Г. В. Кормаков[1]

ИНТЕРПРЕТАЦИЯ ПРОЦЕССА ПЕРЕОБУЧЕНИЯ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ С ИСПОЛЬЗОВАНИЕМ ТРОПИЧЕСКОЙ ГЕОМЕТРИИ*

Введение и основные определения

В данной работе рассматривается исследование структуры параметров искусственных нейронных сетей с функциями активации *ReLU* методами тропической геометрии.

Среди используемых в современных архитектурах ИНС функций активации можно выделить класс функций вида

$$p(\vec{x}) = \max_{i=\overline{1,k}}(\vec{w}_i^T\vec{x} + b_i), \quad \vec{x}, \vec{w}_i \in \mathbb{R}^n, \ b_i \in \mathbb{R}. \tag{1}$$

Приведём основные примеры таких функций активации для $v = \vec{w}^T\vec{x} + b, \quad \vec{x}, \vec{w} \in \mathbb{R}^n, \ b \in \mathbb{R}$:

$$ReLU(v) = \max(0, v)$$

$$LeakyReLU_\alpha(v) = \max(v, \alpha v), \quad \alpha \in (0,1)$$

$$Maxout(\vec{x}) = \max_{j\in[k]}(0, \vec{W}_j^T\vec{x} + b_j), \quad \vec{W} \in \mathbb{R}^{n\times k}, \vec{b} \in \mathbb{R}^k$$

Везде далее будем называть класс данных функций в привязке к наиболее часто используемой функции — *ReLU-подобный* класс функций активации. В полносвязных ИНС функции активации стоят после линейных слоёв, что приводит к максимуму от кусочно-линейных функций (выражение (1)).

Теория тропической математики, разработанная в прошлом веке для анализа экономических и физических эффектов, позволила изучать кусочно-линейные функции с помощью перехода от операции сложения к взятию максимума и от умножения к сложению[2].

Тропическое полукольцо над множеством $\mathbb{R}_{max} = \mathbb{R} \cup \{-\infty\}$ можно задать следующим образом.

Определение 1 (Тропическое полукольцо)**.** *Тропическим полукольцом* $\mathbb{T}$ называется алгебраическая структура $(\mathbb{R}_{max}, \oplus, \odot)$, где $\oplus$ и $\odot$ задают

[1]Факультет ВМК МГУ имени М. В. Ломоносова, e-mail: gkormakov@gmail.com.

*Работа выполнена при поддержке госбюджетной темы НИР № 5.4 ВМК МГУ.

[2]Подробное введение в теорию тропической математики и тропической геометрии дано в [9], [10]

операцию тропического сложения и тропического умножения соответственно.

Везде далее будут использоваться операции следующего вида[1]:

$$x \oplus y = \max\{x, y\}, \quad x \odot y = x + y, \quad \forall x, y \in \mathbb{T} \tag{2}$$

Тропическая степень определяется по аналогии с классическим случаем: $x^{\odot a} = \underbrace{x \odot \cdots \odot x}_{a} = a \cdot x, \quad x \in \mathbb{T},\ a \in \mathbb{N}$. При этом

1. Для элемента $(-\infty)$ положительная степень определяется следующим выражением: $(-\infty)^{\odot a} = \begin{cases} -\infty, & a > 0 \\ 0, & a = 0 \end{cases}, \quad a \in \mathbb{N}.$
2. Обратным элементом по тропическому умножению является число, взятое с противоположным знаком: $x^{\odot(-1)} := -x$. Отсюда определяется операция возведения в отрицательную степень: $\forall a \in \mathbb{N} \quad x^{\odot(-a)} = (-x)^{\odot a}$.
3. Не существует обратного элемента по тропическому умножению для $-\infty$.
4. Для отрицательных степеней не определена степень элемента $-\infty$: $\forall a \in \mathbb{Z},\ a < 0\,, \quad \nexists (-\infty)^{\odot a}$.

Это позволяет ввести термин *тропического полинома*.

Определение 2 (Тропический полином)**.** Пусть имеются d тропических переменных $\vec{x} \in \mathbb{T}^d$, их наборы степеней $\vec{a}_i \in \mathbb{N}^d$ и координаты $c_i \in \mathbb{T},\ i \in \overline{1,n}$. *Тропическим полиномом с n мономами* называется отображение $f : \mathbb{T}^d \to \mathbb{T}$ вида

$$f(\vec{x}) = \bigoplus_{i=1}^{n} \left(c_i \odot \vec{x}^{\odot \vec{a}_i} \right) = \max_{i=1\ldots n} \left\{ c_i + \sum_{j=1}^{d} a_{ij} x_j \right\}, \quad \forall \vec{a}_i \neq \vec{a}_j,\ i \neq j. \tag{3}$$

Векторное возведение в степень $\vec{x}^{\odot \vec{a}}$ записывается поэлементно в виде выражения $\vec{x}^{\odot \vec{a}} = x_1^{\odot a_1} \odot x_2^{\odot a_2} \cdots \odot x_d^{\odot a_d}$.

Не нарушая общности, будем далее считать степени тропического полинома вещественными числами (поскольку они задают коэффициенты линейных функций). Множество тропических полиномов от d переменных обозначается $\mathbb{T}[x_1, \ldots x_d]$. Множество $\mathbb{T}[x_1, \ldots x_d]$ можно считать подмножеством более широкого класса тропических рациональных функций.

Определение 3 (Тропическая рациональная функция)**.** *Тропической рациональной функцией* называется тропическое деление двух

[1]Тропическое полуполе также определяется, например, с помощью операции сложения min над $\mathbb{R} \cup \{+\infty\}$

тропических полиномов (стандартная разность) $h(\vec{x}) = f(\vec{x}) \oslash g(\vec{x}) = f(\vec{x}) - g(\vec{x})$.

Обозначение: $h = f \oslash g \in \mathbb{T}(x_1, \ldots, x_d)$.

Если рассматривать векторное действие тропических полиномов, то вводится термин *тропического отображения*.

Определение 4 (Тропическое отображение)**.** Отображение $F : \mathbb{T}^d \to \mathbb{T}^p, \vec{x} = (x_1, \ldots, x_d) \mapsto (f_1(\vec{x}), \ldots, f_p(\vec{x}))$ называется *тропическим полиномиальным (рациональным) отображением*, если каждая функция $f_i : \mathbb{T}^d \to \mathbb{T}$, $i = \overline{1,p}$ является тропическим полиномом (тропической рациональной функцией).

Обозначение: $F \in Pol(d,p)$ $(F \in Rat(d,p))$

Ключевым фактом для всей области анализа теоретической структуры ИНС с функциями активации вида *ReLU* является следующая теорема.

