

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ
Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра теории функций и стохастического анализа

Аппроксимативные свойства нейронных сетей прямого распространения
сигнала

наименование темы выпускной квалификационной работы

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

Студента 2 курса 218 группы

направления 01.04.02 «Прикладная математика и информатика»

механико-математического факультета

Пескова Игоря Сергеевича

фамилия, имя, отчество

Научный руководитель,
зав. кафедрой, д.ф.-м.н.,
доцент

С. П. Сидоров

подпись, дата

Заведующий кафедрой,
зав. кафедрой, д.ф.-м.н.,
доцент

С. П. Сидоров

подпись, дата

Саратов 2025

В данной работе рассмотрены аппроксимативные свойства нейронных сетей прямого распространения сигнала.

Актуальность. Нейронные сети прямого распространения сигнала актуальны благодаря фундаментальной роли в машинном обучении и их широкому применению в компьютерном зрении, NLP и прогнозировании. Исследование их способности приближать сложные функции важно для оптимизации архитектур, ускорения обучения и повышения интерпретируемости моделей. Несмотря на классические результаты, остаются открытые вопросы — влияние глубины сети, выбор функций активации и эффективность в задачах высокой размерности, что делает тему востребованной как в теории, так и на практике.

Целью работы является теоретически и экспериментально исследовать аппроксимативные свойства нейронных сетей прямого распространения сигнала, реализовать алгоритм вычисления параметров нейронной сети для аппроксимации многочлена, сравнение полученной функций с исходной.

Задачи:

- Изучить теоретические основы нейронных сетей прямого распространения сигнала, включая архитектуру, функции активации и алгоритмы обучения.
- Проанализировать теорему об универсальном аппроксиматоре и ее практические ограничения.
- Исследовать влияние архитектуры нейронной сети (количество слоев, количество нейронов в слоях), выбора функции активации и методов оптимизации на аппроксимативные свойства сети.
- Провести экспериментальное исследование по аппроксимации различных классов функций (полиномиальные, тригонометрические, разрывные) с использованием нейронных сетей различной архитектуры.
- Разработать рекомендации по выбору оптимальных параметров нейронной сети для решения задач аппроксимации.

Структура магистерской работы:

- В первом разделе даны теоретические основы нейронных сетей прямого распространения сигнала.
- Во втором разделе описана постановка задачи аппроксимации, а так-

же реализация алгоритма поиска параметров нейронной сети прямого распространения сигнала.

Основное содержание работы.

Раздел 1 посвящен общим сведениям о сетях.

Нейронная сеть представляет собой совокупность взаимосвязанных вычислительных узлов (нейронов), организованных в слои: входной, скрытые и выходной. Каждый нейрон скрытого или выходного слоя получает входной вектор $\mathbf{x} = (x_1, x_2, \dots, x_n)$ и вычисляет взвешенную сумму:

$$z = \sum_{i=1}^n w_i x_i + b,$$

где w_i — веса связей, b — смещение (bias). Далее применяется функция активации $\phi(z)$, формируя выход нейрона.

Математическая модель нейрона формализуется как композиция линейного преобразования и нелинейной функции активации:

$$y = \phi(z) = \phi \left(\sum_{i=1}^n w_i x_i + b \right),$$

где $\phi : \mathbb{R} \rightarrow \mathbb{R}$ — непрерывная и, как правило, монотонная функция. Примеры: сигмоида, $\tanh$, ReLU.

Согласно универсальной теореме аппроксимации, нейронные сети с одним скрытым слоем и подходящей функцией активации способны приближать любую непрерывную функцию на компакте:

$$\forall \varepsilon > 0, \exists F(\mathbf{x}) \text{ такое, что } \sup_{\mathbf{x} \in K} |f(\mathbf{x}) - F(\mathbf{x})| < \varepsilon.$$

Обучение заключается в минимизации функции потерь:

$$\mathcal{L}(\theta) = \frac{1}{m} \sum_{i=1}^m \|f_{\theta}(x^{(i)}) - y^{(i)}\|^2,$$

где f_{θ} — выход сети с параметрами θ , а m — число обучающих примеров. где η — скорость обучения.

Раздел 1.1.5 содержит теорему о том что: функция активации σ непре-

рывна на R , тогда M - множество всех отображений осуществляемых однослойными нейронными сетями определенными в R^n будет плотно со значениями в R . Доказательство теоремы проходит через гребневые функции.

