

ПРИМЕНЕНИЕ АППАРАТА ТРОПИЧЕСКОЙ МАТЕМАТИКИ В АРХИТЕКТУРЕ НЕЙРОННЫХ СЕТЕЙ

Петрова А.¹, аспирантка, ✉ anna.petrova.16.11@gmail.com
Казакевич В. Г.¹, старший преподаватель, ✉ sokratt@gmail.com,
orcid.org/0000-0002-3150-2105

¹ Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»
им. В. И. Ульянова (Ленина), ул. Профессора Попова, 5, корп. 3, 197022, Санкт-Петербург, Россия

Аннотация

Используя методы тропической математики, можно упростить структуру нейронной сети. Целью настоящей работы является исследование эффективности использования тропических функций в нейронных сетях. Базовая алгебраическая структура в тропической математике — полукольцо с выбором минимума (или максимума) в качестве сложения. Такая структура естественно возникает в работе с некоторыми нейронными сетями, например сетями с функцией активации ReLU. В рамках этого подхода произвольная кусочно-линейная функция с фиксированным типом выпуклости может быть представлена тропическим многочленом, а произвольная кусочно-линейная функция — тропической рациональной функцией. Таким образом, слой нейронной сети с линейной преактивацией и ReLU-активацией можно рассматривать как векторнозначную тропическую рациональную функцию, которая, в свою очередь, может быть представлена двумя тропическими слоями. В работе были реализованы два тропических слоя и построены пять тропических архитектур. Обучение моделей проводилось на датасете, целью которого было оценить вероятность наличия сердечно-сосудистых заболеваний у пациентов по определенному набору характеристик. Все модели имели одинаковые гиперпараметры. Каждая из моделей обучалась в течение 100 эпох с использованием оптимизаторов Adam и SGD. Сравнение результатов показало, что наилучшей точности достигла модель со смешанной архитектурой, состоящая из двух линейных слоев с min-слоем и max-слоем между ними. Такая точность была достигнута при использовании оптимизатора Adam. Классическая модель набрала 77,3%, а тропическая 77,7%.

Ключевые слова: тропическая математика, нейронные сети.

Цитирование: Петрова А., Казакевич В. Г. Применение аппарата тропической математики в архитектуре нейронных сетей // Компьютерные инструменты в образовании. 2023. № 3. С. 18-27. doi: 10.32603/2071-2340-2023-3-18-27

1. ВВЕДЕНИЕ

В последние годы идея использования аппарата тропической математики в архитектуре нейронных сетей набирает популярность. Теоретические оценки показывают, что использование тропических слоев может привести к повышению точности и скорости работы нейронных сетей [1].

Тропическая математика — это раздел математики, изучающий свойства функций, заданных на множестве, в котором определены операции максимума (минимума) и сложения. Она находит широкое применение в различных областях, таких как теория оптимизации, теория графов, теория динамических систем, теория управления и т. д. Естественно, возникает идея использования тропических функций для описания нейронных сетей, в которых встречаются кусочно-линейные функции активации [2]. Целью настоящей работы является исследование эффективности использования тропических функций в нейронных сетях.

2. ОПРЕДЕЛЕНИЯ

Пусть $\mathbb{R}_{\min} = \mathbb{R} \cup \{\infty\}$ и $\mathbb{R}_{\max} = \mathbb{R} \cup \{-\infty\}$. Тогда min-алгебра будет представлять собой полукольцо вида

$$(\mathbb{R}_{\min}, \oplus, \odot),$$

где \oplus — операция тропического сложения, определяемая как $x \oplus y = \min\{x, y\}$, \odot — операция тропического умножения, определяемая как $x \odot y = x + y$, а max-алгеброй будет являться полукольцо вида

$$(\mathbb{R}_{\max}, \boxplus, \odot),$$

где тропическое сложение определяется как $x \boxplus y = \max\{x, y\}$. [3]

