

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теоретических основ  
компьютерной безопасности и  
криптографии

**Основные теоремы нейроинформатики.**

**РЕФЕРАТ**

студента 5 курса 531 группы  
специальности 10.05.01 Компьютерная безопасность  
факультета компьютерных наук и информационных технологий

Сметанкина Никиты Артуровича

Проверил

доцент

\_\_\_\_\_

подпись, дата

И. И. Слеповичев

Саратов 2024

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	3
1 Фундаментальные теоремы.....	5
1.1 Теорема Колмогорова .....	5
1.2 Теорема Вейерштрасса .....	13
1.3 Обобщенная теорема Стоуна .....	13
2 Ключевые теоремы .....	18
2.1 Точное представление многочленов от многих переменных.....	18
2.2 Теоремы о термах.....	21
2.3 Теорема о скрытых параметрах .....	29
2.4 Универсальная теорема аппроксимации .....	31
ЗАКЛЮЧЕНИЕ .....	35
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	36

## ВВЕДЕНИЕ

Современные нейронные сети представляют собой фундаментальный инструмент для моделирования и анализа сложных систем, вдохновлённых структурой и функциями биологических нервных систем. Они находят применение в самых разнообразных задачах, от диагностики заболеваний и обработки сигналов до анализа данных и управления роботами. Эффективность нейронных сетей основывается на их математических свойствах, описываемых рядом ключевых теорем, которые служат теоретическим фундаментом для понимания их возможностей и ограничений.

Основные теоремы этой области представляют собой теоретическую основу, на которой строятся современные методы моделирования, обучения и оптимизации. Они задают формальные границы возможностей нейронных сетей и помогают понять, какие классы задач такие модели способны решать.

Эти теоремы дают ответы на ключевые вопросы, лежащие в основе построения нейросетей, такие как:

1. Возможно ли точное представление функции многих переменных через суперпозицию функций меньшего числа переменных?
2. Можно ли с произвольной точностью аппроксимировать функции многих переменных с помощью более простых функций и операций?

Эти теоремы отражают фундаментальные аспекты работы нейронных сетей, включая их способность приближать сложные функции, ограничения в обучении и зависимость от структуры модели. Они также подчёркивают ключевые вызовы, такие как необходимость эффективных алгоритмов оптимизации и устойчивости решений. Теоретическая основа нейроинформатики не только проливает свет на природу искусственных систем, но и направляет дальнейшее развитие методов, способных справляться с неопределённостью и сложностью реальных данных.

Абстрактное понимание основных теорем нейроинформатики позволяет рассматривать эту область как связующее звено между математикой, информатикой и когнитивными науками. Оно служит фундаментом для разработки универсальных подходов к обработке информации и построению адаптивных систем, которые стремятся к оптимальности в условиях неопределённости.

## 1 Фундаментальные теоремы

Нейронные сети строят вычисления на основе линейных функций, нелинейных функций одной переменной и их всевозможных суперпозиций — функций, формируемых путём последовательного каскадного соединения сетевых элементов. Возникает естественный вопрос: какие классы функций можно точно выразить с помощью таких операций, а какие функции поддаются сколь угодно точной аппроксимации?

Ответы на эти вопросы основаны на фундаментальных теоремах. Теорема Колмогорова о представлении непрерывных функций многих переменных с помощью суперпозиций функций одной переменной и линейных функций задаёт формальные границы выразительных возможностей нейросетей. Обобщённая аппроксимационная теорема Стоуна расширяет это понимание, утверждая, что для произвольных алгебр функций можно построить аппроксимации с заданной степенью точности.

Эти теоремы лежат в основе нейроинформатики, объясняя, как нейронные сети формируют свои вычислительные возможности, и указывают направления для их дальнейшего совершенствования. [7]

### 1.1 Теорема Колмогорова

Функции могут быть заданы не только аналитическими выражениями, но и через уравнения, решения которых позволяют значительно расширить класс вычисляемых функций. Если принять за основу решения ряда простых уравнений, то более сложные задачи могут быть решены посредством их комбинации. Классическим примером является использование радикалов, которые позволяют выразить решения уравнений 2-й, 3-й и 4-й степеней. Например, решение квадратного уравнения  $ax^2 + bx + c = 0$  можно записать через операции сложения, умножения, деления и функцию квадратного корня, что существенно упрощает вычисления.

Однако вопрос об универсальности такого подхода был глубоко изучен в математике. Исследования Абеля и Галуа показали, что уравнения 5-й степени и

выше неразрешимы в радикалах, что потребовало разработки новых подходов к построению решений. Возникла идея расширить набор базовых функций, используя более сложные функции небольшого числа переменных. Этот подход привёл к важной математической проблеме, сформулированной Давидом Гильбертом в начале XX века. Его 13-я проблема задаёт вопрос о том, можно ли представить корень уравнения  $x^7 + ax^3 + bx^2 + cx + 1 = 0$  как суперпозицию непрерывных функций двух переменных.

Впоследствии оказалось, что для уравнений 5-й и 6-й степени это действительно возможно не только для непрерывных, но даже для аналитических функций. Однако дальнейшее развитие этой идеи потребовало ещё большего обобщения. Полезным оказалось отойти от конкретных уравнений и поставить общий вопрос: может ли любая непрерывная функция  $n$  переменных быть представлена через операции сложения, умножения и суперпозиции непрерывных функций двух переменных?

Ответ на этот вопрос был найден А. Н. Колмогоровым. Он доказал, что произвольная непрерывная функция любого числа переменных действительно может быть представлена в такой форме. Это открытие стало важным шагом в математике и теории функций, задав направление для дальнейших исследований в области функционального анализа и теории аппроксимации.

А. Н. Колмогоров, затем и В. И. Арнольд дали окончательное решение поставленной проблемы. Они доказали, что любую непрерывную функцию  $n$  переменных можно представить с использованием операций сложения, умножения и суперпозиции непрерывных функций всего одной переменной. Этот результат не только подтвердил возможность такой универсальной аппроксимации, но и обозначил новый подход к пониманию выразительных возможностей математических моделей.

Каждая непрерывная функция  $n$  переменных, заданная на единичном кубе  $n$ -мерного пространства, представима в

$$f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} h_q \left[ \sum_{p=1}^n \varphi_q^p(x_p) \right] \quad (1)$$

где функции  $h_q(u)$  непрерывны, а функции  $\varphi_q^p(x_p)$ , кроме того, еще и стандартны, т.е. не зависят от выбора функции  $f$ . [2]

В частности, каждая непрерывная функция двух переменных  $x, y$  представима в виде

$$f(x, y) = \sum_{q=1}^5 h_q [\varphi_q^p(x) + \psi_q(y)] \quad (2)$$

Доказательство приведем для случая  $n = 2$ , следуя подходу, предложенному В. И. Арнольда. Возможность представления в форме (2) обосновывается поэтапно.

1. Функции  $\psi_q(x)$  и  $\psi_q(y)$  у представления (2) совершенно не зависят от разлагаемой функции  $f(x, y)$ .

Для определения этих функций нам потребуется предварительно выполнить несколько построений. Представим себе "город", как показано на рисунке 1. Это структура, состоящая из одинаковых "кварталов" (непересекающихся замкнутых квадратов), разделенных узкими "улицами" постоянной ширины.

Уменьшим этот "город" гомотетично в  $N$  раз, выбрав точку  $A_1$  в качестве центра гомотетии. В результате получим уменьшенную копию, которую будем называть "городом ранга 2". Аналогично, "город ранга 3" строится из "города ранга 2" путем его гомотетичного уменьшения с коэффициентом  $\frac{1}{N}$ . Далее, "город ранга 4" получается гомотетичным уменьшением "города ранга 3" в  $N$  раз, и так далее. В общем случае "город ранга  $k$ " получается из исходного "города" (называемого "городом первого ранга") гомотетичным уменьшением с коэффициентом  $N^k$ . При этом центр гомотетии, выбранный в  $A_1$  не играет принципиальной роли для дальнейших построений.

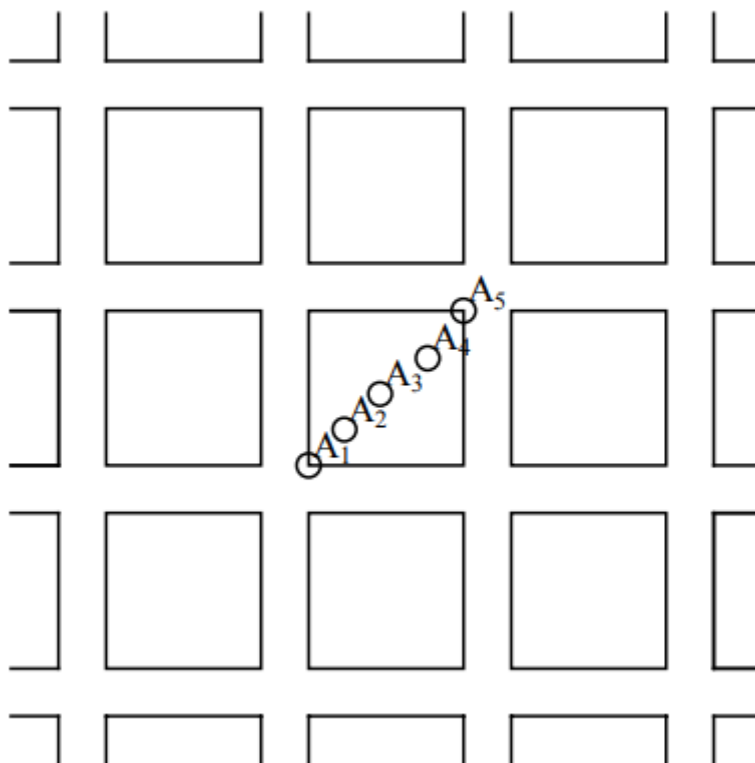


Рисунок 1 – Система кварталов

Построенную систему "городов" будем называть **1-й системой**. "Город первого ранга"  $q$ -й системы ( $q = 2, \dots, 5$ ) получается из исходного "города" (рис. 1) путем параллельного переноса, совмещающего точку  $A_1$  с точкой  $A_q$ .

