

СТАТИСТИЧЕСКИЕ СВОЙСТВА АППРОКСИМАЦИИ ДВОИЧНЫХ ФУНКЦИЙ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

К. В. ЛАТУШКИН, Ю. С. ХАРИН

*НИИ прикладных проблем математики и информатики,
Белорусского государственного университета,
г. Минск, Республика Беларусь*

Введение

В последние годы Искусственные нейронные сети (ИНС) начинают широко использовать в задачах криптологии и кибербезопасности [1–3]. Примерами таких задач являются: аппроксимация дискретных функций в программных датчиках псевдослучайных последовательностей, оценка s -блоков и других криптографических примитивов; распознавание компьютерных атак на информационные системы. Математически эти задачи сводятся к задаче аппроксимации двоичных функций от многих двоичных переменных. Исследованию особенностей этой актуальной задачи посвящена данная публикация.

Математическая модель и постановка задачи

Введем обозначения: $V = \{0, 1\}$ – двоичный алфавит; s – натуральное число; V^s – двоичный гиперкуб; $x = (x_1, \dots, x_s)' \in V^s$ – двоичный вектор-столбец; $I\{B\} \in V$ – индикатор события (Ω, F, P) . На вероятностном пространстве определена случайная двоичная функция

$$y = f(x) = f(x_1, \dots, x_s), x \in V^s, y \in V, \quad (1)$$

задающую классификацию в два класса: $\Omega_0 = \{y = 0\}$, $\Omega_1 = \{y = 1\}$ которую можно интерпретировать как раскраску каждой вершины в один из двух цветов: $y = 0$ или $y = 1$. Функция выбирается равномерно из множества всех $M = 2^{2^s}$ двоичных функций.

Рассматривается задача статистического оценивания неизвестной двоичной функции (1) по случайной выборке объема n из V^s : $X = \{x^{(1)}, \dots, x^{(n)}\} \subseteq V^s$.

Для построения статистической оценки $\hat{f}(x)$ неизвестной функции (1) по выборке X используется двухслойная искусственная нейронная сеть, математически представляемая в виде:

$$\hat{f}(x_1, \dots, x_s) = \sigma \left(b_0 + \sum_{j=1}^m b_j * ReLU \left(a_{0j} + \sum_{i=1}^s a_{ij} x_i \right) \right), \quad (2)$$

где $\{b_i\}$, $\{a_{ij}\}$ – коэффициенты (веса) модели; $ReLU(z) = \max\{0, z\}$; $\sigma(z) = 1/(1 + e^{-z})$.

На рис. 1 представлена графическая схема ИНС, соответствующая (2); здесь $\{H_i^{(1)}\}$ – набор из m нейронов 1-ого (скрытого) слоя, $H^{(2)}$ – нейрон 2-го слоя.

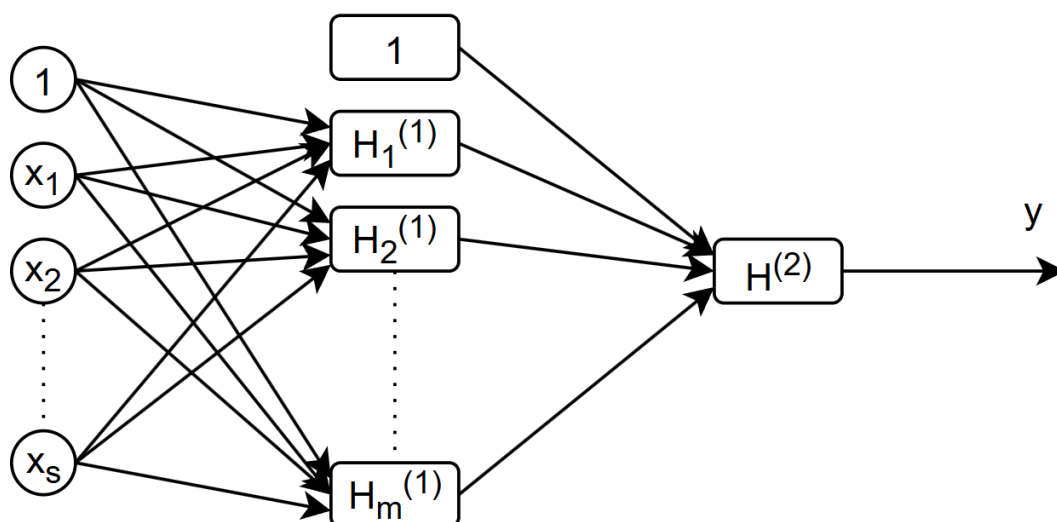


Рисунок 1 – Графическая схема искусственной нейронной сети (ИНС)