Теорема 1. *(О связи ИНС с тропическими полиномами [16]) Полносвязная нейронная сеть с d входами и p выходами, при выполнении следующих ограничений на архитектуру:*

1. *Матрицы весов являются целочисленными*[1].
2. *Векторы весов (свободные члены в линейных слоях) вещественные.*
3. *Функции активации принадлежат ReLU*-подобному *классу.*

является отображением $\nu : \mathbb{R}^d \to \mathbb{R}^p$ *вида:*

$$\nu(\vec{x}) = F(\vec{x}) \oslash G(\vec{x}) = F(\vec{x}) - G(\vec{x}), \tag{4}$$

где $F, G \in Pol(d,p)$ *— тропические полиномиальные отображения. Следовательно,* $\nu \in Rat(d,p)$ *— тропическое рациональное отображение.*

Данное утверждение позволяет использовать для анализа теоретических структур ИНС аппарат тропической геометрии над полиномами.

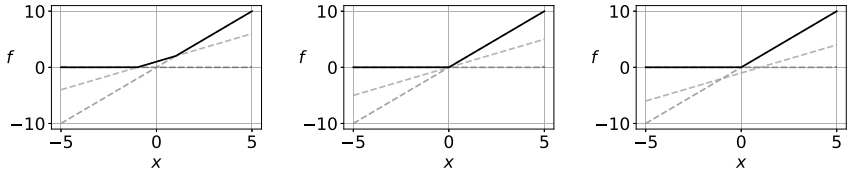
Определение 5 (Тропическая гиперповерхность)**.** *Тропической гиперповерхностью* тропического полинома $f(\vec{x}) = \bigoplus_{i=1}^{n} \left(c_i \odot \vec{x}^{\odot \vec{a}_i}\right)$ называется множество точек недифференциируемости в пространстве:

$$\mathscr{T}(f) = \left\{ \vec{x} \in \mathbb{T}^d : \; c_i \odot \vec{x}^{\odot \vec{a}_i} = c_j \odot \vec{x}^{\odot \vec{a}_j} = f(\vec{x}), \;\; \forall \vec{a}_i \neq \vec{a}_j \right\} \tag{5}$$

Таким образом, тропический полином разделяет исходное пространство на выпуклые области, где тропический полином линеен в обычном смысле.

[1] В последствии отмечается, что ограничение может быть расширено до рациональных весов. Что, фактически, даёт возможность применения на практике.

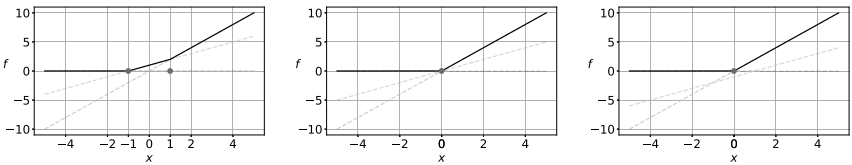
Приведём примеры тропических полиномов и их гиперповерхностей. На рис. 1 изображены тропические полиномы $f_1(x) = \max\{0, 1+x, 2x\}$, $f_2(x) = \max\{0, x, 2x\}$ и $f_3(x) = \max\{0, -1+x, 2x\}$. Важно отметить, что итоговый вид разделяющих прямых может быть одинаковым для полиномов, состоящих из разных мономов (рис. 1b, 1c).



(a) $\max\{0, x+1, 2x\}$ (b) $\max\{0, x, 2x\}$ (c) $\max\{0, x-1, 2x\}$

Рис. 1. Примеры тропических полиномов

Тропические гиперповерхности $\mathscr{T}(f_1), \mathscr{T}(f_2), \mathscr{T}(f_3)$ приведённых на рис. 1 полиномов, состоят из точек нелинейности. На рис. 2 показаны точки, в которых достигается нелинейная часть тропических полиномов с рис. 1.



(a) $\mathscr{T}(f_1) = \{-1, 1\}$ (b) $\mathscr{T}(f_2) = \{0\}$ (c) $\mathscr{T}(f_2) = \{0\}$

Рис. 2. Тропические гиперповерхности для полиномов рис. 1

Следующим связанным с тропическим полиномом объектом является *многогранник Ньютона*.

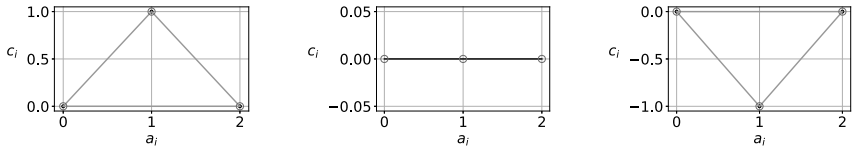
Определение 6 (Многогранники Ньютона)**.** *Многогранником Ньютона* тропического полинома $f(\vec{x}) = \bigoplus_{i=1}^{n} \left(c_i \odot \vec{x}^{\odot \vec{a}_i}\right)$ называется выпуклая оболочка его тропических степеней с коэффициентом, отличным от тропического нуля.

$$\Delta(f) = ConvHull\left\{\vec{a}_i \in \mathbb{R}^d : \; c_i \neq -\infty, \; i = 1, \ldots, n\right\} \tag{6}$$

Однако для полного определения тропического полинома из многогранника Ньютона строят расширенное множество с тропическими коэффициентами (называемым *расширенным многогранником Ньютона*).

$$\mathscr{P}(f) = ConvHull\left\{(\vec{a}_i, c_i) \in \mathbb{R}^d \times \mathbb{T} : \; i = 1, \ldots, n\right\} \tag{7}$$

Для полиномов на рис. 1 многогранники Ньютона выглядят одинаково и совпадают с расширенным многогранником Ньютона полинома на рис. 3b. Для остальных полиномов расширенный многогранник Ньютона даёт дополнительную информацию относительно одномерного случая из-за наличия свободных членов в линейных частях мономов (см. рис. 3a, 3c)



(a) $\mathscr{P}(f_1)$ (b) $\mathscr{P}(f_2)$ (c) $\mathscr{P}(f_3)$

Рис. 3. Расширенные многогранники Ньютона для полиномов рис.1

Ключевым для дальнейшего анализа понятием тропической геометрии является двойственное определение к тропической гиперповерхности — это *двойственное подразделение*[1]. Двойственность данного множества заключается в задании точек пространства, однозначно задающих области линейности тропического полинома. Формальнее, рассматривается *множество верхних граней* выпуклой оболочки расширенного многогранника Ньютона $UF(\mathscr{P}(f))$.