Раздел 1.1.6 посвящен теореме о множестве ридж-функций.

Ридж-функция:

$$\mathcal{R}(A) = \text{span}\{g(\mathbf{a} \cdot \mathbf{x}) : g \in C(\mathbb{R}), \mathbf{a} \in A\}$$

плотно в $C(\mathbb{R}^n)$, в топологии равномерной сходимости на компактах, тогда и только тогда, когда нет нетривиального однородного многочлена, который обращается в нуль на A .

Раздел 1.1.7 содержит информацию о процессе обучения нейронной сети. Подчеркивает, что нейронные сети стали неотъемлемой частью современного мира и широко используются для решения задач, где традиционные алгоритмы неэффективны. Указывается, что это направление стабильно лидирует. Упоминается поиск методов для эффективной параллельной работы НС. Текст дает общее представление о нейронных сетях как мощном и востребованном инструменте, объясняет их природу (адаптивность, аналогия с мозгом), ключевое отличие (обучение vs программирование) и базовый принцип работы.

Раздел 1.1.9 посвящен типам нейронных сетей. Существует несколько типов нейронных сетей, которые применяются в различных задачах:

- **Персептрон**: это самая простая нейронная сеть, состоящая из одного слоя нейронов. Он может решать только задачи линейной классификации.
- **Многослойный персептрон (MLP)**: состоит из нескольких скрытых слоев и может решать более сложные задачи, включая нелинейные классификации.
- **Сверточные нейронные сети (CNN)**: используются в основном для обработки изображений и видео. Они включают сверточные слои, которые эффективно извлекают признаки из изображений.
- **Рекуррентные нейронные сети (RNN)**: предназначены для обработки последовательных данных, таких как текст, речь и временные ряды.

- **Генеративно-сопоставительные сети (GAN)**: используются для генерации новых данных, таких как изображения, которые выглядят как реальные.

Раздел 1.2 посвящен функциям активации. Функции активации придают нейросетям способность к обучению сложных нелинейных зависимостей.

Примеры:

- Сигмоида:

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \quad \sigma'(x) = \sigma(x)(1 - \sigma(x))$$

- Тангенс гиперболический:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad \tanh'(x) = 1 - \tanh^2(x)$$

- ReLU:

$$\text{ReLU}(x) = \max(0, x), \quad \text{ReLU}'(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Выбор функции активации влияет на способность модели к обучению и скорости сходимости.

Раздел 1.3 ВКР содержит информацию о том как используется многочлены в вычислениях.

Многочлен степени n задается выражением:

$$P(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n = \sum_{k=0}^n a_kx^k,$$

где a_k — коэффициенты, x — переменная. Многочлены применяются для интерполяции, аппроксимации и численного анализа.

Коэффициенты a_k определяют поведение и форму полинома. От их выбора зависит точность аппроксимации и устойчивость методов вычислений.

Полином Тейлора используется для приближенного представления гладкой функции:

$$f(x) \approx \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k.$$

Многочлены применяются как функции активации, а также используются для задания весовых коэффициентов или целей аппроксимации.

Для приближения табличных данных минимизируется функция ошибки:

$$\min_{a_0, \dots, a_n} \sum_{i=1}^m \left(y_i - \sum_{k=0}^n a_k x_i^k \right)^2.$$

Один из главных методов нахождения коэффициентов многочлена:

$$\min_{\mathbf{a}} \|X\mathbf{a} - \mathbf{y}\|_2^2,$$

где X — матрица базисных функций, $\mathbf{a}$ — вектор коэффициентов.

Нелинейные модели используют произведения переменных и их степени:

$$P(x_1, x_2, \dots) = a_0 + a_1 x_1 + a_{12} x_1 x_2 + a_{13} x_1^2 x_3 + \dots$$

Многочлены используются для аппроксимации сложных зависимостей между входными и выходными переменными. Пусть задана выборка (x_i, y_i) , тогда полиномиальная модель может быть представлена как:

$$y_i \approx P(x_i) = \sum_{k=0}^n a_k x_i^k,$$

что позволяет эффективно аппроксимировать нелинейные функции.

В нейронных сетях возможно использование многочленов в качестве функций активации. Например:

$$\phi(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3,$$

что позволяет получить гладкие и дифференцируемые активации. Такие функции удобны для анализа и упрощают вычисление градиентов.