Для обучения ИНС (2) в качестве функции потерь выбрана бинарная перекрестная энтропия [4]:

$$H(\hat{y}) = -\frac{1}{n} \sum_{t=1}^n (y^{(t)} \log(\hat{y}^{(t)}) + (1 - y^{(t)}) \log(1 - \hat{y}^{(t)})), \quad (3)$$

где $\hat{y}^{(t)} = \hat{f}(x^{(t)})$ – полученная в процессе обучения оценка $y^{(t)}$, а для оценивания точности обучения метрика *accuracy* – доля правильно классифицированных вершин:

$$accuracy = \frac{1}{n} \sum_{t=1}^n I\{\hat{y}^{(t)} = y^{(t)}\}. \quad (4)$$

Задача заключается в оценивании зависимости числа m нейронов скрытого слоя, достаточных для аппроксимации на заданном уровне точности α , от числа s – количества переменных случайно заданной функции f .

Описание компьютерных экспериментов

Проведены три серии экспериментов.

Серия 1. Выбирается некоторая функция $f(x_1, \dots, x_s)$, которая задает раскраску гиперкуба случайным образом:

$$P\{y = 0\} = P\{y = 1\} = \frac{1}{2}.$$

По случайной выборке X объема $n = 2^s$ строится оценка $\hat{f}(x_1, \dots, x_s)$ ИНС с m нейронами и находится наименьшее $\hat{m}$ при котором достигается частота $\alpha = 0.95$ совпадения значений f и $\hat{f}$ за итераций (эпох) алгоритмом обучения ИНС, реализованным в библиотеке tensorflow.

Для нахождения $\hat{m}$ алгоритмом бинарного поиска перебираются значения из отрезка $[1, n]$ и на каждом выбранном m запускается процесс обучения. Эксперимент повторяется K раз для каждого s и формируется s выборка $M^{(s)} = \{m_1^{(s)}, m_2^{(s)}, \dots, m_K^{(s)}\}$.

Серия 2. В этой серии экспериментов исследуется способность ИНС безошибочно ($\alpha = 1$) аппроксимировать случайную функцию, количество итераций (эпох) обучения увеличивается до .

Серия 3. В этой серии экспериментов исследуются функции «наиболее трудные» (по количеству нейронов $\hat{m}$) для аппроксимации ИНС. Число различных двоичных функций от бинарных переменных равно $M = 2^{2^s}$, что $s = 3$ для составляет $M = 256$. На всех возможных функциях вычисляются и формируется выборка $M^{(s)} = \{m_1, m_2, \dots, m_{2^s}\}$. Для уменьшения влияния начальных значений параметров ИНС на результат обучения $\hat{m}$ вычисляется для каждой функции L раз и выбирается наименьшее.

Отметим, что во всех этих экспериментах для обучения модели ИНС используется адаптивный метод градиентного спуска Adam с параметрами $\beta_1 = 0.9, \beta_2 = 0.999, l_{rate} = 0.001$. Начальные коэффициенты задаются согласно [5, 6] из распределений:

$$a_{ij} \sim N\left(0, \sqrt{\frac{2}{s}}\right), b_i \sim U\left[-\frac{\sqrt{6}}{\sqrt{m+1}}, \frac{\sqrt{6}}{\sqrt{m+1}}\right].$$

Регрессионная оценка зависимости числа нейронов

По результатам серий экспериментов № 1, № 2 для каждого по выборке $M^{(s)}$ построены оценки среднего $\hat{\mu}^{(s)}$ и среднеквадратичного отклонения $\hat{\sigma}^{(s)}$ величины $\hat{m}$. На рисунках 2, 3 представлены столбчатые диаграммы, по горизонтальной оси нанесены значения параметра s , а по вертикальной – оценки $\hat{\mu}^{(s)}$ и $\hat{\sigma}^{(s)}$ в логарифмической шкале (по основанию 10).