Определение 7. *Верхней точкой расширенного многогранника Ньютона* называется такая точка $\vec{x} \in \mathscr{P}(f) \subset \mathbb{R}^d \times \mathbb{T}$, что луч, исходящий из этой точки в положительном направлении по последней координате, пересекает многогранник только в этой точке:

$$\forall \vec{\lambda} \in \vec{0} \times \mathbb{R}_{\geqslant 0} \quad (\vec{x} + \vec{\lambda}) \cap \mathscr{P}(f) = \{\vec{x}\} \tag{8}$$

Верхней гранью $UF(\mathscr{P}(f))$ *расширенного многогранника Ньютона* называется множество его верхних точек:

$$UF(\mathscr{P}(f)) = \left\{ \vec{x} \in \mathscr{P}(f) : \quad \forall \vec{\lambda} \in \vec{0} \times \mathbb{R}_{\geqslant 0} \quad (\vec{x} + \vec{\lambda}) \cap \mathscr{P}(f) = \{\vec{x}\} \right\} \tag{9}$$

Определение 8. Пусть $f(\vec{x}) : \mathbb{T}^d \to \mathbb{T}$ — тропический полином, $\mathscr{P}(f)$ — его расширенный многогранник Ньютона и $UF(\mathscr{P}(f))$ — множество верхних граней многогранника $\mathscr{P}(f)$.

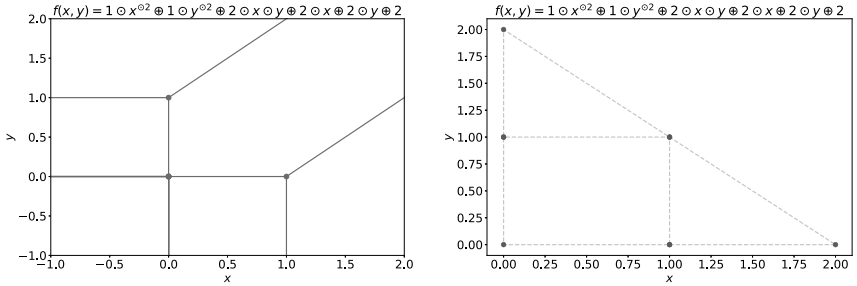
Тогда *двойственным подразделением*, заданным тропическим полиномом $f(\vec{x})$, называется множество проекций последних координат верхних граней из $UF(\mathscr{P}(f))$.

$$\delta(f) = \left\{ \pi(\vec{p}) \subset \mathbb{R}^d : \ \vec{p} \in UF(\mathscr{P}(f)) \subset \mathbb{R}^d \times \mathbb{T} \right\}, \tag{10}$$

где $\pi(\cdot) : \mathbb{R}^d \times \mathbb{T} \to \mathbb{R}^d$ — оператор проецирования последней координаты.

[1](англ.) dual subdivision

Рассмотрим пример построения всех приведённых понятий для одного тропического полинома $f(x,y) = (1 \odot x^{\odot 2}) \oplus (1 \odot y^{\odot 2}) \oplus (2 \odot x \odot y) \oplus (2 \odot x) \oplus (2 \odot y) \oplus 2$. На рис. 4 изображена гиперповерхность для этого тропического полинома (4a) и двойственное к ней подразделение (4b). Гиперповерхностью являются непосредственно точки пересечения прямых (т. е. точки нелинейности), сами прямые на рисунки характеризуют мономы, составляющие тропический полином. Визуализация получения двойственного подразделения приведена на рис. 5 и будет описана ниже. Видно, что каждая точка двойственного подразделения отвечает за регион линейности тропического полинома (и со смещением по координатам будет однозначно задавать эту область).



(a) Гиперповерхность $\mathscr{T}(f)$ (b) Двойств. подразделение $\delta(f)$

Рис. 4. Построение двойственного подразделения для полинома $f(x,y)$

Процесс получения двойственного подразделения (по определению) изображён на рис. 5. Для понимания различия между направлениями координат, переменная x обозначена за x_1, а переменная y за x_2. Из множества всех граней выбираются верхние грани в положительном направлении по последней координате (положительное направление свободного члена обозначено за c), затем для каждой точки грани осуществляется проекция на первые две координаты (a_1 и a_2 соответственно).

Связь ИНС и тропических полиномов (доказанная в теореме 1) позволяет использовать определённые понятия тропической геометрии. Наиболее показательными являются результаты анализа двумерного случая, используемые также в данной работе.

Исследования двумерной геометрии параметров ИНС

В данном разделе приведены основные результаты и алгоритмы, полученные в различных исследованиях для случая нейронных сетей в задачах классификации. Основное внимание уделяется задаче бинарной

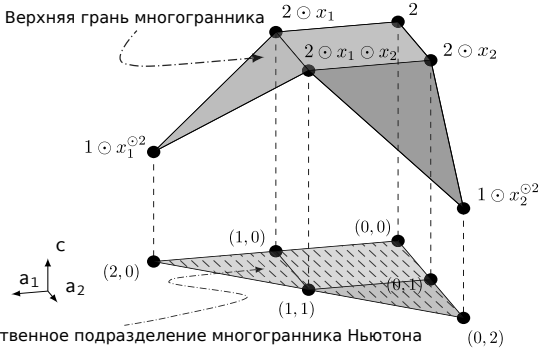


Рис. 5. Процесс формирования двойственного подразделения

классификации как наиболее интерпретируемой (с точки зрения двойственных подразделений).

Рассмотрим нейронную сеть следующего вида: $f(\vec{x}) = \vec{B}\max(\vec{A}\vec{x}+\vec{c}_1,\vec{0})+\vec{c}_2$, где max — поэлементная операция максимума над векторами. Для получения теоретических результатов, на параметры нейронной сети накладывают следующие ограничения: $\vec{A} \in \mathbb{Z}^{p\times n}$, $\vec{B} \in \mathbb{Z}^{2\times p}$, $\vec{c}_1 \in \mathbb{R}^p$, $\vec{c}_2 \in \mathbb{R}^2$. По теореме 1, на выходах ИНС будут получены тропические рациональные отображения $f_1(\vec{x}) = H_1(\vec{x}) \oslash Q_1(\vec{x})$ и $f_2(\vec{x}) = H_2(\vec{x}) \oslash Q_2(\vec{x})$ соответственно, где H_1, H_2, Q_1, Q_2 — тропические полиномы.

Определение 9. Пусть $\vec{x}_1,\ldots,\vec{x}_p \in \mathbb{R}^n$. *Зоноэдром*, образованным векторами $\vec{x}_1,\ldots,\vec{x}_p$ называется сумма Минковского сегментов, образованных данными векторами и началом координат, $[\vec{0},\vec{x}_i] = \{\alpha_i\vec{x}_i| \quad \alpha_i \in [0,1] \subset \mathbb{R}\}$:

$$\mathscr{Z}(\vec{x}_1,\ldots,\vec{x}_p) = \left\{ \sum_{i=1}^{p} \alpha_i\vec{x}_i :\ 0 \leqslant \alpha_i \leqslant 1 \right\}$$

.

Эквивалентно, *зоноэдром сегментов* $\{[\vec{l}_i,\vec{r}_i]\}_{i=1}^p$, $\vec{l}_i,\vec{r}_i \in \mathbb{R}^n$ называют сумму Минковского данных сегментов[1]. Обозначение: $\mathscr{Z}(\{[\vec{l}_i,\vec{r}_i]\}_{i=1}^p)$

Доказательство и формулировка общего результата для случая полносвязной ИНС с произвольным числом выходов и с весами-смещениями приведено в [16] и [1]. В данной работе будем опираться больше на результаты следующей теоремы.

