УДК 004.032.26 **DOI** 10.15622/ia.2022.21.6

Г.Б. Маршалко, Ю.А. Труфанова

ПОЛИНОМИАЛЬНЫЕ АППРОКСИМАЦИИ НЕКОТОРЫХ ФУНКЦИЙ АКТИВАЦИИ НЕЙРОННЫХ СЕТЕЙ

Маршалко Г.Б., Труфанова Ю.А. **Полиномиальные аппроксимации некоторых функций активации нейронных сетей.**

Аннотация. Активное внедрение систем машинного обучения ставит актуальную задачу обеспечения их защиты от различных типов атак, направленных на нарушение свойств конфиденциальности, целостности и доступности как обрабатываемых данных, так и обучаемых моделей. Одним из перспективных направлений защиты является разработка конфиденциальных систем машинного обучения, использующих гомоморфные схемы шифрования для защиты моделей и данных. Однако такие схемы могут обрабатывать только полиномиальные функции, что в свою очередь ставит задачу построения полиномиальных аппроксимаций используемых в нейросетевых моделях нелинейных функций. Целью настоящей работы является построение наиболее точных аппроксимаций некоторых широко используемых функций активаций нейронных сетей, а именно ReLU, логистического сигмоида и гиперблолического тангенса, при ограничениях на степень аппроксимирующего полинома, а также оценка влияния точности такой аппроксимации на результат работы нейронной сети в целом. В отличие от опубликованных ранее работ рассматриваются и сравниваются различные способы построения аппроксимирующих полиномов, вводятся метрики точности приближения, приводится конкретный вид аппроксимирующих полиномов, а также соответствующие значения точности приближения. Проводится сравнение с аппроксимациями, приведенными в опубликованных ранее работах. В заключение для простейшей нейронной сети экспериментально оценено влияние точности приближения аппроксимирующего полинома на величину отклонения значений выходных нейронов такой сети от соответствующих значений выходных нейронов исходной сети. Результаты показывают, что для функции ReLU наилучшее приближение может быть получено с помощью численного метода, а для логистического сигмоида и гиперболического тангенса – с помощью полиномов Чебышева. При этом наилучшее приближение из трех рассмотренных функций получено для функции ReLU. Полученные результаты в дальнейшем могут быть использованы при построении аппроксимаций функций активации в конфиденциальных системах машинного обучения.

Ключевые слова: функция активации, ReLU, гиперболический тангенс, логистический сигмоид, гомоморфное шифрование, BGV, CKKS, нейронная сеть, полиномиальная аппроксимация, конфиденциальное машинное обучение.

1. Введение. Одним из способов защиты систем машинного обучения от широкого спектра существующих атак [1] является построение систем конфиденциального машинного обучения с использованием схем гомоморфного шифрования [2–4]. Примерами используемых схем шифрования являются BGV [5], $CKKS$ [6] и др.

Необходимым условием реализации вычислений при таком методе защиты является замена нелинейных функций активации нейронов их полиномиальными представлениями. Если степень нелинейности таких представлений будет относительно мала, то применение гомоморфных

схем, таких, например, как BGV позволит обеспечить гарантированную защиту данных и/или моделей при сравнительно небольшом снижении скорости работы. Это связано с тем, что современные схемы гомоморфного шифрования основаны на сложности решения задачи обучения с ошибками (LWE) или обучения с ошибками в кольце (RLWE) и используют при формировании шифртекста шумовые последовательности. При выполнении арифметических операций «амплитуда» шума увеличивается и, начиная с некоторого момента, корректное расшифрование становится невозможным. Особенно быстро шум накапливается при умножении. Снизить уровень шума можно с помощью т.н. операции перезагрузки, однако она выполняется достаточно медленно, что существенно снижает скорость шифрования. В этой связи важно ограничивать степень гомоморфно вычисляемого полинома. С другой стороны, для сохранения свойств обученных моделей машинного обучения нам необходимо использовать достаточно точные аппроксимации исходных функций, что, очевидно, накладывает ограничения на минимально допустимое значение степени полинома.

Исследователи в области применения гомоморфного шифрования к нейронным сетям рассматривали и другой способ обработки нелинейных функций. Он заключается в предварительном вычислении значений функции активации для набора дискретных значений и сохранении этих результатов в поисковую таблицу [7]. Однако, в этом случае качество работы нейронной сети напрямую зависит от числа записей в поисковой таблице, то есть данный подход требует больших затрат по памяти. Поэтому ограничимся рассмотрением полиномиальных аппроксимаций для функций активации.

В доступной литературе по данной тематике практически отсутствуют подробные описания методов построения полиномиальных аппроксимаций функций активации нейронных сетей, конкретного вида получающихся полиномов и обоснования точности аппроксимации.

Одной из первых работ, в которой были описаны принципы применения гомоморфных схем шифрования в задачах машинного обучения, была работа исследователей Microsoft [2].

В некоторых работах, в частности [2, 8], было предложено заменить функцию активации $ReLU$ полиномом x^2. Однако, в работе [9] было показано, что такая аппроксимация хорошо работает только на нейронный сетях с небольшим числом слоев. Для глубоких нейронных сетей качество работы при замене функций активации $ReLU$ на $x^2 + x$ выше, чем при замене на x^2, хотя и проигрывает оригинальной сети.

В работе [3] авторы кратко описали подходы, которые могут быть применены для аппроксимации таких широко используемых функций активации как $ReLU$, *логистический сигмоид* и *гиперболический тангенс*. Однако конкретный вид аппроксимирующих полиномов приведен не был.

В репозитории GitHub представлен проект [10], реализующий описанный в работе [3] подход. Авторы проекта представили полиномы, полученные при помощи четырех методов аппроксимации для функций активации $ReLU$ и *гиперболический тангенс* степени, не превышающей 4, причем ни один из графиков полученных функций не соответствует результатам, представленным в работе [3].

В работе [11] представлены результаты исследований по оценке близости полиномиальных приближений нейронных сетей к исходным в зависимости от параметра точности α. Авторы рассматривают функцию ошибки аппроксимированной модели по отношению к исходной, показывая, что она может быть сделана сколь угодно малой: $|F(x) - F_\alpha(x)| \leqslant Const \cdot 2^{-\alpha}$. Здесь $F(x)$ – это функция, описывающая исходную нейронную сеть, $F_\alpha(x)$ – приближение с точностью α, в котором степень аппроксимирующих многочленов определенным образом зависит от данного параметра, $Const$ – некоторая константа. В этой работе авторы в явном виде приводят вид аппроксимирующих полиномов, позволяющих получить существенном более точные нейросетевые модели по сравнению с работой [3]. Вместе с тем полученные ими полиномы имеют степень не менее 7, что, как было указано выше, существенно замедляет скорость работы защищенной модели.