Легко заметить, что можно выбрать ширину "улиц" в "городе" настолько малой, что каждая точка плоскости будет находиться в пределах как минимум трех "кварталов" пяти различных "городов первого ранга".

Аналогичным образом, "город  $k$ -го ранга"  $q$ -й системы ( $k = 2, 3, \dots; q = 2, \dots, 5$ ) получается из "города  $k$ -го ранга" 1-й системы с помощью параллельного переноса, совмещающего точку  $A_1^k$  с точкой  $A_q^k$ . Точки  $A_1^k$  и  $A_q^k$  представляют собой образы точек  $A_1$  и  $A_q$  полученные с помощью гомотетии, переводящей "город первого ранга" 1-й системы в "город  $k$ -го ранга" той же системы.

При этом каждая точка плоскости будет принадлежать кварталам как минимум трех из пяти "городов" любого фиксированного ранга  $k$ .



Функцию  $\phi_q(x, y) = \varphi_q(x) + \psi_q(y)$  ( $q = 1, 2, \dots, 5$ ) определим таким образом, чтобы она разделяла любые два "квартала" в каждом "городе"  $q$ -й системы. Это означает, что множество значений функции  $\phi_q(x, y)$  принимаемых на любом "квартале" "города  $k$ -го ранга"  $q$ -й системы (где  $k$  – произвольное фиксированное число), не должно пересекаться с множеством значений, принимаемых  $\phi_q(x, y)$  на любом другом "квартале" того же "города". Для этого достаточно рассматривать функцию  $\phi_q(x, y)$  только на единичном квадрате, а не на всей плоскости.

Для того, чтобы функция  $\phi_q(x, y) = \varphi_q(x) + \psi_q(y)$  разделяла "кварталы" "города первого ранга", можно наложить следующие условия. Пусть  $\varphi_q(x)$  на проекциях "кварталов" "города" на ось  $x$  принимает значения, которые мало отличаются от различных целых чисел, а  $\psi_q(y)$  на проекциях "кварталов" на ось  $y$  принимает значения, которые мало отличаются от различных кратных  $\sqrt{2}$ . Последнее условие использует тот факт, что  $m + n\sqrt{2} = m' + n'\sqrt{2}$  при целых  $m, n, m', n'$ , выполняется только тогда, когда  $m = m', n = n'$ . При этом наложенные условия не определяют пока еще, разумеется, функций  $\varphi_q(x)$  и  $\psi_q(y)$  (на "улицах" функция  $\phi_q(x, y) = \varphi_q(x) + \psi_q(y)$  вообще пока может задаваться совершенно произвольно).

Используя это, можно подобрать границы значений  $\varphi_q(x)$  и  $\psi_q(y)$  на "кварталах" "города второго ранга" так, чтобы функция  $\phi_q(x, y) = \varphi_q(x) + \psi_q(y)$  разделяла не только "кварталы" "города 1-го ранга", но и "кварталы" "города 2-го ранга".

Программа, описанная выше, может быть реализована, если выбрать  $N$  большим, чтобы "кварталы" более высоких рангов не пересекались с "кварталами" предыдущих. А. Н. Колмогоров выбрал  $N = 18$ . Привлекая подобным же образом к рассмотрению "города" последующих рангов и уточняя каждый раз значения функций  $\varphi_q(x)$  и  $\psi_q(y)$ , мы в пределе получим

непрерывные функции  $\varphi_q(x)$  и  $\psi_q(y)$ , удовлетворяющие поставленным условиям, которые можно дополнительно потребовать монотонными.

2. Функции  $h_q(u)$  разложения (2), напротив того, существенно зависят от исходной функции  $f(x, y)$ .

Для построения этих функций докажем прежде всего, что любую непрерывную функцию  $f(x, y)$  двух переменных  $x$  и  $y$ , заданную на единичном квадрате, можно представить в виде

$$f(x, y) = \sum_{q=1}^5 h_q^{(1)}(\phi_q(x, y)) + f_1(x, y) \quad (3)$$

где  $\phi_q(x, y) = \varphi_q(x) + \psi_q(y)$  – функции, построенные выше, и

$$M_1 = \max|f_1(x, y)| \leq \frac{5}{6} \max|f(x, y)| = \frac{5}{6} M, \quad (3a)$$

$$\max|h_q^{(1)}(\phi_q(x, y))| = \frac{1}{3} M, \quad q = 1, \dots, 5. \quad (3б)$$

Выберем ранг  $k$  настолько большим, чтобы колебание функции  $f(x, y)$  на каждом "квартале" любого из "городов ранга  $k$ " не превосходило  $\frac{1}{6} M$ . Это, разумеется, возможно, так как с ростом ранга  $k$  размеры "кварталов" уменьшаются неограниченно. Далее, пусть  $p_1^{(ij)}$  – определенный "квартал" "города 1-й системы" и выбранного ранга  $k$ . В таком случае непрерывная функция  $\phi_1(x, y)$  принимает на этом "квартале" значения, принадлежащие определенному сегменту  $\Delta_1^{(ij)}$  числовой оси. В силу свойств функции  $\phi_1$  этот сегмент не пересекается с сегментами значений, принимаемых  $\phi_1$  на всех других "кварталах". Положим теперь функцию  $h_1^{(1)}$  на сегменте  $\Delta_1^{(ij)}$  постоянной, равной  $\frac{1}{3}$  значения, принимаемого функцией  $f(x, y)$  в какой-либо внутренней точке  $M_1^{(ij)}$  квартала  $p_1^{(ij)}$  (эту точку можно назвать "центром квартала"). Таким же образом мы определим функцию  $h_1^{(1)}$  на любом другом из сегментов, задаваемых значениями функции  $\phi_1(x, y)$  на "кварталах" "города  $k$ -го ранга" 1-й системы; при этом все значения  $h_1^{(1)}$  будут по модулю не превосходить  $\frac{1}{3} M$ , так как

значение  $f(x, y)$  в "центре" любого "квартала" по модулю не превосходит  $M$ .  
 Доопределим теперь функцию  $h_1^{(1)}(u)$  при тех значениях аргумента  $u$ , при каких она еще не определена, произвольно, с тем лишь, чтобы она была непрерывна, и чтобы выполнялось неравенство (3б). Совершенно аналогично определим и все остальные функции  $h_q^{(1)}(u) (q = 2, \dots, 5)$ .

Докажем теперь, что разность

$$f_1(x, y) = f(x, y) - \sum_{q=1}^5 h_q^{(1)}(\phi_q(x, y))$$

удовлетворяет условию (3а), т.е. что  $|f_1(x_0, y_0)| \leq \frac{5}{6}M$ , где  $(x_0, y_0)$  – произвольная точка единичного квадрата. Как и все точки плоскости эта точка принадлежит по крайней мере трем кварталам "городов ранга  $k$ ", поэтому заведомо найдутся такие три из пяти функций  $h_q^{(1)}(\phi_q(x, y))$ , которые принимают в точке  $(x_0, y_0)$  значение, равное  $\frac{1}{3}$  значения  $f(x, y)$  в "центре" соответствующего "квартала", т.е. отличающееся от  $\frac{1}{3}f(x_0, y_0)$  не более чем на  $\frac{1}{18}M$  (ибо колебание  $f(x, y)$  на каждом квартале не превосходит  $\frac{1}{6}M$ ); сумма этих трех значений  $h_q^{(1)}(\phi_q(x_0, y_0))$  будет отличаться от  $f(x_0, y_0)$  по модулю не более чем на  $\frac{1}{6}M$ . А так как каждое из оставшихся двух  $h_q^{(1)}(\phi_q(x_0, y_0))$  в силу (3) по модулю не превосходит  $\frac{1}{3}M$  то мы получаем:

$$|f_1(x_0, y_0)| = \left| f(x_0, y_0) - \sum_{q=1}^5 h_q^{(1)}(\phi_q(x_0, y_0)) \right| \leq \frac{1}{6}M + \frac{2}{3}M = \frac{5}{6}M,$$

что и доказывает (3а).