Теорема 2. *(О границах принятия решений и геометрической структуре ИНС [1])*

[1]Эквивалентность устанавливается приведением одной из границ сегментов к началу координат и сдвигом итогового зоноэдра на сумму вычитаемых векторов.

Для двуслойной ИНС $f(\vec{x}) : \mathbb{R}^n \to \mathbb{R}^p \to \mathbb{R}^2$ *без свободных членов в слоях, с целочисленными матрицами весов* $\vec{A} \in \mathbb{Z}^{p\times n}$, $\vec{B} \in \mathbb{Z}^{2\times p}$ *и соответствующими тропическими полиномами на выходах* $f_1(\vec{x}) = H_1(\vec{x}) \oslash Q_1(\vec{x})$, $f_2(\vec{x}) = H_2(\vec{x}) \oslash Q_2(\vec{x})$.

Пусть $R(\vec{x}) = H_1(\vec{x}) \odot Q_2(\vec{x}) \oplus H_2(\vec{x}) \odot Q_1(\vec{x})$, $\vec{A}^- = \max(-\vec{A}, 0)$, $\vec{A}^+ = \max(\vec{A}, 0)$, $\vec{B}^- = \max(-\vec{B}, 0)$, $\vec{B}^+ = \max(\vec{B}, 0)$ *и* $\vec{A}_{j,:} \in \mathbb{R}^n$ *обозначает вектор-строку с номером j, тогда*

1. *Множество* $\mathscr{B} = \{\vec{x} \in \mathbb{R}^n : f_1(\vec{x}) = f_2(\vec{x})\}$ *границы принятия решения f является подмножеством тропической гиперповерхности тропического полинома* $R(\vec{x})$*:* $\mathscr{B} \subseteq \mathscr{T}(R(\vec{x}))$.

2. *Двойственное подразделение этого тропического полинома является выпуклой оболочкой зоноэдров из сегментов векторов-параметров:* $\delta(R(\vec{x})) = ConvHull(\mathscr{Z}_{G_1}, \mathscr{Z}_{G_2})$, *где*

$$\mathscr{Z}_{G_1} = \mathscr{Z}\left(\left\{(\vec{B}^+_{1,j} + \vec{B}^-_{2,j}) \cdot [\vec{A}^+_{j,:}, \vec{A}^-_{j,:}]\right\}_{j=1}^{p}\right) + (\vec{B}^-_{1,:} + \vec{B}^+_{2,:}) \cdot \vec{A}^-$$ *— зоноэдр в* $\mathbb{R}^n$, *образованный сегментами* $\left\{(\vec{B}^+_{1,j} + \vec{B}^-_{2,j}) \cdot [\vec{A}^+_{j,:}, \vec{A}^-_{j,:}]\right\}_{j=1}^{p}$ *и сдвинутый на вектор* $(\vec{B}^-_{1,:} + \vec{B}^+_{2,:}) \cdot \vec{A}^-$,

$$\mathscr{Z}_{G_2} = \mathscr{Z}\left(\left\{(\vec{B}^-_{1,j} + \vec{B}^+_{2,j}) \cdot [\vec{A}^+_{j,:}, \vec{A}^-_{j,:}]\right\}_{j=1}^{p}\right) + (\vec{B}^+_{1,:} + \vec{B}^-_{2,:}) \cdot \vec{A}^-$$ *— зоноэдр в* $\mathbb{R}^n$, *образованный сегментами* $\left\{(\vec{B}^-_{1,j} + \vec{B}^+_{2,j}) \cdot [\vec{A}^+_{j,:}, \vec{A}^-_{j,:}]\right\}_{j=1}^{p}$ *и сдвинутый на вектор* $(\vec{B}^+_{1,:} + \vec{B}^-_{2,:}) \cdot \vec{A}^-$.

Результат пункта 1 позволяет анализировать границы принятия решения по гиперповерхности тропического полинома и оценивать ограничения модели ИНС по числу разделяющих поверхностей, чему были посвящены работы [2], [4], [11], [16].

Однако прямой анализ гиперповерхности тропического полинома на практике сопоставим по сложности с исходной задачей определения границ принятия решений. Поэтому чаще применим результат пункта 2.

Утверждение пункта 2 использовалось ранее для

1. Алгоритма разреживания весов (pruning) на основе решения оптимизационной задачи над двойственным подразделением обученной ИНС [14].

2. Исследования вида двойственных подразделений при инициализации ИНС и осуществления атак на обученные ИНС, исходя из вида двойственных подразделений [1].

Важно отметить, что в ИНС без свободных членов в слоях двойственное подразделение становится многогранником Ньютона (поскольку оператор

проекции становится тождественным). Что позволяет поставить в однозначное соответствие нормали граней многогранника с гиперповерхностью тропического полинома.

Основной темой работы является анализ динамики двойственных подразделений (в процессе обучения ИНС) в зависимости от начальной инициализации и выделение характеристик переобучения, исходя только из вида двойственных представлений параметров в процессе обучения.

Динамика двойственных подразделений параметров ИНС

В разделе формируется полученная в рамках данной работы постановка задачи анализа переобучения ИНС с помощью тропической геометрии, описывается набор исследуемых в экспериментальной части характеристик и выдвигается гипотеза о моменте переобучения, исходя только из знаний о динамике исследуемых характеристик.

Везде далее будем рассматривать ИНС с одним скрытым слоем размера P, удовлетворяющих ограничениям теоремы 2. При этом, как отмечалось ранее, ограничения на целочисленность весов модели на практике сводятся до рациональных весов (т. е. до вещественных чисел с машинной точностью).

На сегодняшний день, открытыми остаются следующие вопросы:

1. Как начальная форма зоноэдров (инициализация весов) влияет на конечные представления? (вопрос частично освещался в [1])
2. Как изменяется форма зоноэдров, представляющих тропические полиномы выходов, в процессе обучения?
3. Возможно ли определить момент начала переобучения, опираясь на знание только о геометрии зоноэдров?

Поэтому *целями данной работы* по анализу динамики двойственных подразделений параметров ИНС, удовлетворяющих условиям теоремы 2, являются:

1. Выявление зависимости конечного вида зоноэдров классов от начальной инициализации.
2. Определение этапов изменения зоноэдров классов в процессе обучения.
3. Интерпретация процесса переобучения, исходя только из геометрических характеристик зоноэдров (т. е. имея инофрмацию только о поведении параметров ИНС на обучающей выборке).
4. Обоснование выдвигаемых гипотез о динамике характеристик в процессе обучения на модельных экспериментах.