Указанные выше работы не раскрывают деталей процесса выработки приведенных в них аппроксимирующих полиномов. В связи с этим вопрос о выборе наилучшего метода приближения, равно как и точности получаемых приближений, остается открытым. Настоящая работа посвящена исследованию различных способов приближения нелинейных функций активации, оценке взаимосвязи точности и степени нелинейности таких приближений и получению явных представлений таких приближений.

2. Основные сведения и обозначения.

2.1. Функции активации. Рассматриваются следующие наиболее часто используемые при построении нейронных сетей функции активации (Рис. 1):

– *Rectified linear units (ReLU)*: $ReLU(x) = \max(0, x)$;
– *Логистический сигмоид (Sigmoid)*: $\sigma(x) = \frac{1}{1+e^{-x}}$;
– *Гиперболический тангенс*: $tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.

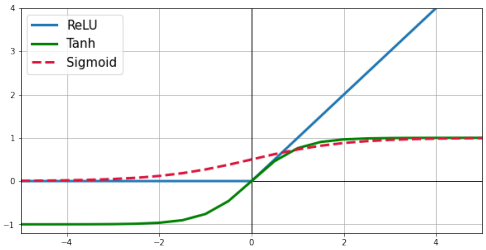

Рис. 1. Исследуемые функции активации

2.2. Методы аппроксимации. Пусть $f(x)$ – функция, которую мы хотим аппроксимировать полиномом $p(x)$ степени, не выше N. Функция $f(x)$ определена на отрезке $[a, b]$. Точки $x_i \in [a, b]$ – узлы интерполяции $(i = 1, \ldots, m;\ m \geqslant N)$.

Будем исследовать следующие методы аппроксимации:

1. *Ряд Тейлора в окрестностях точки* $z \in [a, b]$

Разложение строится численным методом. Основная идея заключается в построении полинома, приближающего исходную функцию на заданном отрезке, методом Крога [12], и разложении этой функции в ряд Тейлора в окрестностях заданной точки.

2. *Численный метод*

Коэффициенты полинома $p(x)$ степени N подбираются при помощи логистической регрессии таким образом, чтобы минимизировать среднюю квадратичную ошибку (mean squared error - MSE) в узлах интерполяции $\{x_i, f(x_i)\}_{i=1}^{m}$.

3. *Аппроксимация полиномами Чебышёва*

Аппроксимация функции $f(x)$ на отрезке $[a, b]$ строится в форме полинома следующего вида [13]:

$$f(x) \simeq \sum_{i=0}^{N} \alpha_i T_i \left(2\frac{x-a}{b-a} - 1 \right),$$

где $T_i(x) = cos(i \arccos(x)), x \in [-1, 1]$ – полиномы Чебышёва, α_i – соответствующие коэффициенты. Для T_i справедливы следующие равенства:

$$T_0(x) = 1, T_1(x) = x,$$

$$T_{i+1}(x) = 2xT_i(x) - T_{i-1}(x), i = 2, 3, \ldots$$

4. *Подход, основанный на аппроксимации производной*

Пусть $p_1(x)$ – полиномиальное приближение для производной $f'(x)$ на отрезке $[a, b]$, полученное одним из методов, описанных ранее (п. 1 – 3), $\deg p_1(x) = (N-1)$. Тогда аппроксимацией функции $f(x)$ на отрезке $[a, b]$ является полином $p(x)$ степени N [3]:

$$p = \int p_1(x)dx.$$

Будем рассматривать следующие способы выбора узлов интерполяции $\{x_i\}_{i=1}^m$:

– равномерное распределение на отрезке $x_i \sim U(a, b)$ (Рис. 2);

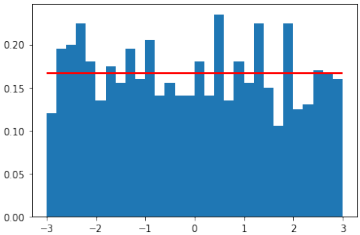

Рис. 2. $x_i \sim U(-3, 3)$

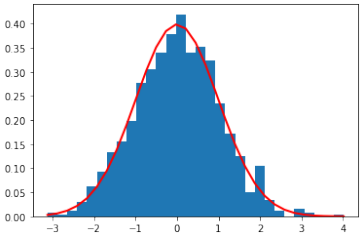

Рис. 3. $x_i \sim N(0, 1) P(x_i \in [-3, 3]) \approx 0,99$

– нормальное распределение $x_i \sim N(\mu, \sigma^2)$ (Рис. 3), где

$$\mu = [(a+b)/2], \sigma = \lceil (b-a)/6 \rceil \Rightarrow P(x_i \in [a, b]) \approx 0,9973;$$

– $x_i = (r_i + 1)\frac{b-a}{2} + a, \quad k = 1, \ldots, m,$

где $m > N$ – число узлов интерполяции (Рис. 4),
r_i – корни многочленов Чебышёва степени не выше N:

$$r_i = -cos\left(\frac{2i-1}{2m}\pi\right).$$

По графикам видно, что концентрация узлов интерполяции:

– примерно одинакова на всем отрезке при равномерном распределении,

– максимальна в середине отрезка и уменьшается к его концам при нормальном распределении,

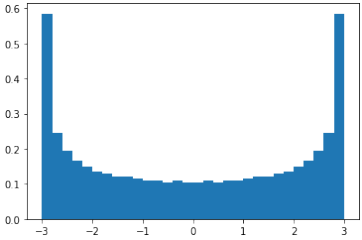

Рис. 4. x_i – корни многочленов Чебышёва, отображённые на отрезок $[-3, 3]$

– максимальна на краях отрезка и уменьшается к середине для корней многочленов Чебышёва.

Следовательно, можно ожидать, что при равномерном распределении точек полином будет стремиться приблизить исходную функцию одинаково хорошо на всем отрезке, при нормальном распределении – в середине отрезка, для корней многочленов Чебышёва – на концах отрезка.

Отметим, что для аппроксимации функции через разложение в ряд Тейлора узлы интерполяции не используются. Нужно знать значение функции и ее производных лишь в одной точке, в окрестностях которой выполняется аппроксимация.

3. Вводимые ограничения.