3. О СТРУКТУРЕ НЕЙРОННЫХ СЕТЕЙ

Многослойная полносвязная нейронная сеть, состоящая из L слоев, может быть описана как отображение $v: \mathbb{R}^d \rightarrow \mathbb{R}^p$, определяемое композицией функций

$$v = \sigma^{(L)} \cdot \rho^{(L)} \cdot \sigma^{(L-1)} \cdot \rho^{(L-1)} \cdot \dots \cdot \sigma^{(1)} \cdot \rho^{(1)}. \quad (1)$$

Функции $\rho^{(1)}, \dots, \rho^{(L)}$ в (1) определяются с помощью весовых матриц и вектора смещений, а функции активации $\sigma^{(1)}, \dots, \sigma^{(L)}$ выбираются и фиксируются заранее. Функция активации — это нелинейная функция, которая применяется к выходу каждого нейрона слоя. Функции активации обычно применяются после применения функции преактивации — линейной функции, такой как умножение вектора входных данных на весовую матрицу и добавление вектора смещения. Внесенная таким образом нелинейность позволяет моделировать более сложные процессы. [4–6]

Многослойная полносвязная нейронная сеть состоит из нескольких слоев нейронов, каждый из которых соединен с каждым нейроном следующего слоя. Входные данные передаются на первый слой, где в каждом нейроне к данным применяются сначала функция преактивации, затем функция активации, после чего результат передается на следующий слой, где происходит аналогичный процесс. Последний слой выдает окончательный результат. [7–9]

Существует множество функций активации, таких как сигмоида, гиперболический тангенс, MaxOut, ReLU, Leaky ReLU, ELU, SoftMax и другие. Каждая функция активации имеет свои преимущества и недостатки, и выбор функции активации зависит от конкретной задачи и структуры сети. Для применения аппарата тропической математики представляет интерес семейство кусочно-линейных функций активации, таких как ReLU, LeakyReLU и MaxOut.

Поскольку такие функции активации непрерывны, нейронная сеть с линейными слоями преактивации и кусочно-линейными непрерывными функциями активации

сама может быть представлена как кусочно-линейная непрерывная (далее PLC) функция. Класс PLC-функций важен благодаря двум свойствам:

1. Композиция PLC-функций также является PLC-функцией, то есть PLC-функции — множество, замкнутое относительно композиции.
2. Любая непрерывная функция на компактном множестве может быть равномерно аппроксимирована PLC-функциями. Другими словами, PLC-функции представляют собой класс *универсальных аппроксиматоров*.

Поскольку и тропическое сложение (выбор минимума или максимума) и тропическое умножение (вещественное сложение) относятся к классу PLC, то любой тропический многочлен также является PLC-функцией. Однако есть два ограничения:

1. Тропические многочлены имеют фиксированную выпуклость и фиксированную монотонность по любой переменной: выпуклые и возрастающие для \max -алгебры и вогнутые и убывающие для \min -алгебры. Это следует непосредственно из определения тропических операций.
2. Градиент PLC-функции, представленной тропическим многочленом, имеет только неотрицательные целые компоненты. Это следует из того, что мономы, составляющие многочлен, имеют целые неотрицательные степени.

Первое ограничение можно снять, введя *тропическую рациональную функцию* как тропическое отношение двух тропических многочленов:

$$v(x) = F(x) \oslash G(x) = F(x) - G(x).$$

Второе ограничение можно снять, если обобщить понятие тропического многочлена до понятия *ряда Пюизо*. Можно доказать, что любая PLC-функция может быть представлена в виде тропической рациональной функции.

По построению тропическая рациональная функция является непрерывной кусочно-линейной функцией, что, в свою очередь, означает, что каждая из областей, на которой функция линейна, является многогранником. Тропическая гиперповерхность (объединение границ этих многогранников) разбивает пространство \mathbb{R}^d на выпуклые ячейки. Тогда граница между областями принятия решений нейронной сети может быть представлена как подмножество тропической гиперповерхности, соответствующей тропическому многочлену, а области принятия решений, на которые тропическая гиперповерхность делит пространство решений, будут являться блоками построения нейронной сети.