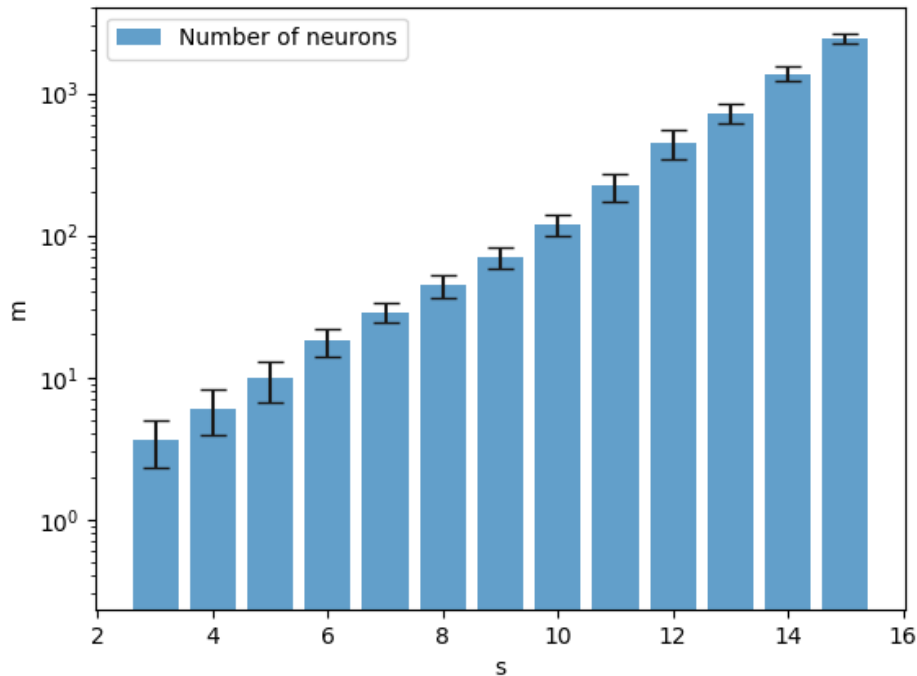


Рисунок 2 – Столбчатая диаграмма оценок среднего и среднеквадратичного отклонения величины $\hat{m}$ в серии экспериментов № 1 для соответствующих s

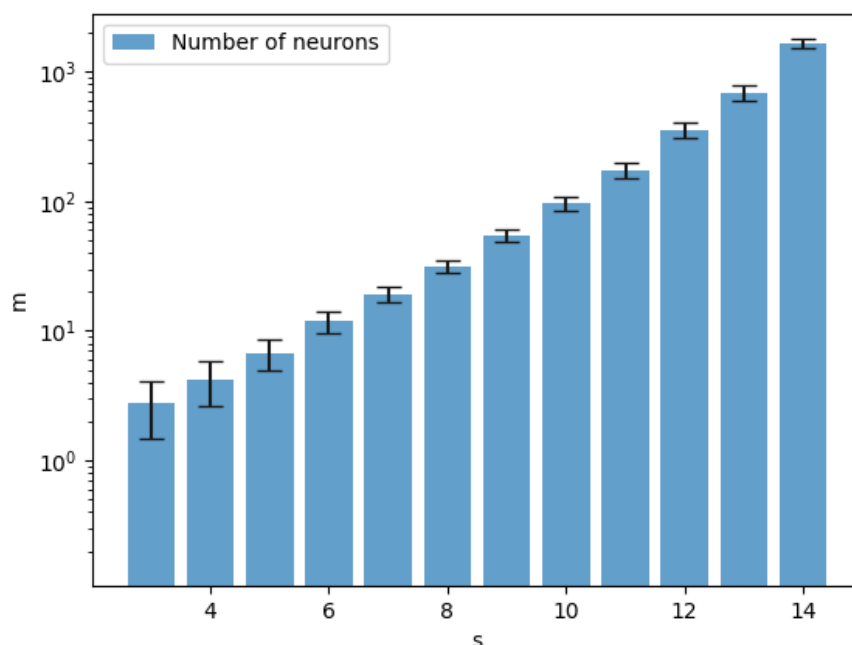


Рисунок 3 – Столбчатая диаграмма оценок среднего и среднеквадратичного отклонения величины $\hat{m}$ в серии экспериментов № 2 для соответствующих s

По результатам проведенных экспериментов зависимость среднего числа нейронов $\mu = \mu(s)$ при аппроксимации двоичных функций (1) с помощью ИНС (2)-(4) допускает следующую нелинейную регрессионную зависимость:

$$\mu(s) \approx (as^4 + bs^3 + cs^2 + ds + e)f^s,$$

где коэффициенты a, b, c, d, e, f найдены при решении задачи минимизации:

$$G(a, b, c, d, e, f) = \frac{1}{s_{max} - s_{min} + 1} \sum_{s=s_{min}}^{s_{max}} \left(\frac{\mu(s) - \hat{\mu}^{(s)}}{\hat{\sigma}^{(s)}} \right)^2 \rightarrow \min_{a, b, c, d, e, f}.$$

Результаты решения задачи минимизации представлены в таблице 1.