3.1. Способы оценки качества аппроксимации. Существуют различные способы оценки качества аппроксимации, такие как:

– средняя абсолютная ошибка

$$MAE(f,p) = \frac{1}{m}\sum_{i=1}^{m}|f(x_i) - p(x_i)|,$$

– средняя квадратичная ошибка

$$MSE(f,p) = \frac{1}{m}\sum_{i=1}^{m}(f(x_i) - p(x_i))^2,$$

– максимальное значение разности

$$max\_loss(f,p) = \max_{\substack{x_i,\\ i=1,\ldots,m}} |f(x_i) - p(x_i)|$$

и др.

Оценки MAE и MSE имеют схожий смысл, поэтому в данной работе будут рассматриваться MSE и max_loss, которые выявляют различные типы отклонений значений исследуемых функций.

В качестве единой метрики для сравнения точности аппроксимации будем использовать

$$\rho(f,p) = \sqrt{MSE(f,p)^2 + max\_loss(f,p)^2}.$$

Данная метрика может рассматриваться как аналог естественной евклидовой метрики в двумерном пространстве, где метрики MSE и max_loss определяют расстояние по соответствующим координатам.

3.2. Выбор отрезка для аппроксимации. Рассмотрим поведение функций аппроксимации (Рис. 1):

– *ReLU* – кусочно-полиномиальная на отрезках $(-\infty, 0], [0, \infty)$, принимает значения из отрезка $[0, \infty)$;

– *Sigmoid* – монотонно возрастает, точка перегиба в 0, $\sigma(x) \in (0,1), \sigma(x) \underset{x\to-\infty}{\longrightarrow} 0, \sigma(x) \underset{x\to+\infty}{\longrightarrow} 1$.

– *Tanh* – монотонно возрастает, точка перегиба в 0, $tanh(x) \in (-1,1), tanh(x) \underset{x\to-\infty}{\longrightarrow} -1, tanh(x) \underset{x\to+\infty}{\longrightarrow} 1$.

Исходя из вышесказанного, очевидно, что необходимо выбирать отрезок аппроксимации, симметричный относительно нуля, вида $[-l, l],\ l \neq 0$. В работе [3, 10] использовалось значение $l = 200$. Мы также используем отрезок аппроксимации $[-200, 200]$ в целях сравнения результатов.

Поскольку множества значений рассматриваемых функций активации различны, для корректного сравнения качества приближения, которое показывают рассматриваемые методы аппроксимации, нормализуем значения функций и полиномов с помощью проекции на отрезок $[0, 1]$:

$$y_{normalized} = \frac{y - y_{min}}{y_{max} - y_{min}} :$$

– $ReLU(x) :\ y_{min} = 0, y_{max} = 200$;
– $\sigma(x) :\ y_{min} = 0, y_{max} = 1$;
– $tanh(x) :\ y_{min} = -1, y_{max} = 1$.

После такой замены значения MSE и max_loss изменятся следующим образом:

$$MSE_{normalized} = \frac{MSE}{(y_{max} - y_{min})^2},$$

$$max\_loss_{normalized} = \frac{max\_loss}{(y_{max} - y_{min})}.$$

Далее в таблицах 1, 2, 3, 4, 5 нижние индексы $MSE_{normalized}$ и $max\_loss_{normalized}$ опущены для краткости.

3.3. Выбор степени полинома. Поскольку, как было отмечено во введении, существующие схемы гомоморфного шифрования ограничивают допустимое количество операций умножения, разработчикам необходимо по возможности ограничивать степень аппроксимирующего полинома. Вместе с тем, конкретные значения степеней будут зависеть от вида нейронной сети и параметров используемой схемы гомоморфного шифрования. В связи с этим далее будут рассматриваться полиномы степеней $2, 3, 4$ аналогично [3, 10] с тем, чтобы иметь возможность сравнения полученных результатов.

4. Полученные результаты. Далее будут использованы следующие обозначения:

– MSE – нормализованная среднеквадратичная ошибка приближения исходной функции полиномом,

– max_loss – нормализованное максимальное значение разности между полученным полиномом и исходной функцией.

Необходимо сделать некоторые замечания относительно построения приближений для функции $ReLU$.

Функция активации $ReLU$ недифференцируема в нуле, поэтому не имеет разложения в ряд Тейлора в окрестностях этой точки. Однако, поскольку для построения разложения в ряд Тейлора используется численный метод, это ограничение можно обойти.

Производной функции $ReLU$ является функция Хевисайда:

$$ReLU'(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geqslant 0 \end{cases} \quad .$$

Эта функция разрывна и недифференцируема в нуле. Поэтому авторами работы [3] было предложено вместо нее аппроксимировать функцию логистического сигмоида, которая имеет асимптоты $y = 0$ при $x \to -\infty$ и $y = 1$ при $x \to +\infty$, то есть приближает функцию Хевисайда.

4.1. Полученные полиномы. В таблицах 1, 2, 3 далее приведены полиномиальные аппроксимации функций $ReLU$, логистического сигмоида и гиперболического тангенса, соответственно. Для аппроксимации производной использовался метод аппроксимации многочленами Чебышёва, так как он проявил себя наилучшим образом среди рассмотренных методов.

