20,0	3,8	3,1		
	1000	5,7	1,2	20,0	3,7	4,1		
	100	2,2	1,5	3,5	2,3	2,1		
0,20	300	3,3	1,3	10,0	3,3	2,7		
0,20	500	5,0	1,3	20,0	3,3 3,5	3,3		
	1000	6,3	1,2	20,0	3,5	3,6		
	100	1,1	1,2	3,5	2,1	1,5		
0.25	300	2,7	1,3	10,0	3,3	2,4		
0,25	500	4,5	1,2	20,0	3,5	3,0		
	1000	6,9	1,3	20,0	3,3	4,4		
	100	1,1	1,6	4,0	1,9	1,5		
0,30	300	2,4	1,3	9,5	2,8	2,2		
	500	3,6	1,3	20,0	3,6	2,8		
	1000	6,9	1,3	20,0	2,8	3,8		
	100	1,3	1,5	5,0	2,0	1,7		
0,35	300	3,1	1,3	10,0	3,3	2,4		
	500	5,2	1,3	20,0	3,8	3,2		
	1000	6,9	1,3	20,0	3,8	4,5		
_	100	1,1	1,5	4,0	2,2	1,4		
0,40	300	3,0	1,3	10,0	3,5	2,5		
0,40	500	4,1	1,2	20,0	3,8	2,9		
	1000	6,9	1,4	20,0	3,0	3,9		

Здесь и далее в таблицах приняты следующие обозначения столбцов: h – число эпох,  $f_1$  – функция потерь Коши,  $f_2$  – биквадратная функция потерь Тьюки,  $f_3$  – функция потерь Geman–McCluer,  $f_4$  – функция потерь Мешалкина,  $f_5$  – функция потерь Charbonnier.

Полученные результаты позволяют сделать вывод о том, что ни для одной из рассмотренных функций нет явной зависимости между значением параметра  $\beta$  и долей засоряющих наблюдений  $\lambda$ . Однако для всех функций наблюдается зависимость значения  $\beta$  от числа эпох обучения. Для ИНС с функциями потерь Коши, Geman–McCluer, Мешалкина и Charbonnier в целом характерен рост наилучшего значения параметра при увеличении числа эпох обучения нейронной сети. Для нейронной сети с биквадратной функцией потерь Тьюки характерно убывание значения  $\beta$  с ростом числа эпох.

На рис. 1 продемонстрирована полученная зависимость точности классификации от значения параметра  $\beta$  для нейронной сети с функции потерь Charbonnier при различном числе эпох обучения и доли засоряющих наблюдений  $\lambda=0,25$ . Нетрудно заметить, что с ростом числа эпох обучений возрастает и наилучшее значение параметра функции, например, для 100 эпох  $\beta=1,5$ , а для 500 эпох  $\beta=3,0$ . Однако слишком большие значения параметра приводят к снижению точности работы нейронной сети.

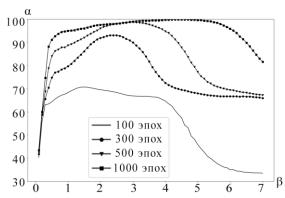


Рис. 1. Зависимость наилучшего значения параметра функции потерь Charbonnier от числа эпох обучения

В табл. 3 приводятся интервалы значений параметра в для ИНС с рассматриваемыми робастными функциями потерь, на которых точность классификации была наилучшей. На основе полученных результатов можно сформировать следующие рекомендации относительно выбора значения параметра робастных функций потерь. Для функций потерь Коши, Charbonnier и Мешалкина значение параметра следует постепенно увеличивать с ростом числа эпох, однако следует помнить о том, что слишком большие значения в могут негативно сказаться на точности классификации. Так, для функции потерь Коши не следует выбирать значения параметра больше 7,0, для функции потерь Charbonnier больше 4,5, для функции потерь Мешалкина больше 4,0. Для биквадратной функции потерь Тьюки, наоборот, с ростом числа эпох обучения следует уменьшить значение β. Что касается функции потерь Geman-McCluer, то с ростом числа эпох наилучшее значение параметра возрастает достаточно резко при малом числе эпох (до 300) можно выбирать в на интервале [2,0; 10,0], а с ростом числа эпох можно рассмотреть значения параметра больше 20.