Во время обучения можно использовать многочлены для регуляризации или инициализации весов. Пример — задание весов в виде:

$$w(x) = \sum_{k=0}^n a_k x^k,$$

что снижает размерность параметров и улучшает обобщающую способность сети.

В RNN можно использовать полиномиальные рекурсии:

$$h_t = \phi \left(\sum_{k=0}^n a_k h_{t-1}^k + W x_t + b \right),$$

что позволяет моделировать сложные временные зависимости с большей гибкостью.

Весовые коэффициенты могут быть заданы не как свободные параметры, а как значения полинома:

$$w_{ij} = P(x_i, x_j) = \sum_{k=0}^d a_k (x_i x_j)^k,$$

что уменьшает количество параметров и облегчает обучение.

Рассмотрим пример сети с одним скрытым слоем, где активация задана полиномом степени 3:

$$\phi(x) = x + 0.5x^2 - 0.1x^3.$$

Такой подход демонстрирует хорошую аппроксимационную способность при минимуме параметров.

В глубоких сетях возможно использование многочленов не только как активаций, но и как аппроксиматоров целевой функции:

$$y \approx P(f(x)) = \sum_{k=0}^n a_k (f(x))^k,$$

где $f(x)$ — выход последнего скрытого слоя.

Раздел 1.4 посвящен тому что Многочлен могут быть использованы в нейронных сетях для различных целей, включая аппроксимацию функций, вычисление весов и улучшение процесса обучения. В частности, многочлены

могут быть использованы для моделирования сложных зависимостей между входами и выходами нейронной сети, что особенно полезно в задачах, где функция, которую пытается аппроксимировать сеть, имеет сложную нелинейную структуру.

Обсуждается также что в нейронных сетях для аппроксимации функций часто используются многочлены в качестве активационных функций. Например, можно использовать полиномы для аппроксимации функции активации нейрона, что улучшает способности сети моделировать сложные зависимости.

Раздел 1.4.4 содержит информацию о том что рекуррентные нейронные сети (RNN) используются для работы с последовательными данными, такими как текст или временные ряды. В таких сетях информация передается через временные шаги, и часто используется механизм памяти для запоминания предыдущих состояний.

Раздел 1.4.7 посвящен тому что в глубоких нейронных сетях (DNN) многослойные структуры позволяют моделировать более сложные функции. Использование многочленов в глубоких сетях может улучшить способность сети обрабатывать сложные данные. Например, многочлены могут использоваться для аппроксимации различных частей нейронной сети, таких как скрытые слои или функции активации.

Раздел 1.5 включает методы вычисления коэффициентов многочленов.

Аналогично обучению нейронных сетей, можно минимизировать ошибку по коэффициентам полинома:

$$a_k^{(t+1)} = a_k^{(t)} - \eta \frac{\partial \mathcal{L}}{\partial a_k}.$$

Раздел 1.5.6 посвящен тому что в некоторых случаях для нахождения коэффициентов многочлена используются численные методы, такие как метод наименьших квадратов с использованием QR-разложения или метод Левенберга-Марквардта. Эти методы обычно используются для более сложных случаев, когда требуется высокая точность или работа с большими объемами данных.

Раздел 2 посвящен сетям прямого распространения и их исследованию.

Цель исследования — показать, что нейронные сети прямого распространения могут эффективно аппроксимировать как алгебраические многочлены, так и их производные. Задается функция:

$$P(x) = \sum_{k=0}^n a_k x^k,$$

и требуется построить сеть $f_\theta(x)$ такую, что:

$$\|P(x) - f_\theta(x)\|_\infty < \varepsilon, \quad \left\| \frac{dP}{dx} - \frac{df_\theta}{dx} \right\|_\infty < \delta.$$

Раздел 2.1 посвящен тому что для решения многих прикладных задач используется способность нейронных сетей аппроксимировать полиномиальную функцию. В статье [3] полиномиальная функция используется как мера емкости искусственной нейронной сети прямого распространения с конечным числом скрытых узлов, при этом под емкостью сети понимается наибольшая степень полиномов, которые данная нейронная сеть приближает с наперед заданной точностью.

Рассматривается архитектура нейронной сети F с одним скрытым слоем.