Таблица 1. Полиномиальные аппроксимации для $ReLU$

Метод	**Полиномы**	**MSE**	**max_loss**	ρ
Численный метод. $x_i \sim U(-200, 200)$	$0.00236x^2+0.5006x+18.5$	0.0013	0.0925	0.0925
	$-2.834 \cdot 10^{-7}x^3 + 0.002358x^2 + 0.5074x + 18.47$	0.0013	0.0924	0.0924
	$-5.178 \cdot 10^{-8}x^4 - 2.154 \cdot 10^{-8}x^3 + 0.004124x^2 + 0.5021x + 11.49$	0.0003	0.0575	0.0575
Численный метод. $x_i \sim N(0, 67^2)$	$0.0030x^2+0.4915x+13.25$	0.0032	0.1734	0.1735
	$-2.011 \cdot 10^{-6}x^3 + 0.003051x^2 + 0.5175x + 13.07$	0.004	0.2385	0.2385
	$-5.114 \cdot 10^{-8}x^4 + 1.103 \cdot 10^{-6}x^3 + 0.004391x^2 + 0.4841x + 10.1$	0.0008	0.0614	0.0614
Численный метод. x_i – корни многочленов Чебышёва.	$0.002122x^2 + 0.5x + 21.22$	0.0015	0.1061	0.1061
	$-7.846 \cdot 10^{-21}x^3 + 0.002122x^2 + 0.5x + 21.22$	0.0015	0.1061	0.1061
	$-4.244 \cdot 10^{-8}x^4 + 6.431 \cdot 10^{-21}x^3 + 0.00382x^2 + 0.5x + 12.73$	0.0004	0.0637	0.0637
Ряд Тейлора	$0.001472x^2 + 0.5x + 40$	0.0063	0.2	0.2
	$-4.466 \cdot 10^{-6}x^3 + 0.003028x^2 + 0.634x + 8.965 \cdot 10^{-15}$	0.0063	0.1503	0.1504
	$-1.03 \cdot 10^{-8}x^4 + 3.6 \cdot 10^{-21}x^3 + 0.002091x^2 + 0.5x + 28.57$	0.0025	0.1429	0.1429
Полиномы Чебышёва	$0.002122x^2 + 0.5x + 21.22$	0.0015	0.1061	0.1061
	$-3.217 \cdot 10^{-21}x^3 + 0.002122x^2 + 0.5x + 21.22$	0.0015	0.1061	0.1061
	$-4.244 \cdot 10^{-8}x^4 - 9.181e \cdot 10^{-21}x^3 + 0.00382x^2 + 0.5x + 12.73$	0.0004	0.0637	0.0637
Аппроксимация производной (функция Хевисайда)	$0.001592x^2 + 0.5x + 36.34$	0.0047	0.1817	0.1818
	$1.017 \cdot 10^{-21}x^3 + 0.001592x^2 + 0.5x + 36.34$	0.0047	0.1817	0.1818
	$-2.653 \cdot 10^{-8}x^4 - 7.138 \cdot 10^{-21}x^3 + 0.003183x^2 + 0.5x + 15.12$	0.0006	0.0756	0.0756
Аппроксимация производной (логистический сигмоид)	$0.001591x^2 + 0.5x + 36.34$	0.0047	0.1817	0.1818
	$-3.311 \cdot 10^{-21}x^3 + 0.001591x^2 + 0.5x + 36.34$	0.0047	0.1817	0.1818
	$-2.652 \cdot 10^{-8}x^4 - 7.138 \cdot 10^{-21}x^3 + 0.003182x^2 + 0.5x + 15.13$	0.0006	0.0756	0.0756

Таблица 2. Полиномиальные аппроксимации для логистического сигмоида

Метод	**Полиномы**	**MSE**	**max_loss**	ρ
Численный метод. $x_i \sim U(-200, 200)$	$-5.352 \cdot 10^{-7}x^2 + 0.003761x + 0.5167$	0.0602	0.4917	0.4953
	$-1.378 \cdot 10^{-7}x^3 - 3.097 \cdot 10^{-7}x^2 + 0.007042x + 0.5127$	0.0328	0.4679	0.4691
	$5.441 \cdot 10^{-11}x^4 - 1.378 \cdot 10^{-7}x^3 - 2.166 \cdot 10^{-6}x^2 + 0.007041x + 0.5201$	0.0329	0.4754	0.4765
Численный метод. $x_i \sim N(0, 67^2)$	$-7.977 \cdot 10^{-7}x^2 + 0.006058x + 0.4916$	0.1333	0.7518	0.7636
	$-2.395 \cdot 10^{-7}x^3 + 1.031 \cdot 10^{-6}x^2 + 0.00917x + 0.487$	0.0512	0.6101	0.6122
	$2.507 \cdot 10^{-10}x^4 - 2.456 \cdot 10^{-7}x^3 - 5.33 \cdot 10^{-6}x^2 + 0.0093x + 0.5$	0.0565	0.8034	0.8053
Численный метод. x_i – корни многочленов Чебышёва.	$2.292 \cdot 10^{-20}x^2 + 0.003183x + 0.5$	0.0641	0.4784	0.4827
	$-1.061 \cdot 10^{-7}x^3 - 9.173 \cdot 10^{-21}x^2 + 0.006365x + 0.5$	0.0340	0.4594	0.4606
	$9.441 \cdot 10^{-25}x^4 - 1.061 \cdot 10^{-7}x^3 - 2.817 \cdot 10^{-20}x^2 + 0.006365x + 0.5$	0.0340	0.4594	0.4606
Ряд Тейлора	$-7.5 \cdot 10^{-6}x^2 + 0.004472x + 0.6$	0.0756	0.5944	0.5992
	$-2.437 \cdot 10^{-8}x^3 - 2.054 \cdot 10^{-19}x^2 + 0.003363x + 0.5$	0.0634	0.4774	0.4815
	$2.232 \cdot 10^{-10}x^4 - 9.958 \cdot 10^{-8}x^3 - 1.071 \cdot 10^{-5}x^2 + 0.006377x + 0.5714$	0.0374	0.5303	0.5316
Полиномы Чебышёва	$-9.932 \cdot 10^{-21}x^2 + 0.003183x + 0.5$	0.0641	0.4784	0.4827
	$-1.061 \cdot 10^{-7}x^3 - 2.141 \cdot 10^{-20}x^2 + 0.006365x + 0.5$	0.0340	0.4594	0.4606
	$-1.248 \cdot 10^{-25}x^4 - 1.061 \cdot 10^{-7}x^3 - 1.901 \cdot 10^{-20}x^2 + 0.006365x + 0.5$	0.0340	0.4594	0.4606
Аппроксимация производной	$-5.818 \cdot 10^{22}x^2 + 0.001592x + 0.3183$	0.1547	0.6697	0.6872
	$-5.305 \cdot 10^{-8}x^3 + 0.004774x + 0.5305$	0.0445	0.4994	0.5014
	$2.182 \cdot 10^{-26}x^4 - 5.305 \cdot 10^{-8}x^3 - 3.103 \cdot 10^{-21}x^2 + 0.004774x + 0.5305$	0.0445	0.4994	0.5014

В таблицах 4 и 5 мы приведем явный вид и оценки точности аппроксимации полиномов из проекта CryptoDL [3, 10].