Таблица 3 Рекомендуемые интервалы значений параметра побастных функций потерь

робастивіх функции потерв							
Функция потерь	100 эпох	101–300 эпох	301–500 эпох				
Биквадратная Тьюки	[1,5; 1,7]	[1,3; 1,6]	[1,2; 1,5]				
Коши	[0,7; 1,3]	[2,4; 3,3]	[3,6; 5,2]				
Geman-McCluer	[2,0; 5,0]	[9,5; 10]	≥ 20				
Charbonnier	[1,2; 1,7]	[2,2; 2,7]	[2,8; 3,3]				
Мешалкина	[1,4; 2,2]	[1,8; 3,3]	[2,2; 3,8]				

Второй этап исследований заключался в сравнении построенных ИНС по точности работы. Для этого при каждой доле засоряющих наблюдений фиксировались значения метрики  $\alpha$ , полученные при наилучших значениях параметра  $\beta$ . В табл. 4 приводятся результаты для всех рассматриваемых робастных функций потерь, а также результаты для ИНС с функцией потерь Хьюбера (столбец  $f_6$ ) и для классической ИНС (столбец  $f_7$ ), полученные авторами ранее в [6].

Таблица 4 Точность классификации

при различных значениях доли выбросов								
λ	h	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$
	100	76,9	77,0	76,7	77,0	76,9	76,7	75,0
0,05	300	91,3	90,6	91,3	90,6	91,3	90,7	89,2
	500	96,8	94,5	96,7	94,5	96,7	94,3	91,4
	100	72,3	72,1	72,3	72,1	72,3	71,9	71,6
0,10	300	89,8	82,1	89,4	82,5	89,8	83,2	77,8
	500	95,7	91,2	95,5	91,5	95,6	91,3	83,5
	100	71,2	71,4	71,6	71,6	71,1	71,5	69,8
0,15	300	86,6	84,0	86,5	84,8	86,6	84,8	77,8
	500	94,3	89,9	94,3	90,8	94,3	90,6	82,4
	100	68,0	67,7	68,0	67,8	68,1	67,7	67,2
0,20	300	79,8	77,9	79,8	78,7	79,8	78,6	72,6
	500	91,3	87,9	91,3	88,3	91,3	89,1	78,9
	100	70,7	69,5	70,6	69,6	70,1	69,3	67,3
0,25	300	93,1	86,7	93,1	87,4	93,1	87,3	76,6
	500	98,6	94,1	98,6	94,3	89,7	94,0	85,2
	100	69,1	67,7	69,1	67,6	69,1	67,3	66,3
0,30	300	87,7	85,6	87,7	84,7	87,5	85,1	76,4
	500	95,5	92,2	94,6	92,4	95,5	92,2	84,8
0,35	100	69,6	67,9	69,6	67,8	69,5	67,7	65,6
	300	89,5	82,9	89,2	82,6	89,5	84,1	74,7
	500	96,0	90,1	95,9	89,9	96,0	90,8	81,8
	100	69,4	69,0	69,4	69,1	69,4	69,0	67,0
0,40	300	86,9	83,1	86,4	84,1	87,1	84,9	74,3
	500	95,4	89,9	95,2	90,9	95,4	90,7	80,3

Значения метрики  $\alpha$  рассматривались на 100, 300 и 500 эпох обучения. С ростом числа эпох точность работы сетей достигала значений 98–100%, что позволяет сделать вывод о потенциально возможном переобучении сетей. Рассматривать такие случаи при сравнении точности работы сетей нецелесообразно, поэтому результаты при большем числе эпох не приводятся.

Анализируя полученные результаты, можно сделать вывод о том, что при малом числе эпох (100) даже при небольших значениях доли засоряющих наблюдений ( $\lambda=0.05$  и  $\lambda=0.10$ ) робастные ИНС дают выигрыш в точности работы в среднем около 1,5%. С ростом числа эпох и доли засоряющих

наблюдений разница в точности работы классической и робастных ИНС становится все более существенной. Так, при классификации сильно зашумленных данных (доля выбросов  $\lambda=0,35$  и  $\lambda=0,40$ ) после 500 эпох обучения робастные сети начинают работать точнее в среднем на 11,7%, а в отдельных случаях позволяют получить выигрыш в точности до 15% (нейронная сеть с функцией потерь Charbonnier).