Введём следующие обозначения:

- Зоноэдры двух классов, для упрощения, обоначим за G_1 и G_2.
- Зоноэдры классов на итерации обучения ИНС с номером t за $G_1(t)$ и $G_2(t)$.
- Множество вершин зоноэдра G обозначим за $V(G)$.

Обсудим теоретические предпосылки анализа переобучения ИНС с помощью тропической геометрии и выдвенем ряд гипотез для экспериментальной проверки.

Переобучение модели, неформально, означает настройку модели на обучающую выборку. Более формально, пусть наша модель выбирается из фиксированного класса $a(\cdot) \in \mathscr{A}$, и её качество измеряется значением эмпирического риска на обучающей выборке $\{(x_i, y_i)\}_{i=1}^N$ по формуле

$$\mathscr{L}\left(a(\cdot)|\{(x_i,y_i)\}_{i=1}^N\right) = \frac{1}{N}\sum_{i=1}^{N} l(a(x_i), y_i),$$

где $l(\cdot)$ —функция потерь. Тогда можно выделить следующие случаи, характеризующие класс моделей:

1. Если класс моделей $\mathscr{A}$ слишком мал, то все модели из этого класса могут *недообучиться* (т. е. будут иметь большое значение эмпирического риска) и выдавать слабые прогнозы на новых данных.
2. Если класс моделей $\mathscr{A}$ слишком широк, то все модели из этого класса могут *переобучиться* (т. е. будут иметь малое значение эмпирического риска и большое значение истинного риска) и также выдавать прогнозы с низким качеством.

Данное определение является достаточно общепринятым и исследовалось как в контексте баланса между смещением и дисперсией выбираемых параметрических моделей [6], [8], так и со стороны теоретических оценок сложности моделей [15].

Основными инструментами обнаружения переобучения является наблюдение за поведением эмпирического риска на обучающей выборке и его значения на валидационной выборке. Тогда, исходя из определений, момент переобучения определяется в точке начала роста эмпирического риска на отложенной выборке. С развитием моделей ИНС и получением большого объёма данных для обучения данный подход показал некоторые неожиданные эффекты — двойной спуск (double descent) [3] и гроккинг (grokking) [13].

Для анализа момента переобучения оперируют термином *обобщающей способности* модели, отражающим баланс между смещением и дисперсией. Формальнее, говорят, что алгоритм обучения обладает способностью к обобщению, если вероятность ошибки на тестовой выборке достаточно мала или хотя бы предсказуема, то есть не сильно отличается от ошибки на обучающей выборке.

Предлагается интерпретировать обобщающую способность модели ИНС, определяя структуру двойственного подразделения параметров ИНС на обучающей выборке. Определим набор характеристик, которые могут говорить о моменте переобучения.

Будем использовать

1. Объём/площадь зоноэдров классов (везде далее $V(G_1)$, $V(G_2)$).
2. Объём/площадь выпуклой оболочки зоноэдров (т. е. объём/площадь двойственного подразделения) $V\left(ConvHull(G_1,G_2)\right) = V(\delta(R(\vec{x})))$.
3. Расстояние Хаусдорфа между множествами вершин зоноэдров $\mathscr{H}(V(G_1),V(G_2))$.

Объёмные характеристики зоноэдров классов призваны показать изменение обобщающей способности на обучающей выборке. Значение объёма/площади двойственного подразделения характеризует наличие пространства между двумя зоноэдрами, в котором потенциально возможно добавление объектов с неоднозначным разделением по классам.

Для разреживания обученных ИНС (процедуры pruning) методами тропической геометрии допустимо использование расстояния Хаусдорфа не между всеми точками многогранника, а только между вершинами, образующими этот многогранник [12].

$$\mathscr{H}(G_1,G_2) = \mathscr{H}(V(G_1),V(G_2)) =$$

$$= \max\left\{\max_{v\in V(G_1)}\rho(v,V(G_2)),\ \max_{u\in V(G_2)}\rho(V(G_1),u)\right\} \qquad (11)$$

Расстояние между двумя зоноэдрами должно определять динамику их расположения относительно друг друга. Обозначим за $\mathscr{H}_t$ расстояние между двумя зоноэдрами в итерацию обучения ИНС с номером t, а $\Delta\mathscr{H}_t = \mathscr{H}_t - \mathscr{H}_{t-1}$. Тогда сформулируем основную гипотезу работы следующим образом.

Утверждение 1. *Момент начала переобучения находится в окрестности итерации с номером k для которой для некоторого $k_0 < k$ одновременно выполнены условия:*

1. $V(G_i(k)) \approx V(G_i(k-1)) \approx \cdots \approx V(G_i(k-k_0)) \approx const, \quad i = 1,2$. *T.e. значение объёмов зоноэдров классов на итерации k остановились на постоянном уровне, начиная с итерации k_0.*
2. $\Delta\mathscr{H}_k \approx \Delta\mathscr{H}_{k-1} \approx \cdots \approx \Delta\mathscr{H}_{k-k_0+1} \approx const$. *T.e. на протяжении последних k_0 итераций вершины зоноэдров не перемещаются относительно вершин другого зоноэдра.*

Вычислительные эксперименты

Продемонстрируем, на сколько верна гипотеза 1 на модельном примере. Для начала, покажем, как на вид зоноэдров классов обученной

ИНС влияет начальная инициализация. Затем смоделируем эксперимент с переобучением и различными видами поведения исследуемых метрик без переобучения. Во всех экспериментах фиксирована архитектура ИНС: полносвязная ИНС с одним скрытым слоем размера P и $ReLU()$ активацией, на входе координаты плоскости, выходы соответствуют двум меткам классов. В качестве оптимизируемой функции потерь взята кросс-энтропия (обозначена за *Cross-Entropy Loss*). Оптимизатор для обучения ИНС — Adam. Для исследования переобучения не использовалась регуляризация (поскольку она призвана решить данную проблему). В визуализации конечного вида зоноэдров при заданной инициализации регуляризация использовалась.

Также фиксированы типы генерируемых наборов данных (рис. 6).

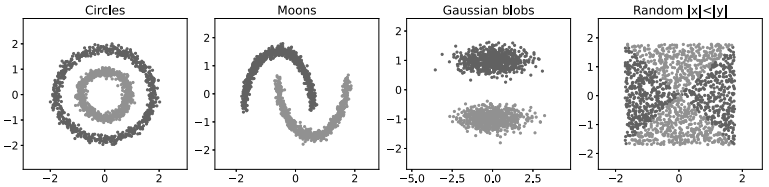


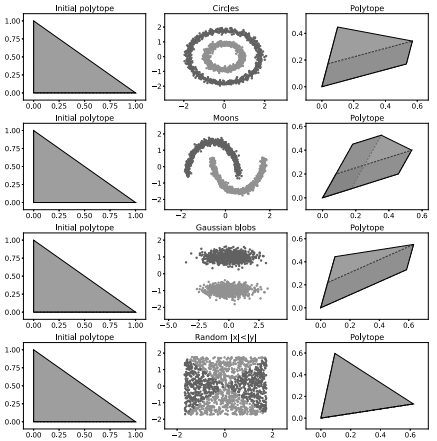
Рис. 6. Типы наборов данных для экспериментов

Зависимость от инициализации. Провизуализируем простейшие способы инициализации — тождественную и случайную (рис. 7). На рисунке 7a слева изображены начальные зоноэдры (*Initial polytope*), выборки, на которых обучалась ИНС, и итоговый вид зоноэдров после обучения *до одного и того же номера итерации* (*Polytope*). Контуры отвечают за зоноэдры классов G_1, G_2, их выпуклая оболочка (двойственное подразделение $\delta(R(\vec{x})) = ConvHull(G_1, G_2)$) отображена заполненной областью.