4. НЕЙРОННЫЕ СЕТИ С ТРОПИЧЕСКИМИ СЛОЯМИ

Как и прежде, мы будем называть *тропическим слоем* слой, в котором будут производиться тропические вычисления.

Проиллюстрируем это понятие на простейшем примере (в терминах \max -алгебры).

Сначала определим входные данные, которые будут подаваться на входной слой нейронной сети. В качестве примера рассмотрим двумерный входной вектор $[x_1, x_2]$.

Затем зададим скрытый слой, для каждого из нейронов которого обозначим весовой коэффициент w_i и смещение b_i .

В каждом нейроне скрытого слоя вектор входных данных тропически умножается на весовую матрицу, а затем результат тропически складывается с вектором смещений:

$$h = [w_{1,1} \odot x_1 \boxplus w_{1,2} \odot x_2 \boxplus b_1; w_{2,1} \odot x_1 \boxplus w_{2,2} \odot x_2 \boxplus b_2].$$

Зададим выходной слой, состоящий из одного нейрона:

$$h = w \odot x \boxplus b.$$

5. ТРОПИЧЕСКИЕ НЕЙРОННЫЕ СЕТИ: ПРЕИМУЩЕСТВА

Тропическая арифметика и алгебра многочленов и рациональных функций — готовый аппарат работы с кусочно-линейными функциями, простой и эффективный. Ниже перечислены основные преимущества использования тропической математики в архитектуре нейронных сетей:

1. Более быстрое обучение. Использование тропических слоев может привести к ускорению обучения нейронной сети, поскольку сокращает количество операций, которые необходимо выполнить для достижения результатов [10].
2. Повышение точности. Использование тропической математики может привести к повышению точности нейронной сети [11, 12].
3. Более простая архитектура. Тропическая математика может помочь упростить архитектуру нейронной сети.

6. АРХИТЕКТУРА НЕЙРОННЫХ СЕТЕЙ: РАССМОТРЕННЫЕ ВАРИАНТЫ

В качестве эталона была выбрана нейронная сеть с классической архитектурой Linear + ReLU + Linear.

Для реализации тропических вычислений использованы стандартные функции фреймворка PyTorch.

Были реализованы два тропических слоя для min-алгебры (Min) и max-алгебры (Max). Весовые функции и функции смещения подбирались автоматически с помощью модуля NN, встроенного в PyTorch [13].

Нейронная сеть обучалась на открытом датасете Heart Disease, используемом для диагностики заболеваний сердца.

Для сравнения производительности нейронной сети с тропическими слоями будут использованы пять различных архитектур:

1. Модель смешанной архитектуры с линейным слоем, min-слоем и max-слоем.
2. Модель со смешанной архитектурой, в которой поочередно используются линейный и max-слой.
3. Модель смешанной архитектуры, в которой тропические слои расположены между двумя линейными слоями.
4. Модель с тропической архитектурой, состоящая из max-слоев.
5. Модель тропической архитектуры с поочередным использованием min- и max-слоев.

Для каждой из 6 моделей были выбраны одинаковые гиперпараметры. Все модели обучались в течение 100 эпох. В качестве оптимизатора были выбраны стандартные оптимизаторы Adam и SGD, а в качестве функции потерь — стандартная функция CrossEntropyLoss.

Основное различие между оптимизаторами заключается в том, что оптимизатор Adam использует скользящее среднее градиентов первого и второго порядка, что позволяет ему адаптироваться к различным скоростям изменения градиента во время обучения. Оптимизатор SGD, напротив, просто обновляет веса на основе градиента функции потерь.

7. РЕЗУЛЬТАТЫ

Полученные данные о точности обучения сведены в таблицы. Результаты обучения с оптимизатором Adam представлены в таблице 1, с оптимизатором SGD — в таблице 2. Подробные графики представлены в [приложении](#).