Применим теперь то же разложение (3) к входящей в (3) функции  $f_1(x, y)$ ; мы получим:

$$f_1(x, y) = \sum_{q=1}^5 h_q^{(2)}(\phi_q(x, y)) + f_2(x, y)$$

или

$$f(x, y) = \sum_{q=1}^5 h_q^{(1)}(\phi_q(x, y)) + \sum_{q=1}^5 h_q^{(2)}(\phi_q(x, y)) + f_2(x, y),$$

где

$$M_2 = \max|f_2(x, y)| \leq \frac{5}{6} M_1 = \left(\frac{5}{6}\right)^2 M$$

$$\text{и } \max|h_q^{(2)}(\phi_q(x, y))| \leq \frac{1}{3} M_1 \leq \frac{1}{3} * \frac{5}{6} M, (q = 1, \dots, 5).$$

Затем мы применим разложение (3) к полученной функции  $f_2(x, y)$  и т.д.; после  $n$ -кратного применения этого разложения мы будем иметь:

$$f(x, y) = \sum_{q=1}^5 h_q^{(1)}(\phi_q(x, y)) + \sum_{q=1}^5 h_q^{(2)}(\phi_q(x, y)) + \\ + f_2(x, y) + \dots + \sum_{q=1}^5 h_q^{(n-1)}(\phi_q(x, y)) + f_n(x, y),$$

где

$$M_n = \max|f_n(x, y)| \leq \left(\frac{5}{6}\right)^n M$$

$$\text{и } \max|h_q^{(s)}(\phi_q(x, y))| \leq \frac{1}{3} * \left(\frac{5}{6}\right)^{n-1} M, (q = 1, \dots, 5; s = 1, \dots, n-1).$$

Последние оценки показывают, что при  $n \rightarrow \infty$  получим:

$$f(x, y) = \sum_{q=1}^5 h_q^{(1)}(\phi_q(x, y)) \\ + \sum_{q=1}^5 h_q^{(2)}(\phi_q(x, y)) + f_2(x, y) + \dots + \sum_{q=1}^5 h_q^{(n)}(\phi_q(x, y)) + \dots$$

где стоящий справа бесконечный ряд сходится равномерно; также и каждый из пяти рядов

$$h_q^{(1)}(\phi_q(x, y)) + h_q^{(2)}(\phi_q(x, y)) + h_q^{(n)}(\phi_q(x, y)) + \dots \quad (q = 1, 2, \dots, 5)$$

сходится равномерно, что позволяет ввести обозначения

$$h_q(u) = h_q^{(1)} + h_q^{(2)} + h_q^{(n)} + \dots \quad (q = 1, 2, \dots, 5).$$

Итак, окончательно получаем:

$$f(x, y) = \sum_{q=1}^5 h_q(\phi_q(x, y)) = \sum_{q=1}^5 h_q[\varphi_q^p(x) + \psi_q(y)], \quad \text{то есть}$$

требуемое разложение (2). [1]

## 1.2 Теорема Вейерштрасса

До этого рассматривалось точное выражение функций нескольких переменных через функции одной переменной. Выяснилось, что в классе непрерывных функций такое представление осуществимо. Однако, помимо задачи точного выражения, существует еще одна важная задача — задача аппроксимации. Более того, можно предположить, что она имеет даже большее значение, поскольку большинство функций вычисляются приближенно, даже если для них существуют «точные» формулы.

Приближение функций многочленами и рациональными функциями имеет еще более долгую историю, чем проблема их точного представления. Одним из важнейших результатов в этой области является знаменитая теорема Вейерштрасса, которая утверждает, что любую непрерывную функцию нескольких переменных  $f(x_1, x_2, \dots, x_n)$  определенную на замкнутом ограниченном множестве  $Q$  можно равномерно приблизить последовательностью полиномов. Это значит, что для любого  $\varepsilon > 0$  существует такой многочлен  $P(x_1, x_2, \dots, x_n)$ , что  $\sup_Q |f(x_1, x_2, \dots, x_n) - P(x_1, x_2, \dots, x_n)| < \varepsilon$ . [3]

## 1.3 Обобщенная теорема Стоуна

Для того чтобы обобщить и усилить теорему Вейерштрасса, следует перейти к более абстрактному языку. Рассмотрим компактное пространство  $X$  и алгебру  $C(X)$  непрерывных функций на  $X$ , принимающих вещественные значения.

Сильным обобщением теоремы о равномерном приближении непрерывных функций многочленами является теорема Стоуна. Она утверждает следующее:

Пусть  $E \subseteq C(X)$  - замкнутая подалгебра в  $C(X)$ , такая что  $1 \in E$  и функции из  $E$  разделяют точки в  $X$  (то есть для любых различных  $x, y \in X$  существует такая функция  $g \in E$ , что  $g(x) \neq g(y)$ ). [4]

Тогда  $E = C(X)$ . Теорема Стоуна обобщает теорему Вейерштрасса по двум направлениям. Во-первых, она касается функций на произвольных компактах, а не только на множестве действительных чисел. Во-вторых, она доказывает новое утверждение, которое имеет значение даже для функций одного переменного, не говоря уже о многомерных функциях: оказывается, что множество многочленов от координатных функций — это не единственное плотное множество. В действительности, плотность характерна для кольца многочленов от любых функций, которые разделяют точки. Это значит, что плотными являются не только многочлены, но и такие объекты, как тригонометрические многочлены или линейные комбинации функций вида  $\exp[-(x - x_0, Q(x - x_0))]$ , где  $(x, Qx)$  - положительно определенная квадратичная форма и другие подобные функции.

Способ конструирования таких обобщений: нужно выбрать произвольный набор функций, которые разделяют точки, затем построить кольцо многочленов от этих функций — и получим плотное множество в  $C(X)$ .

Рассмотрим компактное пространство  $X$  и алгебру  $C(X)$  непрерывных функций на  $X$  с вещественными значениями.

Помимо приближения функций многочленами и их обобщениями через кольца функций, разделяющих точки, в последнее время все большее внимание уделяется приближению функций многих переменных с помощью линейных операций и суперпозиций функций одного переменного. Это приближение реализуется с помощью специальных формальных "устройств" — нейронных сетей. Каждая нейронная сеть состоит из формальных нейронов. Нейрон принимает на вход вектор сигналов  $x$ , вычисляет его скалярное произведение на вектор весов  $\alpha$  и применяет некоторую функцию одного переменного  $\phi(x, \alpha)$ . Результат передается на входы других нейронов или непосредственно на выход.

Таким образом, нейронные сети выполняют вычисления суперпозиций простых функций одного переменного и их линейных комбинаций.

Чтобы получить нужное обобщение, необходимо перейти от изучения колец функций к изучению их алгебр, замкнутых относительно некоторой нелинейной унарной операции.

Пусть  $E \subseteq C(X)$  - линейное пространство,  $C(R)$  - пространство непрерывных функций на действительной оси  $R$ ,  $f \in C(R)$  - нелинейная функция и для любого  $g \in E$  выполнено  $f(g) \in E$ . В этом случае будем говорить, что  $E$  замкнуто относительно нелинейной унарной операции  $f$ .

Очевидным примером является множество функций  $n$  переменных, которые можно точно представить, используя заданную функцию одного переменного и линейные функции. Это множество является линейным пространством, замкнутым относительно нелинейной унарной операции  $f$ .

**Замечание.** Линейное пространство  $E \subseteq C(X)$  замкнуто относительно нелинейной операции  $f(x) = x^2$  тогда и только тогда, когда  $E$  является кольцом. Действительно,  $fg = \frac{1}{2}[(f+g)^2 - f^2 - g^2]$  поэтому для линейного пространства  $E \subseteq C(X)$  замкнутость относительно унарной операции  $f(x) = x^2$  равносильна замкнутости относительно произведения функций.

Согласно приведенному замечанию, теорема Стоуна может быть переформулирована так.

Пусть  $E \subseteq C(X)$  - замкнутое линейное подпространство в  $C(X)$ ,  $1 \in E$ , функции из  $E$  разделяют точки в  $X$  и  $E$  замкнуто относительно нелинейной унарной операции  $f(x) = x^2$ . Тогда  $E = C(X)$ . [1]

Наше обобщение теоремы Стоуна состоит в замене  $f(x) = x^2$  на произвольную нелинейную непрерывную функцию.

**Теорема 1.** Пусть  $E \subseteq C(X)$  - замкнутое линейное подпространство в  $C(X)$ ,  $1 \in E$ , функции из  $E$  разделяют точки в  $X$  и  $E$  замкнуто относительно нелинейной унарной операции  $f \in C(R)$ . Тогда  $E = C(X)$ .

**Доказательство.** Рассмотрим множество всех таких  $p \in C(R)$ , что  $p(E) \subseteq E$ , то есть для любого  $g \in E$  выполнено:  $p(g) \in E$ . Обозначим это множество  $P_E$ . Оно обладает следующими свойствами:

- 1)  $P_E$  - полугруппа относительно суперпозиции функций;
  - 2)  $P_E$  - замкнутое линейное подпространство в  $C(R)$  (в топологии равномерной сходимости на компактах);
  - 3)  $1 \in P_E$  и  $id \in P_E$  ( $id(x) \equiv x$ ).
  - 4)  $P_E$  включает хоть одну непрерывную нелинейную функцию.
- Дальнейшее следует из теоремы 2, которая является, по существу, подготовительной теоремой о полугруппах функций.