Из табл. 4 видно, что все робастные нейронные сети можно условно разделить на две группы. Первая группа будет включать в себя сети с функциями потерь Коши, Geman-McCluer и Charbonnier, вторая сети с функциями потерь Мешалкина, Хьюбера и биквадратной функцией потерь Тьюки. Между собой сети из каждой группы при всех значениях  $\lambda$ сопоставимы по точности работы, однако значение метрики α для сетей, входящих во вторую группу, в среднем на 3,5% ниже, чем для сетей из первой группы. Кроме того, для всех построенных сетей можно заметить снижение точности работы при доле засоряющих наблюдений  $\lambda = 0.20$ , а также резкий рост значения метрики  $\alpha$  при доле выбросов  $\lambda = 0.25$ . Такая разница в точности работы сетей объясняется конфигурацией выборок - тем, насколько далеко друг от друга расположены различные классы объектов. Подробное объяснение этого эффекта приво-

На третьем этапе исследований изучалось влияние объема набора данных, для которого выполнялась классификация, на точность работы построенных нейронных сетей. Поскольку набор «Ирисы Фишера» является достаточно малым, а использование технологии bootstrap привело бы к внесению дополнительного шума в данные, на этом этапе использовался синтетический набор данных, полученный с использованием генератора данных языка Руthon 3.6 [15]. Данный инструмент позволяет генерировать облака нормально распределенных данных, используя которые, можно сформировать признаки объектов для классификации.

В рамках данной работы был сгенерирован набор данных, аналогичный по структуре набору «Ирисы Фишера»: все объекты делились на три сбалансированных класса, каждый объект описывался четырьмя признаками. Наиболее различительными признаками являлись признаки 3 и 4, зашумление выборки выполнялось в соответствии с (2), (3). Рассматривались наборы данных, включающие в себя 150, 450, 600 и 900 объектов (столбец |X|), при доле выбросов  $\lambda = 0.25$ . Для робастных функций потерь значения параметра  $\beta$  выбирались согласно данным ранее рекомендациям. Точность работы нейронных сетей, как и на предыдущем этапе, фиксировалась после 100, 300 и 500 эпох обучения. Результаты данного этапа исследований представлены в табл. 5.

Анализируя представленные в табл. 5 значения метрики α, можно сказать, что в целом с ростом числа объектов в наборе данных скорость обучения робастных нейронных сетей падает. Хотя увеличение числа объектов со 150 до 450 зачастую не сказыва-

ется на точности работы нейронных сетей негативным образом, дальнейшее увеличение числа объектов приводит к снижению точности работы нейронных сетей. Так, по сравнению с набором, состоящим из 450 объектов, точность классификации для набора из 600 объектов снизилась примерно на 1%, а для набора из 900 объектов — на 2%. Это объясняется тем, что с ростом объема данных в целом число нетипичных наблюдений также растет.

Таблица 5 Точность классификации при различном объеме набора данных

при различном объеме наобра данных								
/X/	h	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$
	n	$\beta = 4.5$	$\beta = 1,3$	$\beta = 20,0$	$\beta = 3.3$	$\beta = 2.8$	$\beta = 0.2$	-
	100	88,4	83,8	78,3	84,6	97,8	83,0	74,2
150	300	99,9	88,8	99,9	89,3	99,9	90,9	85,9
	500	99,9	92,7	100,0	93,4	99,9	95,6	91,0
	100	86,5	67,0	92,0	67,0	76,1	67,0	67,0
450	300	97,6	88,9	97,7	89,5	97,3	92,3	87,6
	500	99,2	95,7	99,3	95,9	99,1	97,4	94,1
	100	72,5	68,1	78,6	68,0	67,7	67,5	67,5
600	300	96,4	91,3	96,4	91,9	96,1	93,9	87,3
	500	98,2	95,5	98,3	95,6	98,0	96,5	93,0
	100	68,0	69,6	70,0	69,5	67,0	68,5	68,3
900	300	96,8	88,5	97,1	88,7	96,1	91,3	88,0
	500	98,3	92,2	98,5	92,6	97,9	94,6	91,2

По сравнению с классической ИНС наименьший выигрыш в точности работы на данном этапе дала биквадратная функция потерь Тьюки (от 1 до 2,5%), а наибольший — функции потерь Коши и Geman–McCluer (от 5,2 до 9%). Однако очевидно, что нейронные сети с функциями потерь Коши и Geman–McCluer обучаются быстрее остальных и уже на 300 эпох достигают очень высоких значений точности классификации. Дальнейшее увеличение числа эпох приведет к переобучению этих сетей.

#### Заключение

В ходе выполнения работы были построены пять новых робастных нейронных сетей, в которых использовались следующие функции потерь: Коши, биквадратная функция потерь Тьюки, Мешалкина, Geman-McCluer и Charbonnier. Кроме того, рассматривались построенные ранее классическая ИНС и робастная ИНС с функцией потерь Хьюбера. Была исследована устойчивость моделей к наличию выбросов путем ряда вычислительных экспериментов. Результаты, полученные в ходе проведения исследований, позволили сформировать рекомендации относительно выбора значения параметра В для рассмотренных робастных функций потерь. Для каждой функции были определены наилучшие значения, а также рекомендуемые интервалы значений параметра в зависимости от числа эпох обучения нейронной сети. Было показано, что выбор значений параметра за пределами рекомендуемого интервала приводит к ухудшению точности работы сети.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-37-90077.