Таблица 1. Оптимизатор Adam

Модель	Точность, %
Lin + ReLU + Lin	77,3
Lin+Min+Max	50,0
Lin+Max+Lin+Max	77,3
Lin+Min+Max+Lin	77,7
Max+Max	77,4
Min+Max	73,9

При использовании оптимизатора Adam наилучшую точность (77,7%) показала модель, использующая смешанную архитектуру Lin+Min+Max+Lin. Точность этой архитектуры оказалась на 0,4% выше, чем у классической архитектуры. Кроме того, точность тропической модели, состоящей из двух max-слоев, превышает точность классической модели на 0,1%. Модель со смешанной архитектурой Lin+Max+Lin+Max показала ту же точность, что и классическая модель, — 77,3%. Точность тропической модели Min+Max составила 73,9%, что на 3,4% ниже точности классической модели.

Таблица 2. Оптимизатор SGD

Модель	Точность, %
Lin + ReLU + Lin	76,5
Lin+Min+Max	49,9
Lin+Max+Lin+Max	76,7
Lin+Min+Max+Lin	76,5
Max+Max	76,3
Min+Max	50,0

Точность моделей с использованием оптимизатора SGD оказалась ниже, чем с использованием оптимизатора Adam. Наилучшую точность показала смешанная модель Lin+Max+Lin+Max — 76,7%. Точность классической и Lin+Min+Max+Lin моделей была одинаковой и составила 76,5%. В то же время точность тропической модели Min+Max стабилизировалась, что свидетельствует о прекращении обучения модели при использовании SGD-оптимизатора.

Модель с архитектурой Lin+Min+Max вообще не показала способности к обучению. Точность обучения обоих оптимизаторов оказалась равной 50%. Этот результат можно объяснить тем, что используемые оптимизаторы не приспособлены для работы с сильно негладкими функциями активации. Кроме того, начальные значения входного вектора могли быть выбраны неудачно.

Практически для всех моделей использование оптимизатора Adam оказалось более эффективным, чем использование оптимизатора SGD. В то же время некоторые модели с тропической архитектурой при использовании Adam обучались значительно дольше, чем модели с использованием SGD.

8. ЗАКЛЮЧЕНИЕ

В настоящей работе реализовано шесть архитектур нейронных сетей: пять — с использованием тропических слоев и одна классическая в качестве эталонной. Проведено сравнение скорости и точности обучения для реализованных архитектур на конкретном датасете и с применением двух различных оптимизаторов — Adam и SGD.

По результатам работы наилучшую точность показала модель со смешанной архитектурой, в которой используются два линейных слоя, а между ними расположены min- и max-слои, а также оптимизатор Adam. Кроме того, модель смешанной архитектуры с линейным и максимальным слоем и модель тропической архитектуры с двумя максимальными слоями между ними показали улучшение по сравнению с классической моделью.

Улучшение времени работы нейронной сети с тропическими слоями по сравнению с классической нейронной сетью не достигнуто. Это требует ускорения вычислений, которые реализуются в тропических слоях, например создание динамической библиотеки, ускоряющей тензорные вычисления.

Использование тропических функций для построения нейронных сетей имеет свои преимущества и ограничения и должно рассматриваться в контексте конкретной задачи и требований к нейронной сети.

Помимо полносвязных нейронных сетей, тропические слои могут использоваться в качестве сверточных слоев. Например, в работе [14] использовались сверточные нейронные сети с тропическими слоями, которые показали улучшение точности моделирования по сравнению с обычными сверточными сетями. Дальнейшие исследования будут направлены на применение тропических слоев в более сложных архитектурах, а также на автоматизацию выбора параметров обучения, которые могут быть использованы для существенного повышения точности нейронных сетей, содержащих тропический слой.