Рассматривается также следующая лемма: Производная порядка s функции F существует в каждой точке $x \in \mathbb{R}$, при этом справедливо равенство

$$\frac{d^s F(x)}{dx^s} = \sum_{j=1}^N c_j w_j^s Q_s(\sigma(\theta_j)), \quad (1)$$

Раздел 2.2 посвящен простейшей архитектуре нейронной сети прямого распространения:

- **Входным слоем:** Один входной нейрон $x \in \mathbb{R}$
- **Скрытым слоем:** n нейронов с функцией активации $\sigma(\cdot)$
- **Выходным слоем:** Один выходной нейрон $y \in \mathbb{R}$

Математическая модель

1. **Вход:** x (скаляр)

2. Скрытый слой:

$$z_j = w_j^{(1)}x + b_j^{(1)} \quad \text{для } j = 1, 2, \dots, n$$
$$a_j = \sigma(z_j) = \sigma\left(w_j^{(1)}x + b_j^{(1)}\right)$$

где:

- $w_j^{(1)}$ – весовые коэффициенты скрытого слоя
- $b_j^{(1)}$ – смещения скрытого слоя
- σ – функция активации (ReLU, sigmoid, tanh и др.)

3. Выходной слой:

$$y = \varphi\left(\sum_{j=1}^n w_j^{(2)}a_j + b^{(2)}\right)$$
$$= \varphi\left(\sum_{j=1}^n w_j^{(2)}\sigma\left(w_j^{(1)}x + b_j^{(1)}\right) + b^{(2)}\right)$$

где:

- $w_j^{(2)}$ – весовые коэффициенты выходного слоя
- $b^{(2)}$ – смещение выходного нейрона
- φ – функция активации выходного слоя

Параметры сети

Общее количество параметров (весов и смещений):

$$\text{Параметры} = \underbrace{2n}_{\text{скрытый слой}} + \underbrace{(n+1)}_{\text{выходной слой}} = 3n + 1$$

Особенности

- Информация распространяется строго от входа к выходу (без обратных связей)
- Способна аппроксимировать любую непрерывную функцию (теорема универсальной аппроксимации)
- Обучение осуществляется методом обратного распространения ошибки
- Вычислительная сложность: $O(n)$ операций на один проход

В разделе 2.2 рассматривается сеть следующего вида:

$$f(x) = \sum_{i=1}^m \alpha_i \sigma(w_i x + b_i),$$

где σ — функция активации, m — количество нейронов скрытого слоя. Цель — аппроксимировать полином и его производную:

$$f(x) \approx P(x), \quad f'(x) \approx P'(x).$$

Пусть $\sigma \in C^1(\mathbb{R})$ и не является многочленом. Тогда для любого $P(x)$ существует такая сеть $f(x)$, что:

$$\sup_{x \in [a, b]} |P(x) - f(x)| < \varepsilon, \quad \sup_{x \in [a, b]} |P'(x) - f'(x)| < \delta.$$

Аналогично, если $P(x_1, \dots, x_n)$ — многочлен от нескольких переменных, то существует сеть:

$$f(\mathbf{x}) = \sum_{i=1}^m \alpha_i \sigma(\mathbf{w}_i^\top \mathbf{x} + b_i),$$

которая аппроксимирует $P(\mathbf{x})$ и его градиент $\nabla P(\mathbf{x})$ с произвольной точностью.

В разделе 2.3 рассматривается применение методов для вычисления коэффициентов многочленов с использованием Python. Для этого будут реализованы следующий метод:

1. Аппроксимация многочлена с помощью нейронной сети прямого распространения сигнала

Для аппроксимации многочлена $P(x) = \sum_{k=0}^r a_k x^k$ степени r используется нейронная сеть прямого распространения вида:

$$F(x) = \sum_{j=1}^N c_j \sigma(w_j x + b_j),$$

где $\sigma(\theta) = \theta / (\alpha + |\theta|)$ — сигмоидальная функция с параметром α , $w_j = (j-1)\delta$ — весовые коэффициенты, b_j — параметры смещения, c_j — искомые коэффи-

циенты сети.

Заключение

В ходе проведенного исследования были системно изучены аппроксимативные свойства нейронных сетей прямого распространения сигнала, что позволило получить ряд значимых теоретических и практических результатов. Теоретический анализ подтвердил фундаментальную способность таких сетей к аппроксимации.