**Теорема 2.** Пусть множество  $P \subseteq C(R)$  удовлетворяет условиям 1-4. Тогда  $P = C(R)$ .

**Доказательство** опирается на три леммы.

**Лемма 1.** В условиях теоремы 2 существует дважды непрерывно дифференцируемая функция  $g \in P$ , не являющаяся линейной.

**Доказательство.** Пусть  $v(x) \in C^\infty(R)$ ,  $v(x) = 0$  при  $|x| > 1$ ,  $\int_R v(x) dx = 1$ . Рассмотрим оператор осреднения  $J_\varepsilon f(x) = \int_R f(x+y) \frac{1}{\varepsilon} v(\frac{y}{\varepsilon}) dy$ . Для любого  $\varepsilon > 0$  выполнено:  $J_\varepsilon f(x) \in P$ .

Действительно  $f(x+y) \in E$  для каждого фиксированного  $y$  (т.к. константы принадлежат  $E$  и  $E$  замкнуто относительно линейных операций и суперпозиции функций). Интеграл  $J_\varepsilon f(x)$  принадлежит  $E$ , так как  $E$  является замкнутым линейным подпространством в  $C(R)$ , а этот интеграл - пределом конечных сумм.

Функция  $J_\varepsilon f(x)$  принадлежит  $C^\infty(R)$  так как

$$J_\varepsilon f(x) = \int_R f(x+y) \frac{1}{\varepsilon} v(\frac{y}{\varepsilon}) dy = \int_R f(z) \frac{1}{\varepsilon} v(\frac{z-x}{\varepsilon}) dy$$

(напомним, что  $v$  функция с компактным носителем).



Существует такое  $\varepsilon > 0$ , что функция  $g = J_\varepsilon f$  не является линейной, поскольку  $J_\varepsilon f \rightarrow f$  при  $\varepsilon \rightarrow 0$ , пространство линейных функций замкнуто, а  $f$  не является линейной функцией. Таким образом, в предположениях леммы существует нелинейная функция  $g \in P \cap C^\infty(R)$ , которую можно выбрать в вид  $g = J_\varepsilon f$

**Лемма 2.** Пусть в условиях теоремы 2 существует дважды непрерывно дифференцируемая функция  $g \in P$ , не являющаяся линейной. Тогда функция  $q(x) = x^2$  принадлежит  $P$ .

**Доказательство.** Существует точка  $x_0$ , для которой  $g''(x_0) \neq 0$ . Обозначим  $r(x) = 2(g(x + x_0) - g(x_0) - xg'(x_0))/g''(x_0)$ . Очевидно, что  $r \in P$ ,  $r(0) = 0$ ,  $r'(0) = 0$ ,  $r''(0) = 2$ ,  $r(x) = x^2 + o(x^2)$ . Поэтому  $r(\varepsilon x)/\varepsilon^2 \rightarrow x^2$  при  $\varepsilon \rightarrow 0$ .

Поскольку  $P$  замкнуто, получаем: функция  $q(x) = x^2$  принадлежит  $P$ .

**Лемма 3.** Пусть в условиях теоремы 2 функция  $q(x) = x^2$  принадлежит  $P$ . Тогда  $P$  является кольцом – для любых  $f, g \in P$  их произведение  $fg \in P$ .

**Доказательство.** Действительно,  $fg = \frac{1}{2}[(f + g)^2 - f^2 - g^2]$  и, так как  $P$  замкнуто относительно суперпозиции и линейных операций, то  $fg \in P$ . Доказательство теоремы 2 заканчивается обращением к классической теореме Вейерштрасса о приближении функций многочленами: из лемм 1-3 следует, что в условиях теоремы 2  $P$  является кольцом и, в частности, содержит все многочлены (которые получаются из 1 и  $id$  с помощью умножения и линейных операций). По теореме Вейерштрасса отсюда следует, что  $P = C(R)$ . Теоремы 1, 2 можно трактовать как утверждения о универсальных аппроксимационных свойствах любой нелинейности: с помощью линейных операций и каскадного соединения можно из произвольных нелинейных элементов получить любой требуемый результат с любой наперед заданной точностью. [1, 5]

## **2 Ключевые теоремы**

Далее рассмотрим теоремы, которые, хотя и не обладают такой масштабностью, как фундаментальные теоремы, но играют не менее важную роль в нейроинформатике. Эти теоремы позволяют глубже понять возможности и ограничения нейронных сетей и других методов машинного обучения, а также оптимизировать их использование для различных задач.

Такие теоремы охватывают более специфичные аспекты, хотя эти теоремы и являются частными результатами, их важность в контексте нейроинформатики также велика, поскольку они определяют не только теоретические границы возможностей нейронных сетей, но и конкретные способы их применения в реальных задачах.

### **2.1 Точное представление многочленов от многих переменных**

Точное представление многочленов от многих переменных с помощью одного произвольного многочлена от одного переменного, линейных операций и суперпозиции. Если такая полугруппа полиномов от одного переменного относительно суперпозиции содержит все многочлены первой степени и хотя бы один – более высокой, то она включает все многочлены. На основании этого факта доказано, что всегда возможно представить многочлен от многих переменных суперпозициями произвольного нелинейного многочлена от одного переменного и линейных функций. [1, 6]

Возвращаясь к классическому вопросу о представлении функций многих переменных с помощью функций меньшего числа переменных, следует отметить, что существует два основных направления для исследования:

1. Можно ли получить точное представление функции многих переменных с помощью суперпозиции функций меньшего числа переменных?

2. Можно ли получить сколь угодно точную аппроксимацию функции многих переменных с помощью некоторых более простых функций и операций?

В рамках первого вопроса особое внимание привлекают конструкции, в которых для точного представления всех функций многих переменных

используется один и тот же набор функций одного переменного. Традиционно считается, что для этого нужно использовать весьма специфические функции, как, например, в теореме Колмогорова, где применялись сильно негладкие функции.

С другой стороны, для второго вопроса (о приближении) выбор функций одного переменного для аппроксимации функций многих переменных значительно более свободен. В этом случае можно использовать практически любую нелинейную функцию, при этом достаточно одной такой функции для того, чтобы добиться нужной точности. Как показано в предыдущем разделе, для аппроксимации можно использовать почти любую нелинейную функцию, что значительно упрощает задачу. [1]

Далее доказываются теоремы, относящиеся к первому вопросу (точное представление). В частности, будет доказано, что любой многочлен от многих переменных можно точно представить с помощью суперпозиции произвольного нелинейного многочлена от одного переменного и линейных функций. Это подтверждает, что между первым и вторым вопросом нет принципиальной пропасти.

Пусть  $R[X]$  - кольцо многочленов от одного переменного над полем  $R$ ,  $E \subset R[X]$  - линейное пространство многочленов над  $R$ .

**Предложение 1.** Если  $E$  замкнуто относительно суперпозиции многочленов, содержит все многочлены первой степени и хотя бы один многочлен  $p(x)$  степени  $m > 1$ , то  $E = R[X]$ .

**Доказательство.** Заметим, что степень многочлена  $p'(x) = p(x + 1) - p(x)$  равна  $m - 1$ , и  $p'(x) \in E$ , так как  $E$  содержит многочлены первой степени (поэтому  $x + 1 \in E$ ), замкнуто относительно суперпозиции (поэтому  $p(x + 1) \in E$ ) и линейных операций (поэтому  $p'(x) \in E$ ).

Если  $m > 2$ , то понижаем степень с помощью конечных разностей (переходим к  $p'$ ,  $p''$  и т.д.), пока не получим многочлен второй степени. Вычитая из него линейную часть и умножая на константу, получаем:  $x^2 \in E$ . Поэтому для

любого  $f \in E$  имеем  $f^2 \in E$  (т.к.  $E$  - полугруппа). Дальнейшее очевидно: как неоднократно отмечалось выше, отсюда следует, что для любых  $f, g \in E$  их произведение  $fg \in E$  а с помощью умножения и сложения многочленов первой степени порождается все кольцо  $R[X]$ .

Перейдем к задаче представления многочленов от многих переменных. Обозначим  $R[X_1, \dots, X_n]$  кольцо многочленов от  $n$  переменных над полем  $R$ .

Для каждого многочлена от одного переменного введем множество тех многочленов, которые можно выразить с его помощью, используя суперпозиции и линейные функции. Пусть  $p$  – многочлен от одного переменного,  $E_p[X_1, \dots, X_n]$  – множество многочленов от  $n$  переменных, которое можно получить из  $p$  и многочленов первой степени, принадлежащих  $R[X_1, \dots, X_n]$ , с помощью операций суперпозиции, сложения и умножения на число.

Следующие два предложения дают удобную для дальнейшего характеристику  $E_p[X_1, \dots, X_n]$  и следуют непосредственно из определений.

**Предложение 2.** Множество  $E_p[X_1, \dots, X_n]$  является линейным пространством над  $R$  и для любого многочлена  $g(x_1, \dots, x_n)$  из  $E_p[X_1, \dots, X_n]$   $p(g(x_1, \dots, x_n)) \in E_p[X_1, \dots, X_n]$ .