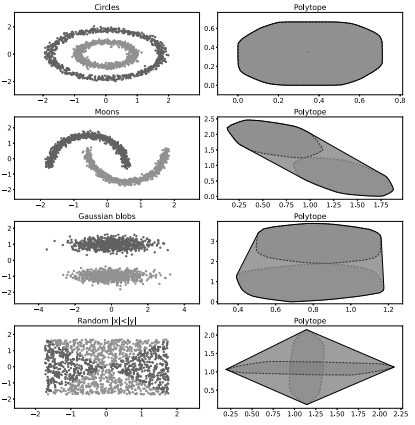
Тождественная инициализация очень сильно ограничивает сложность модели и конечный вид зоноэдров. На приведённых данных с тождественной инициализацей сложно построить разделяющую поверхность, что и отражает конечный вид зоноэдров.

На рис. 7b изображены только исходные выборки и результат после обучения. Начальная инициализация в этом случае является двумя эллипсами. Видна общая тенденция зоноэдров к выделению областей представителей классов. Особенно выделяется случай выборки *Random* $|x| < |y|$, поскольку, при дальнейшем обучении, объёмы зоноэдров классов будут уменьшаться, однако объём выпуклой оболочки будет сохраняться.

Визуализация поведения зоноэдров для более распространённых способов инициализации — Ксавье и Кайминга изображена на рис. 8. Оптимизация также проводилась до одинакового числа итераций. Для



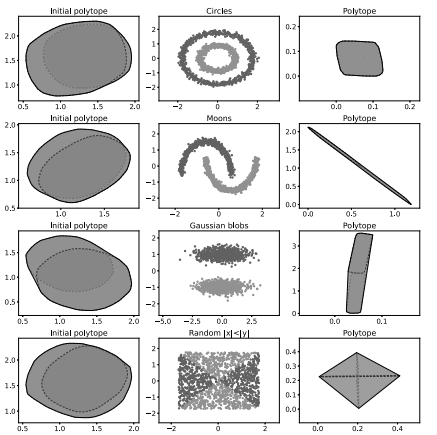
(a) Тождественная



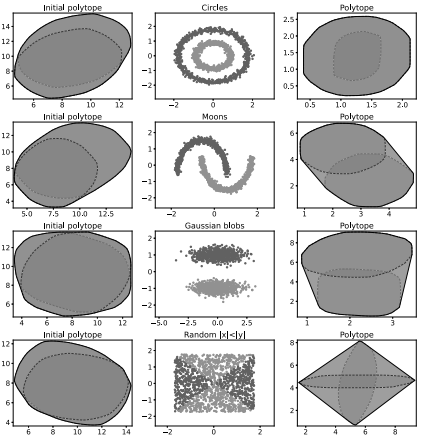
(b) Случайная

Рис. 7. Вид обученных зоноэдров (polytopes) в зависимости от инициализации

инициализации Ксавье (рис. 8a) заметна настройка на обучающую выборку (говорящая о переобучении). Инициализация Кайминга (рис. 8b) показывает качественную разделяющую способность без явного переобучения.



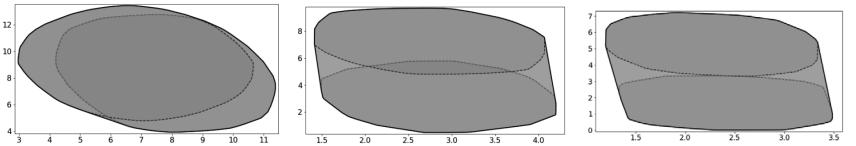
(a) Ксавье



(b) Кайминга

Рис. 8. Вид обученных зоноэдров (polytopes) в зависимости от инициализации

Модельный эксперимент на наиболее простой для инициализации Кайминга выборке — *Gaussian Blobs* воспроизводит момент переобучения на изображении зоноэдров (рис. 9). На итерации 2151 (рис. 9b) функция потерь перестала убывать (и на обучении, и на валидации). Последующие итерации подстраивают структуру многогранников только на обучающую выборку и соответствуют переобучению.



(a) Итерация 0 (b) Итерация 2151 (c) Итерация 3300

Рис. 9. Изображение момента переобучения с инициализацией Кайминга

Анализ переобучения. Уменьшим размер выборки до 150 точек и увеличим дисперсию генерируемых координат в наборах *Moons*, *Circles*, *Gaussian Blobs*. Полученные выборки разделим на обучение (*Train*) и валидацию (*Validation*) в соотношении 100 точек к 50 (рис. 10). Данные ограничения (и отсутствие регуляризации при оптимизации) позволили смоделировать переобучение на наборах данных *Moons* и *Circles* у архитектуры с размером внутреннего слоя $P = 500$.

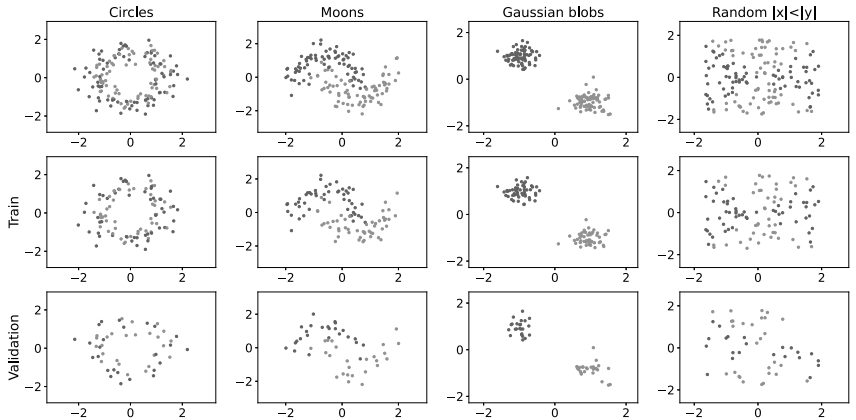
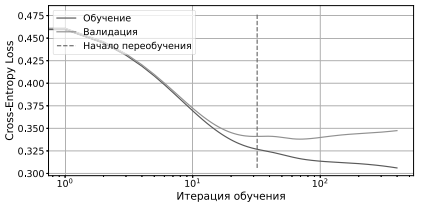


Рис. 10. Вид данных для моделирования переобучения

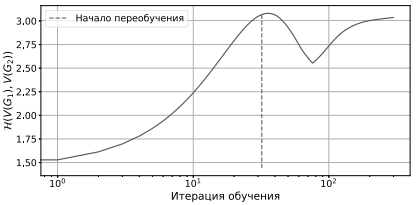
Модель на каждой выборке обучалась 400 итераций, момент начала переобучения определялся по первой итерации возрастания функции потерь на валидационной выборке.