Таблица 3. Полиномиальные аппроксимации для гиперболического тангенса

Метод	**Полиномы**	**MSE**	**max_loss**	ρ
Численный метод. $x_i \sim U(-200, 200)$	$7.093 \cdot 10^{-8}x^2 + 0.007516x - 0.01842$	0.0612	0.4938	0.4976
	$-2.775 \cdot 10^{-7}x^3 - 1.734 \cdot 10^{-7}x^2 + 0.01418x + 0.002386$	0.0338	0.4725	0.4737
	$7.749 \cdot 10^{-11}x^4 - 2.776 \cdot 10^{-7}x^3 - 2.809 \cdot 10^{-6}x^2 + 0.01418x + 0.01252$	0.0338	0.4776	0.4788
Численный метод. $x_i \sim N(0, 67^2)$	$-7.288 \cdot 10^{-7}x^2 + 0.01221x - 0.004985$	0.1369	0.7378	0.7504
	$-4.893 \cdot 10^{-7}x^3 + 5.002 \cdot 10^{-7}x^2 + 0.01845x - 0.01014$	0.0545	0.617	0.6194
	$5.371 \cdot 10^{-10}x^4 - 5.007 \cdot 10^{-7}x^3 - 1.295 \cdot 10^{-5}x^2 + 0.01861x + 0.01763$	0.0596	0.8218	0.824
Численный метод. x_i – корни многочленов Чебышёва.	$-5.393 \cdot 10^{-21}x^2 + 0.006366x$	0.0651	0.4869	0.4913
	$-2.122 \cdot 10^{-7}x^3 + 2.024 \cdot 10^{-21}x^2 + 0.01273x + 6.64 \cdot 10^{-17}$	0.0350	0.4742	0.4755
	$3.092 \cdot 10^{-24}x^4 - 2.122 \cdot 10^{-7}x^3 - 1.32 \cdot 10^{-19}x^2 + 0.01273x + 9.531 \cdot 10^{-16}$	0.0350	0.4742	0.4755
Ряд Тейлора	$-1.5 \cdot 10^{-5}x^2 + 0.008944x + 0.2$	0.0766	0.5944	0.5993
	$-4.874 \cdot 10^{-8}x^3 - 8.165 \cdot 10^{-19}x^2 + 0.006726x + 2.429 \cdot 10^{-14}$	0.0645	0.4862	0.4905
	$4.464 \cdot 10^{-10}x^4 - 1.992 \cdot 10^{-7}x^3 - 2.143 \cdot 10^{-5}x^2 + 0.01275x + 0.1429$	0.0384	0.5454	0.5467
Полиномы Чебышёва	$-1.835 \cdot 10^{-20}x^2 + 0.006366x + 3.309 \cdot 10^{-16}$	0.0651	0.4869	0.4912
	$-2.122 \cdot 10^{-7}x^3 - 1.835 \cdot 10^{-20}x^2 + 0.01273x + 3.309 \cdot 10^{-16}$	0.0350	0.4742	0.4755
	$1.048 \cdot 10^{-24}x^4 - 2.122 \cdot 10^{-7}x^3 - 4.442 \cdot 10^{-20}x^2 + 0.01273x + 8.387 \cdot 10^{-17}$	0.0350	0.4742	0.4755
Аппроксимация производной	$1.115 \cdot 10^{-21}x^2 + 0.003183x$	0.1227	0.4933	0.5083
	$-1.061 \cdot 10^{-7}x^3 + 0.009549x$	0.0446	0.4806	0.4826
	$5.091 \cdot 10^{-26}x^4 - 1.061 \cdot 10^{-7}x^3 - 6.206 \cdot 10^{-21}x^2 + 0.009549x$	0.0446	0.4806	0.4826

Таблица 4. Полиномиальные аппроксимации для ReLU из проекта [3, 10]

Метод	**Полиномы**	**MSE**	**max_loss**	ρ
Полиномы Лежандра	$0.0021x^2 + 0.5x + 21.22$	0.0015	0.1061	0.1061
	$-3.517 \cdot 10^{-16}x^3 + 0.0021x^2 + 0.5x + 21.22$	0.0015	0.1061	0.1061
	$-4.244 \cdot 10^{-8}x^4 + -3.517 \cdot 10^{-16}x^3 + 0.0038x^2 + 0.5x + 12.732$	0.0004	0.0637	0.0637
Численный метод	$0.0016x^2 + 0.5x + 38.557$	0.0057	0.1928	0.1929
	$-1.6093 \cdot 10^{-14}x^3 + 0.0016x^2 + 0.5x + 38.557$	0.0057	0.1928	0.1929
	$-1.73 \cdot 10^{-8}x^4 - 3.61 \cdot 10^{-14}x^3 + 0.0025x^2 + 0.5x + 27.453$	0.0021	0.1373	0.1373
Полиномы Чебышёва	$0.00234x^2 + 0.5x + 18.75$	0.0013	0.0938	0.0938
	$9.363 \cdot 10^{-20}x^3 + 0.00234x^2 + 0.5x + 18.75$	0.0013	0.0938	0.0938
	$-5.127 \cdot 10^{-8}x^4 + 9.363 \cdot 10^{-20}x^3 + 0.0041x^2 + 0.5x + 11.719$	0.0003	0.0586	0.0586
Аппроксимация производной (сигмоид)	$0.0013x^2 + 0.5x$	0.0106	0.2344	0.2346
	$-9.677 \cdot 10^{-21}x^3 + 0.002x^2 + 0.5x$	0.0043	0.2344	0.2346

В сводных таблицах 6 и 7 приведены наилучший метод аппроксимации и соответствующее значение метрики ρ для каждой рассмотренной степени аппроксимирующего полинома и каждой рассмотренной функции активации, полученные в данном исследовании и использованные в проекте [3, 10], соответственно. В целом, полученные результаты согласуются в смысле порядка ошибки аппроксимации. Вместе с тем, в настоящем исследования удалось получить несколько более точную аппроксимацию для функции $ReLU(x)$, чем в проекте [3, 10].

Полученные аппроксимации для логистического сигмоида существенно проигрывают в точности функции $ReLU$, однако оказались несколько лучше, чем аппроксимации для гиперболического тангенса.