**Предложение 3.** Для данного  $p$  семейство линейных подпространств  $L \subseteq R[X_1, \dots, X_n]$ , содержащих все многочлены первой степени и удовлетворяющих условию - если  $g(x_1, \dots, x_n) \in L$ , то  $p(g(x_1, \dots, x_n)) \in L$ , замкнуто относительно пересечений. Минимальным по включению элементом этого семейства является  $E_p[X_1, \dots, X_n]$ . Для любого линейного подпространства  $E \subseteq R[X_1, \dots, X_n]$  рассмотрим множество алгебраических унарных операций, которые переводят элементы  $E$  в элементы  $E$ :  $P_E = \{p \in R[X] \mid p(g(x_1, \dots, x_n)) \in E \text{ для любого } g(x_1, \dots, x_n) \in E\}$ .

**Предложение 4.** Для любого линейного подпространства  $E \subseteq R[X_1, \dots, X_n]$  множество полиномов  $P_E$  является линейным пространством над  $R$ , замкнуто

относительно суперпозиции и содержит все однородные многочлены первой степени.

Если линейное пространство  $E$  содержит 1, а  $P_E$  включает хотя бы один многочлен, степени  $m > 1$  (т.е. нелинейный), то  $P_E = R[X]$ .

**Доказательство.** Замкнутость  $P_E$  относительно суперпозиции следует из определения, все однородные полиномы первой степени входят в  $P_E$ , поскольку  $E$  является линейным пространством, отсюда также следует, что  $P_E$  является линейным пространством. Наконец, если  $1 \in E$  и  $P_E$  содержит многочлен степени  $m > 1$ , то  $1 \in P_E$ , тогда  $P_E = R[X]$  по предложению 1.

**Теорема 3.** Пусть  $E \subseteq R[X_1, \dots, X_n]$  – линейное подпространство,  $P_E$  содержит хотя бы один многочлен степени  $m > 1$  и  $1 \in E$ , тогда  $E$  является кольцом (с единицей).

**Доказательство.** По предложению 4 в условиях теоремы  $P_E = R[X]$ . В частности,  $x^2 \in P_E$ . Это означает, что для любого  $f \in E$  также и  $f^2 \in E$ . Поэтому для любых  $f, g \in E$  получаем:  $fg \in E$ .

**Теорема 4.** Для любого многочлена  $p$  степени  $m > 1$   $E_p[X_1, \dots, X_n] = R[X_1, \dots, X_n]$

**Доказательство.** Заметим, что  $E = E_p[X_1, \dots, X_n]$  – линейное подпространство в  $R[X_1, \dots, X_n]$ ,  $P_E$  содержит хотя бы один многочлен ( $p$ ) степени  $m > 1$  и  $E$  содержит все многочлены первой степени (и поэтому также 1). В силу теоремы 3,  $E$  является кольцом, а так как оно содержит все многочлены первой степени, то совпадает с  $R[X_1, \dots, X_n]$ , поскольку, используя умножение и сложение можно из этих многочленов получить любой. Таким образом, из  $p$  и многочленов первой степени с помощью операций суперпозиции, сложения и умножения на число можно получить все элементы  $R[X_1, \dots, X_n]$ . [1, 6]

## 2.2 Теоремы о термах

Сложная функция может быть задана с использованием суперпозиции простых функций, которые принадлежат заранее определенному конечному

множеству  $F$ . Эти простые функции определяются исключительно их принадлежностью к множеству  $F$ , и не требуются дополнительные специфические свойства, отличающие их от других функций в общем случае.

Теория выражений, определяющих сложные функции, является простейшим разделом математической логики, в которой сами эти выражения называются термами.

Термы — это правильно построенные выражения в некотором формальном языке. Чтобы задать такой язык, необходимо определить его алфавит. Он состоит из трех множеств символов:

- 1)  $C$  - множество символов, обозначающих константы;
- 2)  $V$  - множество символов, обозначающих переменные;
- 3)  $F$  - множество функциональных символов,  $F = \bigcup_{k=1}^{\infty} F_k$ , где  $F_k$  - множество символов для обозначения функций  $k$  переменных. [8]

Термы определяются индуктивно:

- 1) любой символ из  $C \cup V$  есть терм;
- 2) если  $t_1, \dots, t_k$  - термы и  $f \in F_k$ , то  $f t_1, \dots, t_k$  - терм.

Множество термов  $T$  представляет собой объединение:  $T = \bigcup_{i=1}^{\infty} T_i$ , где  $T_0 = C \cup V$ ,

$$T_{i+1} = \bigcup_{k=1}^{\infty} \bigcup_{f \in F_k} \{f t_1, \dots, t_k \mid t_1, \dots, t_k \in \bigcup_{j=0}^i T_j\} \quad (T_1 \subseteq T_2 \subseteq \dots \subseteq T_i \subseteq T_{i+1} \subseteq \dots)$$

Удобно разбить  $T$  на непересекающиеся множества — слои  $S_{i+1} = T_{i+1} \setminus T_i$  ( $S_0 = T_0$ ). Элементы  $S_i$  будем называть термами глубины  $i$  или  $i$ -слойными термами. Множеству  $S_i$  принадлежат выражения, обозначающие те функции от термов предыдущих слоев  $S_0, \dots, S_{i-1}$ , которые сами им не принадлежат.

Для работы с термами полезны две ключевые теоремы, которые позволяют эффективно манипулировать этими выражениями.

**Теорема 1 (о построении термов).** Каждый терм  $t$  единственным образом представляется в виде  $ft_1, \dots, t_k$  где  $f$  - первый символ в  $t$ ,  $f \in F$ , число  $k$  определяется по  $f$  ( $f \in F_k$ ), а  $t_1, \dots, t_k$  - термы.

Эта теорема является точной формулировкой эквивалентности используемой бесскобочной и обычной записи.

Пусть  $u$  и  $v$  - выражения, то есть последовательности символов алфавита. Скажем, что  $u$  входит в  $v$ , если существуют такие выражения  $p$  и  $q$  (возможно, пустые), что  $v$  совпадает с  $puiq$ . [1, 8]

**Теорема 2 (о вхождении терма в терм).** Пусть  $f \in F_k$ ,  $t_1, \dots, t_k$  - термы,  $t$  представляется в виде  $ft_1, \dots, t_k$ ,  $\tau$  - терм и  $\tau$  входит в  $t$ . Тогда или  $\tau$  совпадает с  $t$ , или  $\tau$  входит в одно из  $t_i$  ( $i = 1, \dots, k$ ).

Доказываются эти теоремы элементарной индукцией по длине термов. В доказательстве теоремы 2 выделяется лемма, представляющая и самостоятельный интерес. [1]

**Лемма 1.** Каждое вхождение любого символа в терм  $\tau$  начинает вхождение некоторого терма в  $\tau$ . Определим отношение между термами  $t_1 \leq t_2$  индуктивным образом «сверху вниз» - по глубине вхождения:

- 1)  $t \leq t$ ;
- 2) если  $t$  совпадает с  $ft_1, \dots, t_k$ ,  $f \in F_k$  и  $t_1, \dots, t_k$  - термы, то  $t_1, \dots, t_k \leq t$ ;
- 3) если  $t_1 \leq t$  и  $t \leq t_2$ , то  $t_1 \leq t_2$ .

Согласно теореме 2,  $t_1 \leq t_2$ . тогда и только тогда, когда  $t_1$  входит в  $t_2$ .

Для каждого терма  $t$  определим множество входящих в него термов  $S^t = \{\tau | \tau \leq t\}$ . Если  $t \in S_i$ , то при  $0 \leq k \leq i$  непусты множества  $S_k^t = S^t \cap S_k$ . При этом множество  $S_i^t$  состоит из одного элемента - исходного терма  $t$ .

Свяжем с термом  $t$  ориентированный граф  $G_0^t$  с вершинами, взаимнооднозначно соответствующими термам из  $S^t$ . Будем одинаково обозначать вершины и соответствующие им термы. Пара вершин  $(\tau_1, \tau_2)$  образует ориентированное от  $\tau_1$  к  $\tau_2$  ребро графа  $G_0^t$ , если терм  $\tau_2$  имеет вид

$f, t_1, \dots, t_k, f \in F_k$ ,  $t_1, \dots, t_k$  - термы и один из них  $t_i$  ( $i = 1, \dots, k$ ) совпадает с  $\tau_1$ .  
Вершины графа  $G_0^t$  удобно располагать по слоям  $S_i^t$ .

Для произвольного графа  $G$  будем обозначать  $v(G)$  множество вершин, а  $e(G)$  - множество ребер  $G$ .

Возьмем для примера выражение для сложной функции

$$\varphi(x_1, x_2, x_3) = f_5 \left( f_3(x_1, f_1(x_1, x_2)), f_4(f_1(x_1, x_2), f_2(x_2, x_3)) \right) \quad (4)$$

В принятой выше бесскобочной префиксной записи оно имеет вид

$$f_5 f_3 x_1 f_1 x_1 x_2 f_4 f_1 x_1 x_2 f_2 x_2 x_3 \quad (4')$$

где все функциональные символы принадлежат  $F_2$ . Граф  $G_0^t$  для этого терма изображен на рисунке 2.