На рис. 11 приведены результаты визуализации момента переобучения на наборе данных *Moons*. Классический критерий начала переобучения показывает первое увеличение функции потерь на валидационной выборке до начала явного её возрастания (рис. 11a). Динамика расстояния Хаусдорфа (рис. 11b) меняет характер именно на моменте срабатывания критерия начала переобучения.

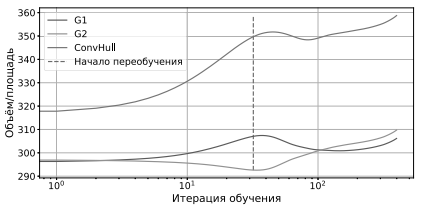
Изменение расстояния Хаусдорфа между вершинами зоноэдров останавливается на некотором константном уровне именно в момент начала явного роста функции потерь на валидационной выборке (рис. 11d). Также объёмы зоноэдров и их выпуклой оболочки (рис. 11c) останавливаются на некотором (близком к константному) уровне. Что согласуется с гипотезой 1.



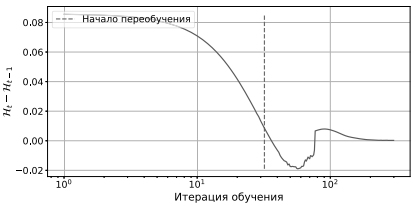
(a) Функции потерь



(b) Расстояние Хаусдорфа



(c) Площади зоноэдров



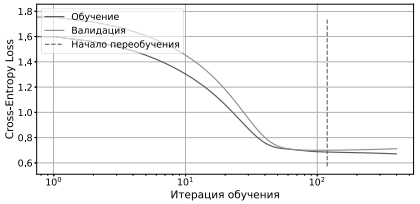
(d) Изменение расстояния Хаусдорфа

Рис. 11. Визуализация момента переобучения на наборе *Moons*

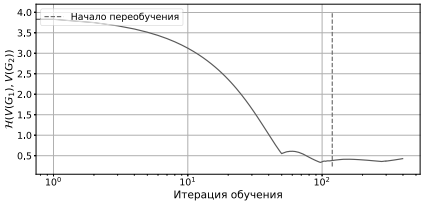
На рис. 12 показаны схожие результаты, согласующиеся с гипотезой 1, для набора данных *Circles*.

На наборах *Gaussian Blobs* и *Random* $|x| < |y|$ показаны результаты без переобучения. Модель на наборе *Gaussian Blobs* (рис. 13) явно отображает качественное обучение по функции потерь (рис. 13a). Изменение расстояния Хаусдорфа (рис. 13d) и площади зоноэдров показывают близкое к константному поведение в начале обучения, что свидетельствует об удачной инициализации параметров, согласующейся с обучающей выборкой.

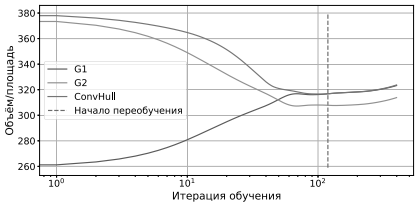
Данный факт накладывает ограничение на применимость гипотезы 1 — *при начале обучения на константной динамике*



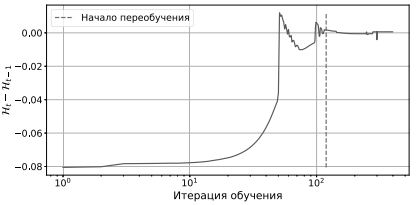
(a) Функции потерь



(b) Расстояние Хаусдорфа



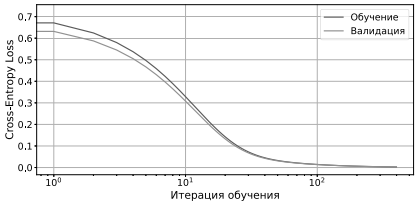
(c) Площади зоноэдров



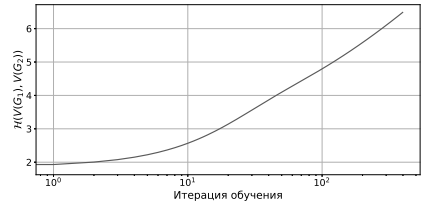
(d) Изменение расстояния Хаусдорфа

Рис. 12. Визуализация момента переобучения на наборе *Circles*

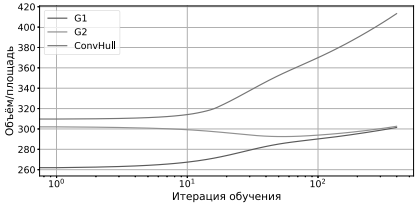
характеристик данное предположение не применимо.



(a) Функции потерь



(b) Расстояние Хаусдорфа

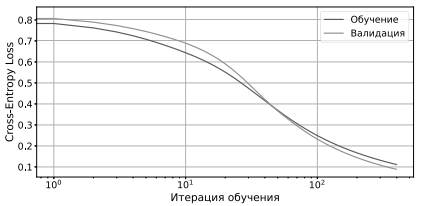


(c) Площади зоноэдров

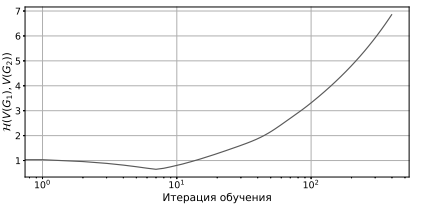
(d) Изменение расстояния Хаусдорфа

Рис. 13. Визуализация момента переобучения на наборе *Gaussian Blobs*

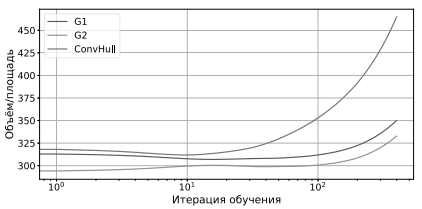
На наборе данных *Normal* $|x| < |y|$ (рис. 14) модель также не показывает переобучения, однако на графике функции потерь (рис. 14a) присутствует момент максимального сближения функций потерь на обучающей выборке и валидационной с дальнейшим более быстрым убыванием валидационной функции потерь. Данный факт не говорит о необходимости завершения обучения, позволяя оптимизировать модель далее. На графиках изменения расстояния Хаусдорфа и площадей зоноэдров (рис. 14d и 14c соответственно) также отображено начало оптимизации с константной характеристикой, затем происходит резкое возрастание объёмов и расстояния между зоноэдрами.