#### Литература

- 1. Ланкин Ю.П. Нейросетевой анализ сложноорганизованных экологических данных / Ю.П. Ланкин, Т.Ф. Басканова, Т.И. Лобова // Современные проблемы науки и образования. 2012. № 4 [Электронный ресурс]. Режим доступа: https://www.science-education.ru/ru/article/view?id=6754, свободный (дата обращения: 05.10.2021).
- 2. Манжула В.Г. Нейронные сети Кохонена и нечеткие нейронные сети в интеллектуальном анализе данных / В.Г. Манжула, Д.С. Федяшов // Фундаментальные исследования. 2011. № 4. С. 108–115 [Электронный ресурс]. Режим доступа: https://www.fundamentalresearch.ru/ru/article/view?id=21239, свободный (дата обращения: 05.10.2021).
- 3. Глубокие нейросети. Ч. І: Подготовка данных [Электронный ресурс]. Режим доступа: https://www.mql5.com/ru/articles/3486, свободный (дата обращения: 05.10.2021).
- 4. Fan J. Local Polynomial Modelling and Its Applications. / J. Fan, I. Gijbels. UK; London: Chapman & Hall, 1996. 360 p.
- 5. Huber J.P. Robust statistics, 2nd edition. New Jersey, Hoboken: Wiley, 2009. 370 p.
- 6. Сивак М.А. Построение робастных нейронных сетей с различными функциями потерь / М.А. Сивак, В.С. Тимофеев // Системы анализа и обработки данных. 2021. № 2 (82). С. 67–83.
- 7. Демиденко Е.З. Линейная и нелинейная регрессии. М.: Финансы и статистика, 1981. 304 с.
- 8. Black M.J. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. / M.J. Black, A. Rangarajan // International Journal of Computer Vision. 1996. Vol. 19. P. 57–91.
- 9. Fair Loss: Margin-Aware Reinforcement Learning for Deep Face Recognition / В. Liu, W. Deng, Y. Zhong, M. Wang, J. Hu, X. Tao, Y. Huang [Электронный ресурс]. Режим доступа: https://openaccess.thecvf.com/content\_ICCV\_2019/papers/Liu\_Fair\_Loss\_Margin-Aware\_Reinforcement\_Learning\_for\_Deep\_Face\_Recognition\_ICCV\_2019\_paper.pdf, свободный (дата обращения: 19.10.2021).
- 10. Сивак М.А. Исследование применимости робастных функций потерь в нейронных сетях // Сб. науч. трудов НГТУ. -2020. -№ 4 (99). -C. 50–58.
- 11. Bishop C. Neural Networks for Pattern Recognition. US, New York: Oxford University Press, 1995. 502 p.
- 12. UCI Machine Learning Repository [Электронный ресурс]. Режим доступа: http://www.ics.uci.edu/mlearn/MLRepository.html, свободный (дата обращения: 05.10.2021).
- 13. Classification: Accuracy [Электронный ресурс]. Режим доступа: https://developers.google.com/machine-learning/crash-course/classification/accuracy, свободный (дата обращения: 05.10.2021).
- 14. Ивахненко А.Г. Помехоустойчивость моделирования /А.Г. Ивахненко, В.С. Степашко. Киев: Наук. думка, 1985. 216 с.
- 15. How to Generate Test Datasets in Python with scikit-learn [Электронный ресурс]. Режим доступа: https://machinelearningmastery.com/generate-test-datasets-python-scikit-learn/, свободный (дата обращения: 05.10.2021).

#### Сивак Мария Алексеевна

Аспирант каф. теоретической и прикладной информатики (ТПИ) Новосибирского государственного технического университета (НГТУ)

К. Маркса пр-т, 20, г. Новосибирск, Россия, 630073

Тел.: +7-951-388-14-89

Эл. почта: pepelyaeva@ami.nstu.ru

#### Тимофеев Владимир Семенович

Докт. техн. наук, профессор каф. ТПИ НГТУ К. Маркса пр-т, 20, г. Новосибирск, Россия, 630073

Тел.: +7-913-952-26-46

Эл. почта: v.timofeev@corp.nstu.ru

### Sivak M.A., Timofeev V.S.

## Adjusting robust neural networks for solving the classification problem

The paper highlights the problem of building and adjusting robust neural networks applying different loss functions for solving the classification problem. The considered functions are those of Cauchy, Meshalkin, Geman-McCluer, Charbonnier and Tukey's Biweight losses. The accuracy of classification is examined for the different values of outliers' fraction, for several values of learning epochs count and for datasets with various sizes. For all obtained networks the parameter values that maximize the accuracy, are defined. The best practices for choosing the parameter values depending on epoch count are also defined for all the loss functions. The ordinary neural network (with quadratic loss) and the robust neural network applying the Huber loss are also considered. The analysis of the results shows that the use of robust approach can significantly increase the learning rate and the classification accuracy, however, choosing the incorrect parameter value can decrease the accuracy of classification.