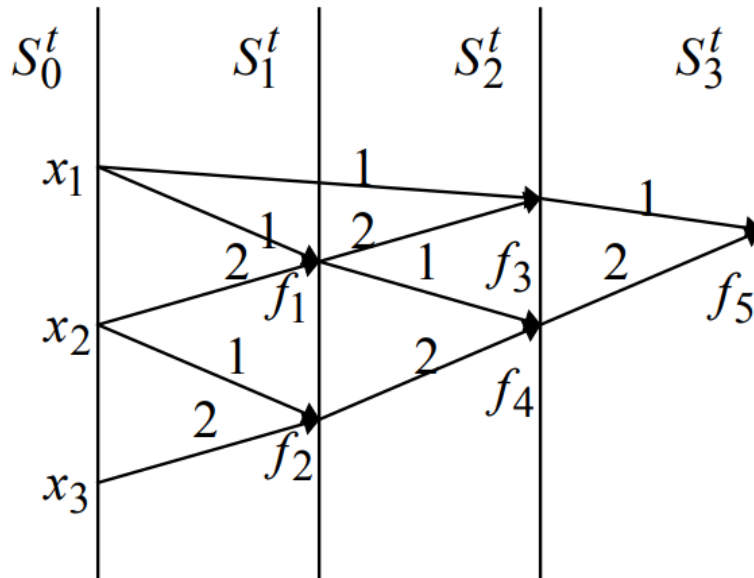


Рисунок 2 – Пример графа  $G_0^t$

Для того, чтобы терм однозначно восстанавливался по графу, необходимы еще два дополнения.

1. Сопоставим каждой вершине  $\tau \in v(G_0^t)$  метку  $p(\tau)$  - символ алфавита. Если вершина принадлежит нулевому слою  $S_0^t$ , то ей соответствует терм, совпадающий с символом из  $C \cup V$ . Этот символ и сопоставляется вершине в качестве метки. Если вершина принадлежит  $S_i^t$  ( $i > 0$ ), то меткой служит



функциональный символ: вершине  $\tau$  сопоставляется  $f \in F$ , если  $\tau$  имеет вид  $f t_1, \dots, t_k$ , где  $f \in F_k$ , а  $t_1, \dots, t_k$  - термы.

2. Каждому ребру  $(\tau', \tau) \in e(G_0^t)$ , приходящему в вершину  $\tau$ , сопоставим метку  $P(\tau', \tau)$  - конечное множество натуральных чисел (номеров): пусть терм  $\tau$  имеет вид  $f t_1, \dots, t_k$  где  $f \in F_k$ , а  $t_1, \dots, t_k$  - термы, тогда ребру  $(\tau', \tau)$  сопоставляется множество тех  $i$  ( $1 \leq i \leq k$ ), для которых  $\tau'$  совпадает с  $t_i$ . На практике в большинстве случаев эта метка состоит из одного номера, но возможны и другие варианты - так же, как функции вида  $f(x, x)$ . Для графических иллюстраций удобно ребра  $(\tau', \tau)$ , имеющие в своей метке  $P(\tau', \tau)$  больше одного номера, рисовать как пучок ребер, идущих от вершины  $\tau'$  к вершине  $\tau$  - по одному такому ребру для каждого номера из  $P(\tau', \tau)$ ; этот номер и будет меткой соответствующего ребра из пучка.

Граф  $G_0^t$  вместе со всеми метками будем обозначать  $G^t$ . На рисунке 2 указаны соответствующие метки для разобранных примера.

Итак, для всякого терма  $t$  построен ориентированный граф  $G_0^t$  и две функции: первая сопоставляет каждой вершине  $\tau \in v(G_0^t)$  символ алфавита  $p(\tau) \in C \cup V \cup F$ , вторая (обозначим ее  $P$ ) - каждому ребру  $(\tau', \tau) \in e(G_0^t)$  - конечное множество натуральных чисел  $P(\tau', \tau)$ . Отмеченный граф - набор  $(G_0^t, p, P)$  обозначаем  $G^t$ . Функции  $p$  и  $P$  удовлетворяют следующему ограничению:

Если для данного  $\tau \in S^t$  множество входящих ребер  $(\tau', \tau)$  не пусто, то  $p(\tau) = f^\tau \in F_k$  (является  $k$ -местным функциональным символом при некотором  $k$ ) и семейство множеств  $\{P(\tau', \tau) | (\tau', \tau) \in e(G_0^t)\}$  при фиксированном  $\tau$  образует разбиение множества номеров  $\{1, \dots, k\}$ , то есть  $P(\tau', \tau) \cap P(\tau'', \tau) = \emptyset$  при  $\tau' \neq \tau''$ ,  $\forall \tau \bigcup_{\tau': (\tau', \tau) \in e(G_0^t)} P(\tau', \tau) = \{1, \dots, k\}$ . [1, 8]

На этом завершается изложение основных формальных конструкций. Прежде, чем переходить к интерпретации, сформулируем теорему об эквивалентности графического и формульного представления термов.

Пусть  $G$  - конечный ориентированный граф, не имеющий ориентированных циклов, и в  $G$  существует и единственна такая вершина  $\tau^*$ , к которой от любой вершины ведет ориентированный путь. Пусть, далее, заданы две функции:  $p$  - на множестве вершин  $G$  со значениями в множестве символов алфавита и  $P$  - на множестве ребер  $G$  со значениями в множестве конечных наборов натуральных чисел и выполнено условие.

**Теорема 3.** Существует и единственен терм  $t$ , для которого  $G^t = (G, p, P)$ . Доказательство проводится в два этапа. Сначала в  $G$  устанавливается послойная структура: строится разбиение множества вершин  $G: s_0 \cup s_1 \cup \dots \cup s_k$ . Множество  $s_0$  состоит из тех вершин, к которым не ведет ни одного ребра - из-за отсутствия ориентированных циклов такие вершины существуют. Множество  $s_{i+1}$  состоит из тех вершин, к которым ведут ребра только из элементов  $s_0 \cup s_1 \cup \dots \cup s_i$ . Последний непустой элемент в последовательности  $s_0 \cup s_1 \cup \dots \cup s_k$  состоит из одной вершины  $\tau^*$ , все предшествующие элементы этой последовательности непусты, а объединение всех  $s_i$  содержит все вершины  $G$ .

Доказательство основного утверждения теоремы проводится индукцией по числу слоев  $k$ .

Интерпретация сопоставляет терму сложную функцию. Она строится так. Задается некоторое множество  $D$  - область интерпретации. Каждой константе  $c$ , входящей в интерпретируемый терм  $t$ , сопоставляется элемент из  $D$  ( $c \in D$ ), каждому  $k$ -местному функциональному символу  $f$ , входящему в  $t$ , сопоставляется функция  $k$  переменных  $f: D^k \rightarrow D$ . Каждой переменной, входящей в интерпретируемый терм  $t$ , сопоставляется переменная, пробегающая  $D$ . В результате терму  $t$  сопоставляется функция  $n$  переменных  $F^t: D^n \rightarrow D$ , где  $n$  - число различных символов переменных, входящих в  $t$ . Эта «сложная» функция получается суперпозицией «простых» функций, соответствующих функциональным символам.

Если  $t \in S_0$ , то есть терм является константой или переменной, то вместо сложной функции  $F^t$  получаем константу или тождественную функцию  $id$ , переводящую значение переменной в него же. Если  $t \in S_1$ , то соответствующая функция является «простой». Истинные сложные функции появляются, начиная со слоя  $S_2$ .

Заданная интерпретация терма  $t$  одновременно определяет интерпретацию каждого терма, входящего в  $t$ . Представим процесс вычисления сложной функции  $F^t$  с помощью отмеченного графа  $G^t$ . Строится два представления: статическое и динамическое.

1) Статическое представление: все результаты промежуточных вычислений находятся на графе  $G^t$ , который отображает структуру вычислений на каждом уровне.

2) Динамическое представление: здесь вычисления происходят шаг за шагом, слой за слоем, начиная с самых простых термов и продвигаясь к более сложным.

Пусть задана интерпретация терма  $t$  и определены значения всех переменных, входящих в  $t$ . Тогда для любого терма  $\tau$ , входящего в  $t$ , также задана интерпретация и определены значения всех функций  $F^\tau$  ( $\tau \leq t$ ). Каждой вершине  $\tau$ , входящей в граф  $G_0^t$ , сопоставляется значение функции  $F^\tau$  - элемент  $D$ . При этом вершинам нулевого слоя соответствуют значения переменных и констант, а единственной вершине последнего (выходного) слоя - значение  $F^t$ . Будем называть элементы  $D$ , соответствующие вершинам, значениями этих вершин и обозначать их  $Z(\tau)$ .

Для каждой вершины  $\tau$ , принадлежащей ненулевому слою, можно выписать уравнение функционирования, связывающее значения вершин. Пусть  $p(\tau) = f^\tau, f^\tau \in F_k$  и  $in(\tau)$  - совокупность входящих в  $\tau$  ребер. Напомним, что совокупность меток ребер, оканчивающихся в  $\tau$ ,  $\{P(\tau', \tau) | (\tau', \tau) \in in(\tau)\}$  образует разбиение множества  $\{1, 2, \dots, k\}$ .