Как было отмечено во введении, задача выбора аппроксимирующего полинома является по меньшей мере двухпараметрической (минимизация и ошибки аппроксимации, и степени полинома). Поэтому представляет интерес сравнение ошибок при разных степенях. Для каждой функции активации найдем полиномы степеней 2,3,4, дающие наилучшее приближе-

Таблица 5. Полиномиальные аппроксимации для гиперболического тангенса из проекта [3, 10]

Метод	**Полиномы**	**MSE**	**max_loss**	ρ
Полиномы Лежандра	$0.00637x$	0.0651	0.4869	0.4912
	$-2.1219 \cdot 10^{-7}x^3 + 0.0127x$	0.035	0.4742	0.4755
	$-2.1219 \cdot 10^{-7}x^3 + 0.0127x$	0.035	0.4742	0.4755
Численный метод	$0.00587x$	$0,0696$	0.4879	0.4928
	$-1.2015 \cdot 10^{-7}x^3 + 0.0095x$	0.0445	0.4806	0.4827
	$-1.2015 \cdot 10^{-7}x^3 + 0.0095x$	0.0445	0.4806	0.4827
Полиномы Чебышёва	$0.00745x$	$0,0611$	0.4847	0.4885
	$-2.734 \cdot 10^{-7}x^3 + 0.0141x$	0.0338	0.4716	0.4728
	$-2.734 \cdot 10^{-7}x^3 + 0.0141x$	0.0338	0.4716	0.4728

Таблица 6. Сводная таблица аппроксимаций функций активации

Функция	**Степень полинома**			**Лучший результат**
	2	3	4	
$ReLU$	Численный метод. $x_i \sim U(-200, 200)$, $\rho = 0.0925$	Численный метод. $x_i \sim U(-200, 200)$, $\rho = 0.0924$	Численный метод. $x_i \sim U(-200, 200)$, $\rho = 0.0575$	Численный метод. $x_i \sim U(-200, 200)$, $deg = 4$, $\rho = 0.0575$
$Sigmoid$	Полиномы Чебышёва, $\rho = 0.4827$	Полиномы Чебышёва, $\rho = 0.4606$	Полиномы Чебышёва, $\rho = 0.4606$	Полиномы Чебышёва, $deg = 3$, $\rho = 0.4606$
$Tanh$	Полиномы Чебышёва, $\rho = 0.4912$	Численный метод. $x_i \sim U(-200, 200)$ $\rho = 0.4737$	Полиномы Чебышёва, $\rho = 0.4755$	Численный метод. $x_i \sim U(-200, 200)$, $deg = 3$, $\rho = 0.4737$

ние. Для каждого из них рассчитаем ошибку аппроксимации и вычислим модули их попарных разностей. Полученные результаты представлены в таблице 8.

Можно видеть, что для:

– функции $ReLU$ использование полинома третьей степени практически не дает выигрыша по сравнению с использованием полинома

Таблица 7. Сводная таблица аппроксимаций функций активации из проекта [3, 10]

Функция	**Степень полинома**			**Лучший результат**
	2	3	4	
$ReLU$	Полиномы Чебышёва, $\rho = 0.0938$	Полиномы Чебышёва, $\rho = 0.0938$	Полиномы Чебышёва, $\rho = 0.0586$	Полиномы Чебышёва, $deg = 4$, $\rho = 0.0586$
$Tanh$	Полиномы Чебышёва, $\rho = 0.4885$	Полиномы Чебышёва, $\rho = 0.4728$	Полиномы Чебышёва, $\rho = 0.4728$	Полиномы Чебышёва, $deg = 3$, $\rho = 0,4728$

Таблица 8. Таблица модуля разности ошибки аппроскимации для наилучших аппроксимирующих многочленов различных степеней

Функция	**Степени полиномов**		
	$2-3$	$2-4$	$3-4$
$ReLU$	0.001	0.035	0.034
$Sigmoid$	0.0221	0.0221	0
$Tanh$	0.0239	0.0188	0.0051

второй степени, хотя они уступают по точности полиному четвертой степени;

– логистического сигмоида, отсутствует разница между ошибками аппроксимаций 3 и 4 степеней;

– гиперболического тангенса наилучшим выбором является использование полинома третьей степени.

Эти результаты целесообразно использовать для оптимизации реализаций гомоморфных вычислений.

5. Оценка ошибки нейронной сети. Представляет интерес оценка влияния использования полученных аппроксимаций функций активации на работу нейронной сети в целом.

Одним из сценариев использования гомоморфного шифрования является защита уже обученной модели, например, от извлечения данных об обучающей выборке. В этом сценарии сначала обучается стандартная модель, после этого происходит замена нелинейных функций полиномами. Затем к нейронной сети применяется гомоморфное шифрование, после чего она может быть использована для безопасной обработки данных.

В этой связи представляет интерес оценка ошибки результата вычисления выходного значения зашифрованной нейронной сети относительно исходной сети, вызванной заменой функций активации приближениями. В работах [2, 3, 10] такая задача в явном виде не рассматриваются – авторы экспериментально оценивают точность классификации сети на стандартных наборах данных.

В работе [11] авторы предлагают общий подход, основанный на представлении нейронной сети в виде композиции элементарных преобразований, для каждого их которых вычисляется оценка ошибки аппроксимации. Показано, что модуль разности значения выхода исходной сети и ее аппроксимации может быть ограничен величиной $Const \cdot 2^{-\alpha}$, где α – параметр, а $Const$ – некоторая константа. Возникает задача оценки значения этой константы.

Так в работе [14] отмечается, что подход, использованный в [11] и основанный на вычислении композиции преобразований, может приводить к завышению оценки. Это следует из того факта, что если функция использует какую-либо переменную несколько раз, а такое справедливо для нейронных сетей с большим числом слоев, ошибки, вносимые данной переменной, потенциально могут компенсироваться. Для более точной оценки может быть использован описанный в той же монографии аналитический способ, однако он требует вычисления частных производных от функции, описывающей нейронную сеть целиком, что в общем случае может являться трудоемкой задачей. Вместо этого для конкретной нейронной сети мы можем оценить такую ошибку экспериментально, вычислив оценку величины $max\_loss(F, F^*)$. Здесь F, F^* – функции, описывающие исходную и приближенную нейронные сети соответственно, а максимум берется по множеству входных аргументов.

Рассмотрим простейшую сверточную нейронную сеть для распознавания изображений цифр от 0 до 9 из базы данных MNIST [15] и приближенный вариант такой сети, для которых вычислим указанное значение max_loss. Каждое изображение является монохромным и представляет собой квадрат со стороной 28 пикселей. Будем считать, что каждый пиксель принимает значение от -200 до 200, что позволяет пользоваться полученными выше оценками.

Нейронная сеть реализуется последовательным применением:

1. Сверточного слоя с ядром свертки 5×5 пикселей, преобразующего матрицу $A = (a_{ij})_{28\times 28}$ в матрицу $B = (b_{ij})_{24\times 24}$ с последующим масштабированием значений элементов в указанный выше интервал ($b_{ij} \in [-200, 200]$).

2. Слоя функции активации $ReLU$, которая применяется к каждому элементу полученной матрицы.

3. Слоя субдескритизации, дающего на выходе матрицу $C = (c_{ij})_{12\times 12}$, $c_{ij} = 1/4(ReLU(b_{2i,2j}) + ReLU(b_{2i+1,2j}) + ReLU(b_{2i,2j+1}) + ReLU(b_{2i+1,2j+1}))$, $c_{ij} \in [-200, 200]$.