Список литературы

1. Zhang L., Naitzat G., Lim L.-H. Tropical Geometry of Deep Neural Networks // Proc. of International Conference on Machine Learning (PMLR), 2018. 2018. Vol. 80. P. 5824–5832.
2. Luo Y., Shiqing F. Min-Max-Plus Neural Networks // arXiv, preprint arXiv:2102.06358, 2021.
3. Казарян М. Э. Тропическая геометрия. М.: МНЦМО, 2012. 43 с.
4. Ceyhan O. Algorithmic Complexities in Backpropagation and Tropical Neural Networks // arXiv, preprint arXiv:2101.00717, 2021.
5. Smyrnis G., Maragos P. Tropical polynomial division and neural networks // arXiv, preprint arXiv:1911.12922, 2019.
6. Hartman P. On functions representable as a difference of convex functions // Pacific Journal of Mathematics. 1959. Vol. 9, № 3. P. 707–713.
7. Smyrnis G., Maragos P., Retsinas G. Maxpolynomial Division with Application To Neural Network Simplification // Proc. of ICASSP 2020 — IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, May 2020. 2020. P. 4192–4196. doi:10.1109/icassp40776.2020.9053540
8. Nikolaos D., Maragos M. P. Advances in the training, pruning and enforcement of shape constraints of Morphological Neural Networks using Tropical Algebra // arXiv, preprint arXiv:2011.07643v1, 2020.
9. Ritter G. X., Sussner P. An introduction to morphological neural networks // Proc. of 13th Int. Conf. on Pattern Recognition, IEEE, Vienna, Austria, 1996, 1996. Vol. 4. P. 709–717. doi:10.1109/ICPR.1996.547657
10. Charisopoulos V., Maragos P. A Tropical Approach to Neural Networks with Piecewise Linear Activations // arXiv, preprint arXiv:1805.0874, 2018.
11. Fournier G. A tropical approach to neural networks // M.S. thesis, Dept. Mathematics, Université de Lille, Lille, France, 2019.

12. Alfarra M. H. A. Applications of Tropical Geometry in Deep Neural Networks // M.S. thesis, Dept. Computer, Electrical and Mathematical Science and Engineering, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia, 2019.
13. PyTorch Contributors, Pytorch documentation // Pytorch.org, 2023. URL: <https://pytorch.org/docs/stable/index.html> (date: 24.04.2023).
14. Fan S., Liu L., Luo Y. An Alternative Practice of Tropical Convolution to Traditional Convolutional Neural Networks // Proc. of 2021 The 5th International Conference on Compute and Data Analysis, Sanya, China. 2021. P. 162–168. doi:10.1145/3456529.3456557