Уравнение функционирования для вершины  $\tau$ , принадлежащей ненулевому слою, имеет вид

$$Z(\tau) = f^\tau(z_1, \dots, z_k), z_i = Z(\tau') \text{ при } i \in P(\tau', \tau), (\tau', \tau) \in in(\tau) \quad (5)$$

В силу уравнения функционирования (5), если для входящих в  $\tau$  ребер  $(\tau', \tau) \in in(\tau)$  известны значения  $Z(\tau')$  и задана интерпретация символа  $p(\tau) = f^\tau$  - метки вершины, то можно найти значение вершины  $Z(\tau)$ . На основании этого (очевидного) замечания строится динамическое представление вычисления сложной функции.

С каждой вершиной графа  $\tau$ , принадлежащей ненулевому слою, ассоциируется автомат, вычисляющий функцию  $f^\tau(z_1, \dots, z_k)$ , где  $f^\tau \in F_k$  - метка вершины  $\tau$ . Автоматы срабатывают по слоям в дискретные моменты времени (такты) - автоматы  $i$ -го слоя в  $i$ -й момент времени. В начальный момент сформированы значения вершин нулевого слоя - известны значения переменных и констант. Они поступают на входы автоматов первого слоя в соответствии с нумерацией аргументов. После  $i$ -го такта функционирования определены значения вершин, принадлежащих слоям  $S_0, \dots, S_i$ . На  $i + 1$ -м такте автоматы  $i + 1$ -го слоя вычисляют значения вершин  $i + 1$ -го слоя, получая входные сигналы с предыдущих слоев по правилу (5) - в соответствии с метками входящих ребер.

**Теорема 4.** Пусть задана интерпретация всех символов, отмечающих вершины графа  $G$ , определены значения независимых переменных (а также констант), соответствующих вершинам входного слоя  $v_{in}$  и значения независимых переменных двойственного функционирования  $\mu(\tau)$  для вершин выходного слоя  $v_{out}$ . Тогда для любого  $\theta \in v_{in}$

$$\mu(\theta) = \sum_{\tau \in v_{in}} \mu(\tau) \frac{\partial F^\tau}{\partial x_\theta} \cdot (6)$$

где  $F^\tau$  - соответствующая вершине  $\tau$  функция от независимых переменных и констант, отмечающих вершины нулевого слоя  $G$ ,  $x_\theta$  - соответствующая

вершине  $\theta \in v_{in}$  переменная или константа, а производные в (6) берутся при фиксированных значениях прочих переменных и констант.

Доказательство проводится индукцией по числу слоев. Для графов из двух слоев - нулевого и первого - теорема очевидна. Спуск от  $i + 1$ -слойных графов к  $i$ -слойным производится с помощью формулы для дифференцирования «двухслойной» сложной функции нескольких переменных.

$$\frac{\partial f(A_1(x_1, \dots, x_n), A_2(x_1, \dots, x_n), \dots, A_k(x_1, \dots, x_n))}{\partial x_i} = \sum_{j=1}^k \frac{\partial f}{\partial_j^A f} \times \frac{\partial A_j}{\partial x_i}$$

Прохождение вершины графа и прилегающих к ней ребер при прямом и обратном процессах проиллюстрировано на рисунке 3.

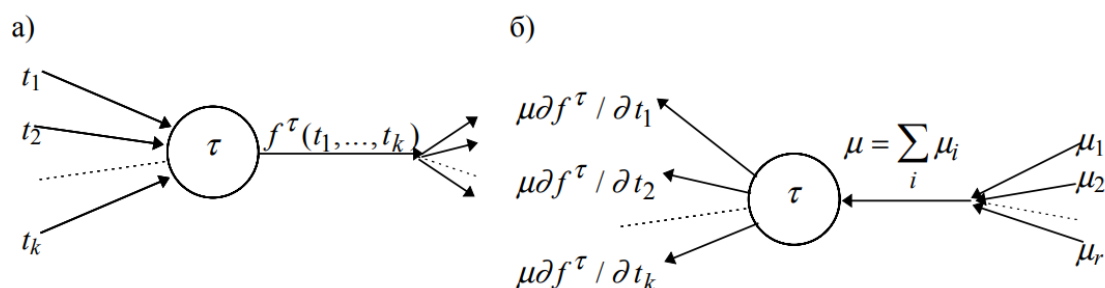


Рисунок 3 – Прохождение вершины  $\tau$  в прямом (а) и обратном (б) направлении

## 2.3 Теорема о скрытых параметрах

**Скрытые параметры** — это параметры, которые не наблюдаются напрямую, но оказывают влияние на результаты модели или системы. В контексте теории регрессии и анализа данных скрытые параметры могут быть связаны с невидимыми переменными, которые влияют на наблюдаемые признаки объектов, но не могут быть измерены напрямую.

Ряд алгоритмов решения проблемы скрытых параметров можно построить на основе следующей теоремы. Пусть  $n$  - число свойств,  $N$  - количество объектов,  $\{x^i\}_{i=1}^N$  - множество векторов значений признаков. Утверждается, что группа объектов подчиняется уравнениям регрессии ранга  $r$ , если все векторы

$\{x^i\}_{i=1}^N$  принадлежат  $n - r$ -мерному линейному многообразию. Обычно в реальных задачах выполняется условие  $N > n$ . Если же  $n \geq N$ , то векторы  $\{x^i\}_{i=1}^N$  принадлежат  $N - 1$ -мерному линейному многообразию и нетривиальные регрессионные связи возникают лишь при ранге  $r > n - N + 1$ . Ранг регрессии  $r$  определяет количество независимых линейных связей между исследуемыми признаками объектов. Он отражает коразмерность линейного подпространства, в котором лежат векторы признаков объектов. Конечно, в реальных данных нужно учитывать погрешности, но в данном контексте будем рассматривать точные связи.

Следующая теорема о скрытых параметрах позволяет преобразовать задачу, связанную с зависимостью между свойствами одного объекта, в задачу, где связь исследуется между свойствами разных объектов, позволяя "транспонировать" задачу регрессии. Это превращает задачу о качественной неоднородности выборки в задачу поиска такой группы объектов, через признаки которых разные свойства конкретного объекта выражаются одинаково и оптимально.

**Теорема.** Пусть для некоторого  $r > 0$  существует такое разбиение  $\{x^i\}_{i=1}^N$  на группы  $\{x^i\}_{i=1}^N = \bigcup_{j=1}^k Y_j$ , что  $r > n - N_j + 1$  (где  $N_j$  - число элементов в  $Y_j$ ), и для каждого класса  $Y_j$  выполняются уравнения регрессии ранга  $r$ . Тогда для каждого объекта  $x^i$  из  $\{x^i\}_{i=1}^N$  найдется такое множество  $W_i$  (опорная группа объекта  $x^i$ ) из  $k$  объектов, что  $n - r + 1 \geq k$  и для некоторого набора коэффициентов  $\lambda_y$

$$x^i = \sum_{y \in W_i} \lambda_y y, \quad \sum_{y \in W_i} \lambda_y = 1. \quad (7)$$

Последнее означает, что значение каждого признака объекта  $x^i$  может быть представлено как линейная функция от значений этого признака для

объектов опорной группы. При этом линейная функция остаётся одинаковой для всех признаков.

Линейная зависимость (7) обладает особенностью, что она инвариантна к изменениям единиц измерения свойств и сдвигам начала отсчета. Например, если координаты всех векторов признаков подвергнутся неоднородным линейным преобразованиям вида  $x_j \rightarrow a_j x_j + b_j$ , где  $j$  - номер координаты, то линейная связь (7) останется неизменной. Инвариантность относительно изменения масштаба обеспечивается линейностью и однородностью связи, а инвариантность относительно сдвига начала отсчета — тем, что сумма коэффициентов  $\lambda_y$  равна 1. [1, 2]

Сформулированная теорема позволяет преобразовать задачу регрессии (поиск зависимостей между признаками одного объекта) в транспонированную задачу, где требуется найти линейную зависимость признаков одного объекта от признаков других объектов, а также определить опорные группы, для которых эта зависимость является наилучшей.

Доказательство основывается на следующем факте: на каждом  $k$ -мерном линейном многообразии для любого набора из  $q$  точек  $y_1, y_2, \dots, y_q$  при  $q > k + 1$  выполнено соотношение  $\sum_{j=1}^q \lambda_j y_j = 0$  для некоторого набора  $\lambda_j$ ,  $\sum_{j=1}^q \lambda_j = 0$  и некоторые  $\lambda_j \neq 0$ . С математической точки зрения теорема о скрытых параметрах представляет собой вариант утверждения о равенстве ранга матрицы, вычисляемого по строкам, рангу, вычисляемому по столбцам.