4. Слоя регуляризации, исключающего случайным образом 20% нейронов, в результате чего в общей сложности остается 116 элементов,

подлежащих дальнейшей обработке. Без ограничения общности будем считать, что выбираются первые 116 элементов: $D = (d_1, ..., d_{116})$, $d_{ij} \in [-200, 200]$.

5. Полносвязного слоя из 116 нейронов с функцией активации $ReLU$, $E = (e_1, ..., e_{116}).e_i = ReLU(\sum_{j=1}^{116} \gamma_{ij} d_j)$, $e_i \in [0, 200]$.

6. Финального полносвязного слоя из 10 нейронов с многомерной логистической функцией ($Softmax$): $F = (f_1, ..., f_{10})$, $f_i = \frac{e^{z_i}}{\sum_{j=1}^{10} e^{z_j}} \in [0, 1]$, где $z_i = \sum_{j=1}^{116} \kappa_{ij} e_j$, при этом значения z_i нормируются: $z_i \in [-1, 1]$.

Эксперимент по оценке возникающей при аппроксимации функций $ReLU$ ошибки рассмотренной нейронной сети в целом заключался в следующем:

1. Случайным образом вырабатывались весовые коэффициенты нейронов.

2. Случайным образом вырабатывалось 10^6 входных матриц A.

3. Для каждой из матриц вычислялось значение $|f_1(A) - f_1(A)^*|$, где f_1 и f_1^* – значение первого выходного нейрона для исходной нейронной сети и ее аппроксимации на одной и той же входной матрице A.

4. Вычислялось значение функции $max\_loss(f_1, f_1^*) = \max_A |f_1(A) - f_1(A)^*|$.

Аппроксимация проводилась с помощью полинома четвертой степени, дающего наилучшее приближение (см. Таблицу 1). Функция max_loss не аппроксимировалась, поскольку она не содержит информации о параметрах нейронной сети.

Проведенный эксперимент показал, что $max\_loss(f_1, f_1^*) = 0.200293$.

Таким образом, даже для рассмотренной сравнительно простой сети использование аппроксимаций функций активации может приводить к достаточно большим искажениям выходных значений.

6. Заключение. Полученные результаты не позволили в точности получить приближения, опубликованные в широко цитируемой работе [3], или используемые в доступной в сети Интернет программной реализации [10], подготовленной на ее основе. Однако точность получаемых аппроксимаций в целом согласуется с опубликованными ранее результатами. Для функции $ReLU(x)$ удалось получить представление аппроксимирующего полинома, обладающего большей точностью, чем в рассмотренных источниках, в смысле использованной метрики точности приближения ρ. Получен вид аппроксимирующих полиномов для функции логистический сигмоид.

Полученные результаты показывают, что наилучшее полиномиальное приближение из трех рассмотренных функций активации можно построить для функции $ReLU(x)$ с помощью численного метода, что обусловлено ее поведением на числовой прямой, а именно неограниченным возрастанием. Точность полученных с помощью полиномов Чебышева наилучших полиномиальных аппроксимаций функций $\sigma(x)$ и $tanh(x)$ ниже из-за наличия асимптот при $x \longrightarrow \pm\infty$. Таким образом, для снижения ошибок классификации при защите нейронных сетей с помощью алгоритмов гомоморфного шифрования целесообразно использовать сети, состоящие из нейронов с функцией активации $ReLU(x)$.

Для простейшей сверточной нейронной сети, включающей два слоя нелинейных преобразований с функцией активации $ReLU(x)$, приведена экспериментальная оценка отклонения в значении одного выходного нейрона ее приближенного варианта относительно исходного.

В отличие от опубликованных ранее работ нами исследованы различные подходы к построению аппроксимаций функций активации. Результаты работы (полученные приближения или методы построения наилучших приближений для рассмотренных функций) целесообразно использовать при реализации методов конфиденциального машинного обучения.

Литература

1. Pitropakis N., Panaousis E., Giannetsos T., Anastasiadis E., Loukas G. A taxonomy and survey of attacks against machine learning // Comput. Sci. Rev. 2019. Vol. 34. pp. 100 – 199.
2. Dowlin N., Gilad-Bachrach R., Laine K., Lauter K., Naehrig M., Wernsing J. CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy // Proceedings of the 33rd International Conference on Machine Learning (ICML). 2016. pp. 201 – 210.
3. Hesamifard, E., Takabi, H., Ghasemi, M. CryptoDL: Deep neural networks over encrypted data // arXiv preprint:1711.05189. 2017.
4. Juvekar C., Vaikuntanathan V., and Chandrakasan A. GAZELLE: A low latency framework for secure neural network inference // 27th USENIX Security Symposium. USENIX Association. 2018. pp. 1651—1669.
5. Brakerski Z., Gentry C., Vaikuntanathan V. (Leveled) Fully Homomorphic Encryption Without Bootstrapping // ACM Trans. Comput. Theory. vol. 6. 2014. pp. 13:1–13:36.
6. Cheon J.H., Kim A., Kim M., Song Y. Homomorphic encryption for arithmetic of approximate numbers // Proceedings of the International Conference on the Theory and Applications of Cryptology and Information Security. 2017. pp. 409 – 437.
7. Crawford L. H., Gentry C., Halevi S., Platt D., and Shoup V. Doing realwork with FHE: The case of logistic regression // 6th Workshop Encrypted Comput. Appl. Homomorphic Cryptogr. (WAHC), New York, USA. 2018. pp. 1-–12.
8. Ghodsi Z., Gu T., Garg S. Safetynets: Verifiable execution of deep neural networks on an untrusted cloud // Advances in Neural Information Processing Systems. 2017. pp. 4672—4681.

9. Ramy E.A., Jinhyun S., Salman Avestimehr. On Polynomial Approximations for Privacy-Preserving and Verifiable ReLU Networks // arXiv preprint:2011.05530, 2020.
10. Репозиторий проекта CryptoDL URL: https://github.com/inspire-lab/CryptoDL (дата обращения: 01.09.2021).
11. Lee J., Lee E., Lee J.-W., Kim Y., Kim Y.-S., No J.-S. Precise Approximation of Convolutional NeuralNetworks for Homomorphically Encrypted Data // arXiv preprint:2105.10879, 2021.
12. Krogh F.T. Efficient Algorithms for Polynomial Interpolation and Numerical Differentiation, Math. Comput. 1970. vol. 24. no. 109. pp. 185 – 190.
13. Молчанов И.Н. Машинные методы решения прикладных задач. Алгебра, приближение функций. Киев: Наук. думка, 1987. С. 288.
14. Taylor J.R. An introduction to error analysis, University Science Books Mill Valley, California, 1982. P. 344.
15. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition // Proceedings of the IEEE. 1998. vol. 86, no. 11. pp. 2278–2324.