Приложение

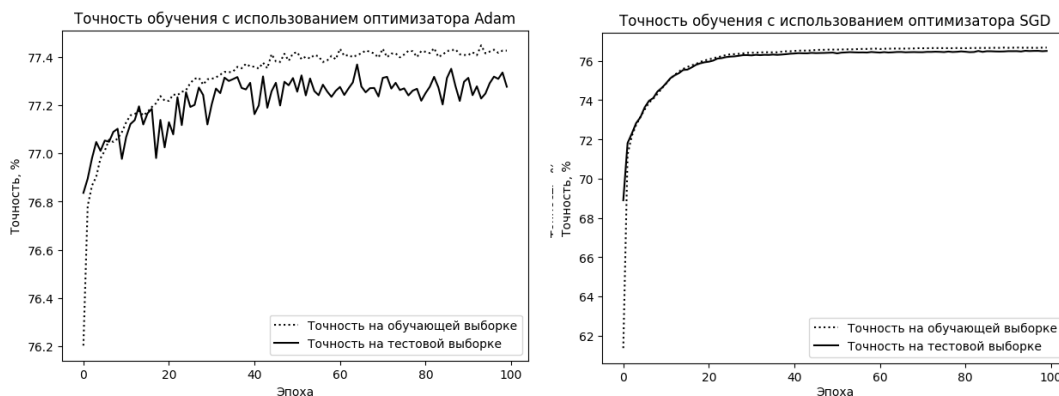


Рис. 1. Классическая модель



Рис. 2. Модель Lin+Max+Lin+Max

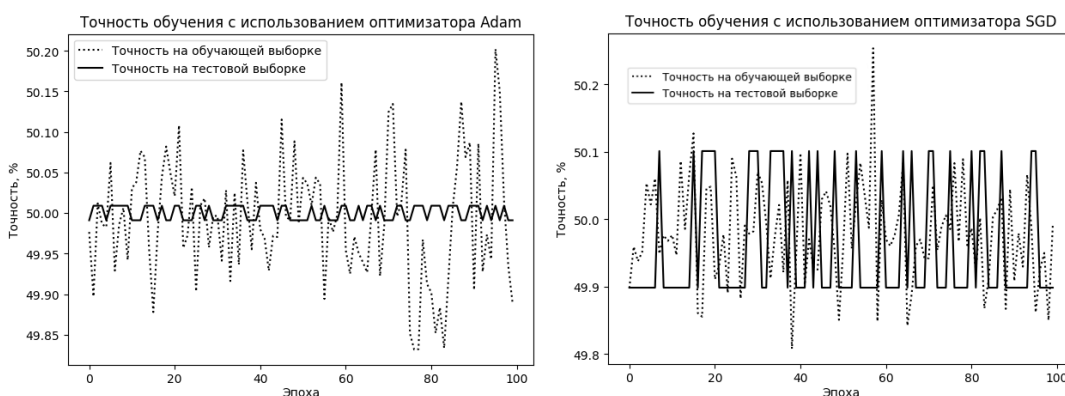


Рис. 3. Модель Lin+Min+Max

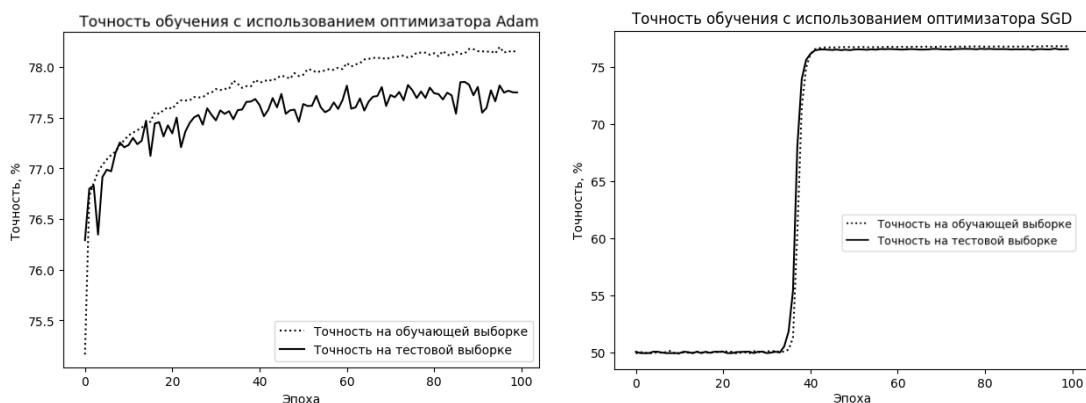


Рис. 4. Модель Lin+Min+Max+Lin

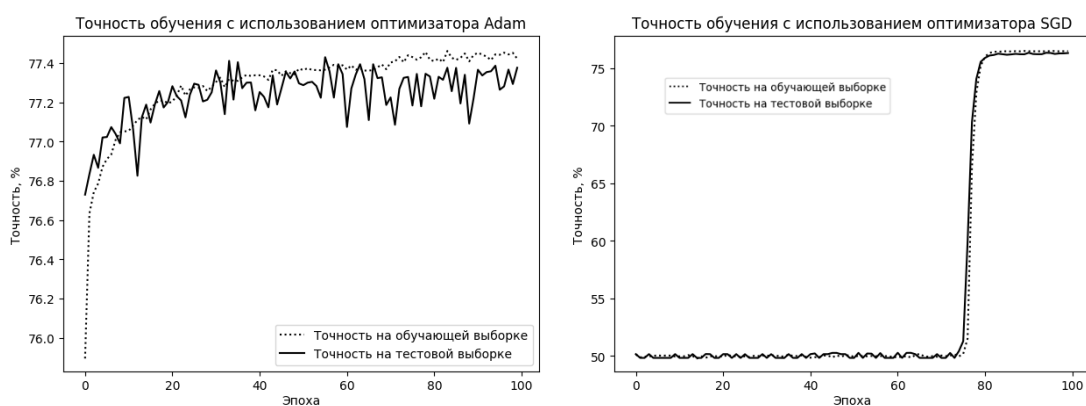


Рис. 5. Модель Max+Max

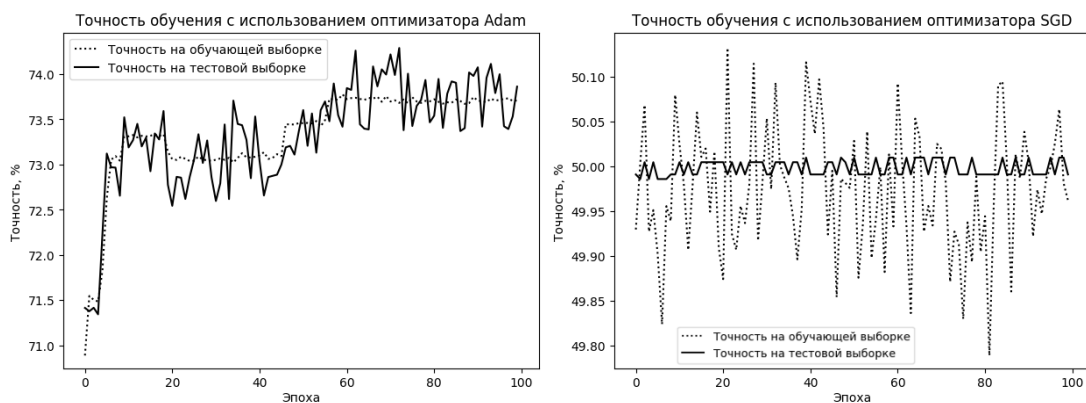


Рис. 6. Модель Min+Max

Поступила в редакцию 15.06.2023, окончательный вариант — 20.07.2023.

Петрова Анна, аспирантка 1 года кафедры алгоритмической математики СПбГЭТУ «ЛЭТИ»,
✉ anna.petrova.16.11@gmail.com

Казакевич Виктория Григорьевна, старший преподаватель кафедры алгоритмической математики СПбГЭТУ «ЛЭТИ», ✉ sokratt@gmail.com

Applications of Tropical Mathematics to Neural Network Architecture

Petrova A.¹, Postgraduate Student, ✉ anna.petrova.16.11@gmail.com
Kazakevich V. G.¹, Senior Lecturer, ✉ sokratt@gmail.com, orcid.org/0000-0002-3150-2105

¹Saint Petersburg Electrotechnical University,
5, building 3, st. Professora Popova, 197022, Saint Petersburg, Russia

Abstract

Using the methods provided by tropical mathematics we can simplify the structure of a neural network, which increases its explainability, without decreasing its accuracy. This paper aims to explore the use of tropical functions in neural networks and compare their efficiency with classical ones. Theoretical framework of tropical mathematics is a semiring with idempotent addition, which is a natural approach to piecewise-linear neural networks, e.g. networks with ReLU activation. Within this approach, piecewise-linear convex function is a tropical polynomial, and general piecewise-linear functions are tropical rational functions. Thus a layer of a neural network with linear preactivation and ReLU activation can be viewed as a vector-valued tropical rational function, which in turn can be represented by two tropical layers. Two tropical layers were implemented, and five tropical architectures were constructed. The models were trained on a heart disease dataset, aiming to determine the presence of heart disease. All models had the same hyperparameters. Each of the models was trained for 100 epochs using Adam and SGD optimizers. The results of the comparison showed that the best accuracy was achieved by a mixed-architecture model using two linear layers. The comparison results showed that the best accuracy was achieved by a mixed-architecture model with two linear layers with a min-layer and a max-layer in between. This accuracy was achieved by using an Adam optimizer. The classical model scored 77.3% and the tropical 77.7%.