## 2.4 Универсальная теорема аппроксимации

Теорема утверждает, что нейронная сеть прямого распространения с одним скрытым слоем, содержащим конечное число нейронов, может аппроксимировать любую непрерывную функцию на компактных подмножествах  $R^n$  при наличии соответствующих функций активации. Этот результат имеет глубокие последствия для проектирования, возможностей и понимания нейронных сетей. [11]

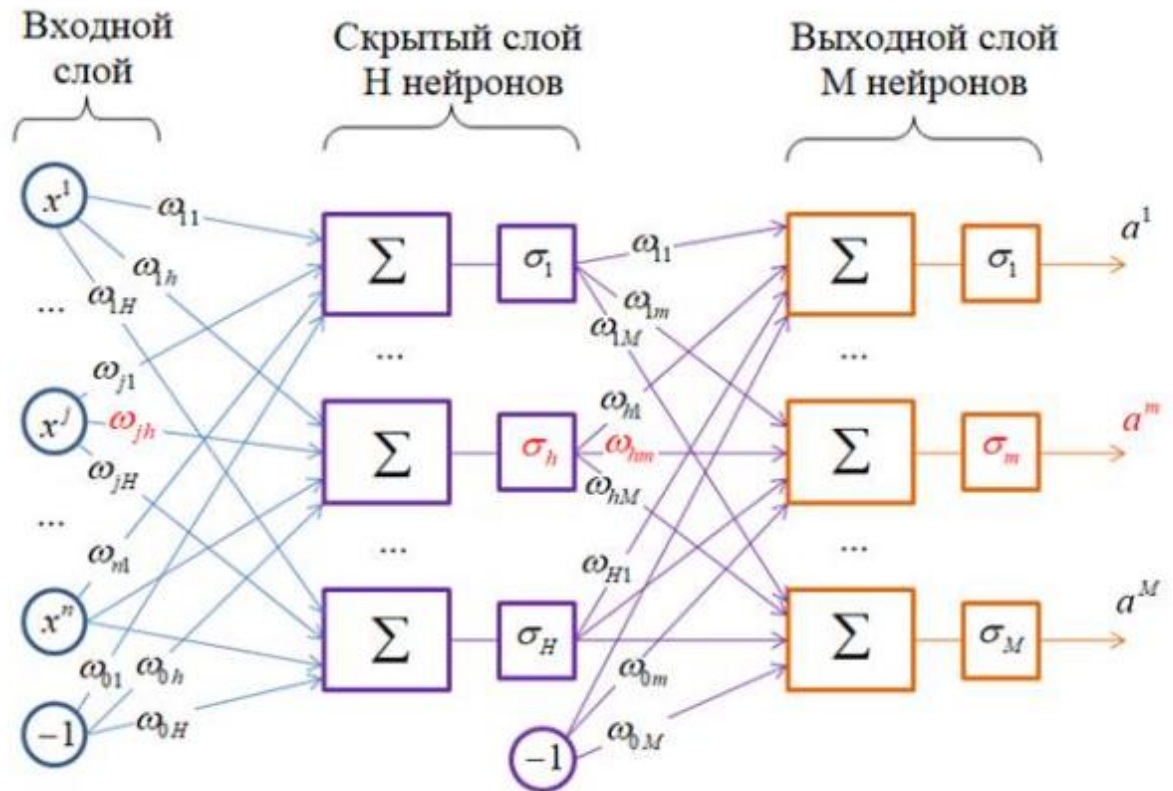


Рисунок 4 - Общая схему полносвязной нейронной сети прямого распространения

Здесь на вход подается вектор из  $n + 1$  признаков:  $x = [-1, x^1, x^2, \dots, x^n]^{-T}$

Затем, на каждом сумматоре вычисляется взвешенная сумма признаков с некоторыми весовыми коэффициентами:  $\Sigma_h = \sum_{j=0}^n \omega_{jh} \cdot x^j, h = 1, 2, \dots, N$

Эта сумма пропускается через нелинейную функцию (функцию активации) и формируется выходное значение соответствующего нейрона

скрытого слоя:  $u^h = \sigma_h(\Sigma_h) = \sigma_h\left(\sum_{j=0}^n \omega_{jh} \cdot x^j\right), h = 1, 2, \dots, N$

Получаем вектор промежуточных значений:  $u = [-1, u^1, u^2, \dots, u^N]^{-T}$

После этого выходы  $\{u^h\}$  нейронов скрытого слоя подаются на следующий слой (выходной в нашем примере) и формируются  $M$  выходных значений нейронной сети:



$$a^m = \sigma_m \left( \sum_{h=0}^H \omega_{hm} \cdot u^h \right), \quad m = 1, 2, \dots, M$$

Вот в двух словах принцип работы полносвязной нейронной сети прямого распространения. Легко видеть, что каждый нейрон здесь представляет линейный алгоритм, подобно линейному классификатору:

$$a(x) = \text{sign}(\langle \omega, x \rangle) = \sigma(\langle \omega, x \rangle)$$

То есть, нейронная сеть – это суперпозиция линейных алгоритмов, но с нелинейными функциями активаций. Почему так важно, чтобы функция  $\sigma(\cdot)$  была нелинейной? Если взять ее линейной, например,  $\sigma(x) = x$ , то мы получим суперпозицию обычных линейных алгоритмов, которые вырождаются в один линейный алгоритм. То есть, нейронная сеть с линейными функциями активаций можно представить одним нейроном, в виде модели:  $a(x) = \sigma(\langle \omega, x \rangle)$

Чтобы суперпозиция имела смысл, должны использоваться нелинейные функции активации. Исключение составляют выходные нейроны, там иногда (в задачах регрессии) используются линейные активационные функции.

Вообще, существует определенный набор для функций  $\sigma(x)$ . Наиболее часто используют, следующие:

- ступенчатая:  $\sigma(x) = \begin{cases} 1, & x \geq \theta \\ 0, & x < \theta \end{cases}$
- сигмоидная:  $\sigma(x) = \frac{1}{1 + \exp(-x)}$
- гиперболический тангенс:  $\sigma(x) = \text{th}(x)$
- ReLU:  $\sigma(x) = \max(0, x)$

И в выходных нейронах:

- линейная:  $\sigma(x) = x$
- softmax:  $\sigma(x) = \frac{\exp(x^m)}{\sum_{j=1}^M \exp(x^j)}$

Конечно, существует и множество других функций активаций. [9, 10]

Теорема универсальной аппроксимации была независимо доказана Джорджем Цибенко в 1989 году и Куртом Хорником в 1991 году. Доказательство Цибенко конкретно касалось сетей с сигмовидными функциями активации, а работа Хорника распространила результат на более широкий класс функций активации. [10]

## **ЗАКЛЮЧЕНИЕ**

Основные теоремы нейроинформатики представляют собой фундаментальные положения, определяющие теоретические возможности и ограничения нейронных сетей. Они позволяют формализовать подходы к решению сложных задач обработки информации, подчеркивая важность выбора архитектуры, параметров и методов обучения. Эти теоремы обеспечивают глубокое понимание принципов работы нейронных сетей и создают основу для разработки более мощных и универсальных моделей.

Однако, несмотря на значительный прогресс, остаются нерешённые проблемы, связанные с глобальной оптимизацией, интерпретацией результатов и адаптацией моделей к изменяющимся условиям. Эти вызовы указывают на необходимость дальнейших исследований, направленных на расширение возможностей существующих теоретических подходов.

Закреплённые в теоремах положения не только демонстрируют силу нейронных сетей, но и подчёркивают их зависимость от математических и вычислительных аспектов. Будущее нейроинформатики связано с углублением понимания этих взаимосвязей и развитием методов, способных преодолевать существующие ограничения, открывая новые горизонты в области искусственного интеллекта.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Горбань, А.Н. Нейроинформатика / А.Н. Горбань, В.Л. Дунин-Барковский, А.Н. Кирдин и др. - Новосибирск: Наука. Сибирское предприятие РАН, 1998. - 296с.
- 2 Ismayilova, A. On the Kolmogorov neural networks / A. Ismayilova, V. E Ismayilov // Neural Networks. – 2023. – P. 14.
- 3 Puknys, L. The Stone-Weierstrass theorem / L. Puknys // The University of Chicago Mathematics. – 2017. – P. 20.
- 4 Young, M. The Stone-Weierstrass Theorem / M. Young // Queen's University at Kingston Winter Term. – 2006. – P. 10.
- 5 Фёдорова, В. П. Теорема Стоуна–Вейерштрасса и пространства мер / В. П. Фёдорова. - Математические заметки, 2002. - 417–427с.
- 6 Горбань, А.Н. Обобщенная аппроксимационная теорема и точное представление многочленов от нескольких переменных суперпозициями многочленов от одного переменного / А.Н. Горбань. - Изв. вузов. Матем., 1998. – 6-9с.
- 7 Gorban, A.N. Neuroinformatics: What are us, where are we going, how to measure our way? / A.N. Gorban // Institute of Computational Modelling of SB RAS. – 2003. – P. 9.
- 8 Дадеркин, Д. О. Эффективные алгоритмы построения термов минимальной вычислительной сложности / Д. О. Дадеркин - Вестник ТвГУ. Серия: Прикладная математика, 2016. – 21-33с.
- 9 Бустинг и нейронные сети: Нейронные сети. Краткое введение в теорию [Электронный ресурс] // 2022. – URL: <https://proproprogs.ru/ml/ml-neuronnye-seti-kratkoe-vvedenie-v-teoriyu> (Дата обращения 20.12.2024). – Загл. с экрана. – Яз. рус.
- 10 Smorodinov, A.D. Theoretical foundations of artificial neural network application to approximation and interpolation problems / A.D. Smorodinov,

T.V. Gavrilenko, V.A. Galkin // Russian Journal of Cybernetics, – 2023. – P. 41–53

- 11 Горбань, А.Н. Обобщенная аппроксимационная теорема и вычислительные возможности нейронных сетей / А.Н. Горбань. - Сиб. журн. вычисл. матем, 1998. – 11-24с.