**Keywords:** artificial neural network, error back-propagation algorithm, outliers, computational experiment, robust technique, loss function, machine learning, classification problem. **doi:** 10.21293/1818-0442-2021-24-3-26-32

#### References

- 1. Lankin Yu.P., Baskanova T.F., Lobova T.I. [Neural network analysis of complicated ecological data]. *Sovremennye problemy nauki i obrazovaniya*, 2012, no. 4 (in Russ.) Available at: https://www.science-education.ru/ru/article/view?id=6754, free (Accessed: October 05, 2021).
- 2. Manzhula V.G., Fedyashov D.S. [Kohonen neural networks and fuzzy neural networks in data mining]. *Fundamentalnye issledovaniya*, 2011, no. 4, pp. 108–115 (in Russ.) Available at: https://www.fundamental-research.ru/ru/article/view?id=21239, free (Accessed: October 05, 2021).
- 3. Glubokie neiroseti (Chast' 1). Podgotovka dannyh. [Deep neural networks (Part 1). Data preprocessing.]. Available at: https://www.mql5.com/ru/articles/3486, free (Accessed: October 05, 2021).
- 4. Fan J., Gijbels I. *Local Polynomial Modelling and Its Applications*, Chapman & Hall, UK; London, 1996, 360 p.
- 5. Huber J.P. Robust statistics, 2nd edition, Wiley, Hoboken, New Jersey, 2009, 370 p.

- 6. Sivak M.A., Timofeev V.S. [Building robust neural networks using different loss functions]. *Analysis and Data Processing Systems*, 2021, no. 2 (82), pp. 67–82 (in Russ.).
- 7. Demidenko E.Z. *Linejnaja i nelinejnaja regressii* [Linear and non-linear regression]. Moscow, Finansy i statistika Publ., 1981. 304 p. (in Russ.).
- 8. Black M.J., Rangarajan A. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision, *International Journal of Computer Vision*, 1996, vol. 19, pp. 57–91.
- 9. B. Liu, W. Deng, Y. Zhong, M. Wang, J. Hu, X. Tao, Y. Huang. Fair Loss: Margin-Aware Reinforcement Learning for Deep Face Recognition. Available at: https://openaccess.thecvf.com/content\_ICCV\_2019/papers/Liu\_Fair\_Loss\_Margin-Aware\_Reinforcement\_Learning\_for\_Deep \_Face\_Recognition\_ICCV\_2019\_paper.pdf, free (Accessed: October 19, 2021).
- 10. Sivak M.A. [The research on using robust loss functions for neural networks]. *Transaction of Scientific Papers of the Novosibirsk State Technical University*, 2020, no. 4 (99), pp. 50–58 (in Russ.).
- 11. Bishop C. *Neural Networks for Pattern Recognition*. US, New York: Oxford University Press, 1995. 502 p.
- 12. UCI Machine Learning Repository. Available at: http://www.ics.uci.edu/mlearn/MLRepository.html, free (Accessed: October 05, 2021).
- 13. Classification: Accuracy. Available at https://developers.google.com/machine-learning/crash-

- course/classification/accuracy, free (Accessed: October 05, 2021)
- 14. Ivahnenko A.G., Stepashko V.S. *Pomehoustojchivost' modelirovanija* [Noise-resistant modelling]. Kiev, Nauk. Dumka Publ., 1985. 21 p. (in Russ.).
- 15. How to Generate Test Datasets in Python with scikitlearn. Available at: https://machinelearningmastery.com/generate-test-datasets-python-scikit-learn/, free (Accessed: October 05, 2021).

#### Maria A. Sivak

Postgraduate Student, Department of Theoretical and Applied Computer Science, Novosibirsk State Technical University (NSTU) 20, K. Marksa pr., Novosibirsk, Russia, 630073

Phone: +7-951-388-14-89 Email: pepelyaeva@ami.nstu.ru

#### Vladimir S. Timofeev

Doctor of Science in Engineering, Professor, Department of Theoretical and Applied Computer Science, NSTU 20, K. Marksa pr., Novosibirsk, Russia, 630073

Phone: +7-913-952-26-46 Email: v.timofeev@corp.nstu.ru