Keywords: *tropical mathematics, neural networks.*

Citation: A. Petrova and V. G. Kazakevich, "Applications of Tropical Mathematics to Neural Network Architecture," *Computer tools in education*, no. 3, pp. 18-27, 2023 (in Russian); doi: 10.32603/2071-2340-2023-3-18-27

References

1. L. Zhang, G. Naitzat, and L.-H. Lim, "Tropical Geometry of Deep Neural Networks," in *Proc. of International Conference on Machine Learning (PMLR)*, 2018, vol. 80, pp. 5824–5832, 2018.
2. Y. Luo and F. Shiqing, "Min-Max-Plus Neural Networks," in *arXiv*, [Online], preprint arXiv:2102.06358, 2021.
3. M. Kazaryan, *Tropical geometry*, Moscow: MCNMO, 2012 (in Russian).
4. O. Ceyhan, "Algorithmic Complexities in Backpropagation and Tropical Neural Networks," in *arXiv*, [Online], preprint arXiv:2101.00717, 2021.
5. G. Smyrnis and P. Maragos, "Tropical polynomial division and neural networks," in *arXiv*, [Online], preprint arXiv:1911.12922, 2019.

6. P. Hartman, “On functions representable as a difference of convex functions,” *Pacific Journal of Mathematics*, vol. 9, no. 3, pp. 707–713, 1959.
7. G. Smyrnis, P. Maragos, and G. Retsinas, “Maxpolynomial Division with Application To Neural Network Simplification,” in *Proc. of ICASSP 2020 — IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, May 2020*, pp. 4192–4196, 2020; doi:10.1109/icassp40776.2020.9053540
8. D. Nikolaos and M. P. Maragos, “Advances in the training, pruning and enforcement of shape constraints of Morphological Neural Networks using Tropical Algebra,” in *arXiv*, [Online], preprint arXiv:2011.07643v1, 2020.
9. G. X. Ritter and P. Sussner, “An introduction to morphological neural networks,” in *Proc. of 13th Int. Conf. on Pattern Recognition, IEEE, Vienna, Austria, 1996*, vol. 4, pp. 709–717, 1996; doi:10.1109/ICPR.1996.547657
10. V. Charisopoulos and P. Maragos, “A Tropical Approach to Neural Networks with Piecewise Linear Activations,” in *arXiv*, [Online], preprint arXiv:1805.0874, 2018.
11. G. Fournier, “A tropical approach to neural networks,” M.S. thesis, Dept. Mathematics, Université de Lille, Lille, France, 2019.
12. M. H. A. Alfarrá, “Applications of Tropical Geometry in Deep Neural Networks,” M.S. thesis, Dept. Computer, Electrical and Mathematical Science and Engineering, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia, 2019.
13. PyTorch Contributors, “Pytorch documentation,” in *Pytorch.org*, 2023. [Online]. Available: <https://pytorch.org/docs/stable/index.html>
14. S. Fan, L. Liu, and Y. Luo, “An Alternative Practice of Tropical Convolution to Traditional Convolutional Neural Networks,” in *Proc. of 2021 The 5th International Conference on Compute and Data Analysis, Sanya, China*, pp. 162–168, 2021; doi:10.1145/3456529.3456557

Received 15-06-2023, the final version — 20-07-2023.

Anna Petrova, Postgraduate Student of the Algorithmic Mathematics Department, Saint Petersburg Electrotechnical University, ✉ anna.petrova.16.11@gmail.com

Victoria Kazakevich, Senior Lecturer of Algorithmic Mathematics Department, Saint Petersburg Electrotechnical University, ✉ sokratt@gmail.com