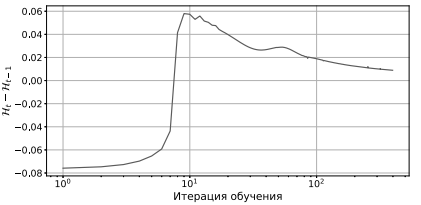
(a) Функции потерь



(b) Расстояние Хаусдорфа



(c) Площади зоноэдров



(d) Изменение расстояния Хаусдорфа

Рис. 14. Визуализация момента переобучения на наборе *Random* $|x| < |y|$

Обсуждение и выводы

Проведённые эксперименты показывают согласованность с гипотезой на модельных примерах при отсутствии константной динамики характеристик зоноэдров в начале оптимизации.

Стоит отметить, что для формирования однозначного вывода о применимости теории обнаружения переобучения ИНС на основе тропической геометрии без знаний о валидационной выборке требуется проведение экмпериментов с реальными данными. В пределах модельного эксперимента найдены подтверждения гипотезы 1.

Ключевым преимуществом предлагаемого подхода относительно известных является анализ только структуры параметров ИНС на обучающей выборке, что позволяет анализировать сложность модели и

делать вывод о переобучении без отделения от выборки валидационной части.

В текущей постановке применимость продемонстрирована на двумерных примерах с отделением случая инициализации, согласующейся с характером данных. Основными ограничениями на масштабируемость метода являются

1. Скорость построения зоноэдров. В реализации, написанной в данной работе, сложность построения квадратична по числу сегментов (из-за вычисления соответствующей суммы Минковского). В работе [1] обсуждается метод ускорения построения зоноэдра с использованием матрицы-генератора сегментов, однако приводится только теоретическая возможность улучшения алгоритма, поскольку число вершин зоноэдра ассимптотически линейно зависит от числа сегментов.

2. Необходимость формализации алгоритма обнаружения переобучения в общем случае.

Для проведения экспериментов, изложенных в работе, создан программный модуль визуализации тропических полиномов и анализа динамики зоноэдров параметров ИНС с использованием библиотеки pytope[1] языка программирования python.

Ключевым результатом работы является предлагаемая постановка задачи анализа переобучения модели ИНС по двойственному подразделению тропического полинома. Применимость данного подхода продемонстрирована на модельных наборах данных *Moons*, *Circles*, *Gaussian Blobs*, *Random* $|x| < |y|$.

С точки зрения практической применимости, предлагаемая постановка позволяет исследовать процесс переобучения модели ИНС на основе геометрии двойственных подразделений. Это даёт возможность как создания критерия остановки обучения, так и анализа более сложных эффектов, связанных с переобучением.

Открытыми видятся проблемы

1. Построения формальной теории обнаружения переобучения по двойственному подразделению тропического полинома параметров ИНС в общем случае.

2. Проведения экспериментов на немодельных данных.

3. Исследования более сложных процессов — двойного спуска (double descent) [3] и гроккинга (grokking) [13] на моделях ИНС с ограничениями на функции активации.

[1] https://pypi.org/project/pytope/

Литература

1. *Alfarra M.* Applications of Tropical Geometry in Deep Neural Networks // MS Thesis, 2020. KAUST Research Repository http://hdl.handle.net/10754/662473.
2. *Alfarra M., Bibi A., Hammoud H., Gaafar M., Ghanem B.* On the Decision Boundaries of Deep Neural Networks: A Tropical Geometry Perspective // CoRR. 2020. Vol. abs/2002.08838. https://arxiv.org/abs/2002.08838.
3. *Belkin M., Hsu D., Ma S., Mandal S.* Reconciling modern machine learning practice and the bias-variance trade-off // arXiv. 2018. https://arxiv.org/abs/1812.11118.
4. *Charisopoulos V., Maragos P.* A Tropical Approach to Neural Networks with Piecewise Linear Activations // arXiv. 2018. https://arxiv.org/abs/1805.08749.
5. *Fukushima K.* Visual Feature Extraction by a Multilayered Network of Analog Threshold Elements // IEEE Transactions on Systems Science and Cybernetics. 1969. Vol. 5, no.4. —Pp. 322-333. https://doi.org/10.1109/TSSC.1969.300225.
6. *Geman S., Bienenstock E., Doursat R.* Neural Networks and the Bias/Variance Dilemma // Neural Computation. 1992. Vol. 4, no.1. —Pp. 1-58. https://doi.org/10.1162/neco.1992.4.1.1.
7. *Glorot X., Bordes A., Bengio Y.* Deep Sparse Rectifier Neural Networks // Journal of Machine Learning Research. 2010. Vol. 15.
8. *Hastie T., Tibshirani R., Friedman J., Franklin J.* The Elements of Statistical Learning: Data Mining, Inference, and Prediction // Math. Intell. 2004. Vol. 27. —Pp.83-85. https://doi.org/10.1007/BF02985802.
9. *Itenberg I., Mikhalkin G., Shustin E.* Tropical Algebraic Geometry // Birkhauser Basel. 2010. Vol. IX, 104. https://doi.org/10.1007/978-3-0346-0048-4.
10. *Maclagan D., Sturmfels B.* Introduction to Tropical Geometry // Graduate Studies in Mathematics. 2015. Vol. 161.
11. *Maragos P., Charisopoulos V., Theodosis E.* Tropical Geometry and Machine Learning // Proceedings of the IEEE. 2021. Vol. 109, no.5. Pp.728-755. https://doi.org/10.1109/JPROC.2021.3065238.
12. *Misiakos P., Smyrnis G., Retsinas G., Maragos P.* Neural Network Approximation based on Hausdorff distance of Tropical Zonotopes //

International Conference on Learning Representations. 2022. https://openreview.net/forum?id=oiZJwC_fyS.

13. *Power A., Burda Y., Edwards H., Babuschkin I., Misra V.* Grokking: Generalization beyond overfitting on small algorithmic datasets // ICLR MATH-AI Workshop. 2021.

14. *Smyrnis G., Maragos P., Retsinas G.* Maxpolynomial Division with Application To Neural Network Simplification // ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2020. Pp.4192-4196. https://doi.org/10.1109/ICASSP40776.2020.9053540.

15. *Vapnik V. N.* Statistical learning theory // New York. Wiley. 1998.

16. *Zhang L., Naitzat G., Lim L.-H.* Tropical Geometry of Deep Neural Networks //CoRR. 2018. Vol. abs/1805.07091. https://dblp.org/rec/journals/corr/abs-1805-07091.bib.

17. *Кормаков Г. В.* Динамика двумерной тропической геометрии структур параметров искусственных нейронных сетей // Сборник тезисов международной конференции студентов, аспирантов и молодых учёных «Ломоносов-2022». 2022.