Таблица 1 – Результаты решения задачи минимизации

	Серия № 1	Серия № 2
a	0.0194066	0.0008
b	-0.451091	-0.0219983
c	3.97759	0.225223
d	-13.9166	-0.9509
e	18.7154	2.11477
f	1.19617	1.55643
G_{min}	0.0222633	0.0375392

О «наиболее трудных» для аппроксимации функций.

На рисунке 4 представлена гистограмма выборки $M^{(s)}$, полученной из серии экспериментов № 3, для $s = 3$ при переборе всех $M = 2^{2^s}$ двоичных функций.

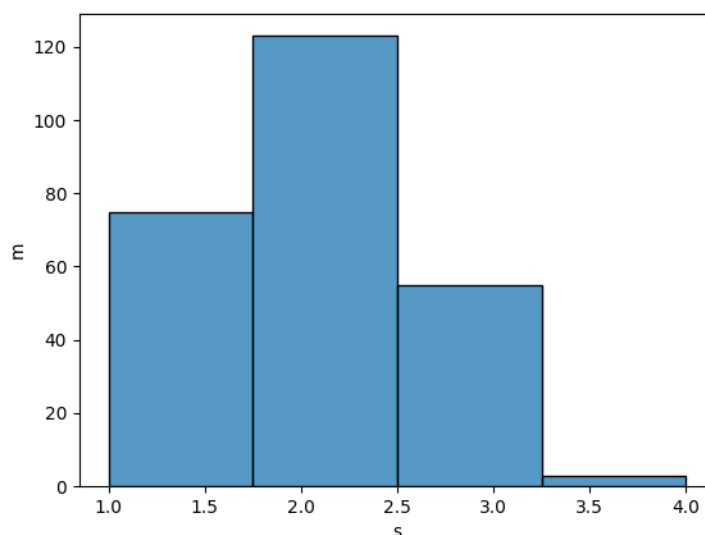


Рисунок 4 – Гистограмма для выборки

В таблице (2) представлены 10 «наиболее трудных» для аппроксимации функций при $s = 3$, где $y_0 = f(0, \dots, 0)$, $y_1 = f(0, \dots, 1)$, $y_{2^s-1} = f(1, \dots, 1)$.

Таблица 2 – «Наиболее трудные» для аппроксимации функции при $s = 3$

Число нейронов t	Значения функции f $Y=(y_0, y_1, \dots, y_{(2^s-1)})$
3	$Y=(1, 1, 0, 1, 0, 1, 1, 0)$
3	$Y=(1, 1, 0, 1, 1, 0, 0, 0)$
3	$Y=(1, 1, 0, 1, 1, 0, 1, 1)$
3	$Y=(1, 1, 1, 0, 0, 0, 1, 1)$
3	$Y=(1, 1, 1, 0, 0, 1, 1, 1)$
3	$Y=(1, 1, 1, 0, 0, 1, 1, 0)$
3	$Y=(1, 1, 1, 0, 1, 0, 1, 1)$
4	$Y=(0, 1, 0, 0, 0, 0, 1, 0)$
4	$Y=(0, 1, 1, 1, 1, 0, 0, 1)$
4	$Y=(1, 0, 0, 1, 0, 1, 1, 0)$

Литература

1. Gohr A. Improving attacks on round-reduced speck32/64 using deep learning // Advances in cryptology, CRYPTO-2019. – 2019. – P.150–179.
2. Deep learning-based physical side-channel analysis / S. Picek, G. Perin, L. Mariot, L. Wu, L. Batina // ACM Computing Surveys, Vol. 55(11). – 2023. P.1–35.

3. Boanca, S. Exploring patterns and assessing the security of pseudorandom number generators with machine learning / S. Boanca // International Conference on Agents and Artificial Intelligence. – 2024. – Vol.3. – P.186–193.
4. Николенко, С. Глубокое обучение. Погружение в мир нейронных сетей / С. Николенко, А. Кадури, Е. Архангельская. – СПб.: Питер, 2018. – 480 с.
5. Glorot X., Bengio Y. Understanding the Difficulty of Training Deep Feedforward Neural Networks // International conference on artificial intelligence and statistics, 2010. – P. 249–256.
6. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification / K. He et al. // Proc. ICCV 2015, 2015. – P. 1026–1034.