Маршалко Григорий Борисович — эксперт, Технический комитет по стандартизации "Криптографическая защита информации". Область научных интересов: защита информации, криптография, биометрическая идентификация. Число научных публикаций – 32. marshalko_gb@tc26.ru; ул. Отрадная, 2В-1, 127273, Москва, Россия; р.т.: +7(495)737-6192.

Труфанова Юлия Анатольевна — эксперт, Технический комитет по стандартизации "Криптографическая защита информации". Область научных интересов: защита информации, биометрическая идентификация, машинное обучение. Число научных публикаций – 1. trufanova_ua@tc26.ru; ул. Отрадная, 2В-1, 127273, Москва, Россия; р.т.: +7(495)737-6192.

DOI 10.15622/ia.2022.21.6

G.B. MARSHALKO, J.A. TRUFANOVA

POLYNOMIAL APPROXIMATIONS FOR SEVERAL NEURAL NETWORK ACTIVATION FUNCTIONS

Marshalko G.B., Trufanova J.A. **Polynomial Approximations for Several Neural Network Activation Functions.**

Abstract. Active deployment of machine learning systems sets a task of their protection against different types of attacks that threaten confidentiality, integrity and accessibility of both processed data and trained models. One of the promising ways for such protection is the development of privacy-preserving machine learning systems, that use homomorphic encryption schemes to protect data and models. However, such schemes can only process polynomial functions, which means that we need to construct polynomial approximations for nonlinear functions used in neural models. The goal of this paper is the construction of precise approximations of several widely used neural network activation functions while limiting the degree of approximation polynomials as well as the evaluation of the impact of the approximation precision on the resulting value of the whole neural network. In contrast to the previous publications, in the current paper we study and compare different ways for polynomial approximation construction, introduce precision metrics, present exact formulas for approximation polynomials as well as exact values of corresponding precisions. We compare our results with the previously published ones. Finally, for a simple convolutional network we experimentally evaluate the impact of the approximation precision on the bias of the output neuron values of the network from the original ones. Our results show that the best approximation for ReLU could be obtained with the numeric method, and for the sigmoid and hyperbolic tangent – with Chebyshev polynomials. At the same time, the best approximation among the three functions could be obtained for ReLU. The results could be used for the construction of polynomial approximations of activation functions in privacy-preserving machine learning systems.

Keywords: activation function, ReLU, tanh, sigmoid, homomorphic encryption, BGV, CKKS, neural network, polynomial approximation, privacy-preserving machine learning.

Marshalko Grigory — Expert, Technical committee for standardization "Cryptography and security mechanisms". Research interests: information security, cryptography, biometric identification. The number of publications – 32. marshalko_gb@tc26.ru; 2B-1, Otradnaya St., 127273, Moscow, Russia; office phone: +7(495)737-6192.

Trufanova Julia — Expert, Technical committee for standardization "Cryptography and security mechanisms". Research interests: information security, biometric identification, machine learning. The number of publications — 1. trufanova_ua@tc26.ru; 2B-1, Otradnaya St., 127273, Moscow, Russia; office phone: +7(495)737-6192.

References

1. Pitropakis N., Panaousis E., Giannetsos T., Anastasiadis E., Loukas G. A taxonomy and survey of attacks against machine learning // Comput. Sci. Rev. 2019. Vol. 34. pp. 100 – 199.
2. Dowlin N., Gilad-Bachrach R., Laine K., Lauter K., Naehrig M., Wernsing J. CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy // Proceedings of the 33rd International Conference on Machine Learning (ICML). 2016. pp. 201 – 210.

3. Hesamifard, E., Takabi, H., Ghasemi, M. CryptoDL: Deep neural networks over encrypted data // arXiv preprint:1711.05189. 2017.
4. Juvekar C., Vaikuntanathan V., and Chandrakasan A. GAZELLE: A low latency framework for secure neural network inference // 27th USENIX Security Symposium. USENIX Association. 2018. pp. 1651—1669.
5. Brakerski Z., Gentry C., Vaikuntanathan V. (Leveled) Fully Homomorphic Encryption Without Bootstrapping // ACM Trans. Comput. Theory. vol. 6. 2014. pp. 13:1–13:36.
6. Cheon J.H., Kim A., Kim M., Song Y. Homomorphic encryption for arithmetic of approximate numbers // Proceedings of the International Conference on the Theory and Applications of Cryptology and Information Security. 2017. pp. 409 – 437.
7. Crawford L. H., Gentry C., Halevi S., Platt D., and Shoup V. Doing realwork with FHE: The case of logistic regression // 6th Workshop Encrypted Comput. Appl. Homomorphic Cryptogr. (WAHC), New York, USA. 2018. pp. 1-–12.
8. Ghodsi Z., Gu T., Garg S. Safetynets: Verifiable execution of deep neural networks on an untrusted cloud // Advances in Neural Information Processing Systems. 2017. pp. 4672—4681.
9. Ramy E.A., Jinhyun S., Salman Avestimehr. On Polynomial Approximations for Privacy-Preserving and Verifiable ReLU Networks // arXiv preprint:2011.05530, 2020.
10. CryptoDL repository URL: https://github.com/inspire-lab/CryptoDL (accessed: 01.09.2021).
11. Lee J., Lee E., Lee J.-W., Kim Y., Kim Y.-S., No J.-S. Precise Approximation of Convolutional NeuralNetworks for Homomorphically Encrypted Data // arXiv preprint:2105.10879, 2021.
12. Krogh F.T. Efficient Algorithms for Polynomial Interpolation and Numerical Differentiation, Math. Comput. 1970. vol. 24. no. 109. pp. 185 – 190.
13. Molchanov I.N. Mashinnie metody reshenia prikladnikh zadach. Algebra, priblizhenie funkcij [Machine methods of solving applied tasks. Algebra, function approximation]. Kiev: Nauk. dumka. 1987. P. 288. (in Russ.)
14. Taylor J.R. An introduction to error analysis, University Science Books Mill Valley, California, 1982. P. 344.
15. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition // Proceedings of the IEEE. 1998. vol. 86, no. 11. pp. 2278–2324.