Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра теории функций и
стохастического анализа

# ПРИМЕНЕНИЕ МЕТОДА ФРАНКА-ВУЛЬФА В ИЗУЧЕНИИ ЕМКОСТИ НЕЙРОННОЙ СЕТИ ПРЯМОГО РАСПРОСТРАНЕНИЯ СИГНАЛА

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

Студентки 2 курса 218 группы
направления 01.04.02 — Прикладная математика и информатика
механико-математического факультета
Сысуевой Светланы Дмитриевны

Научный руководитель
зав. каф., д. ф.-м. н., доцент ____________ С. П. Сидоров

Заведующий кафедрой
д. ф.-м. н., доцент ____________ С. П. Сидоров

Саратов 2018

# ВВЕДЕНИЕ

Выпускная квалификационная работа посвящена изучению ёмкости нейронной сети прямого распространения сигнала. Под ёмкостью понимается наибольшая степень многочленов, которые искусственная нейронная сеть может приближать с любой наперед заданной точностью. Основным подходом к изучению ёмкости в данной работе является анализ разреженной структуры сети, получаемой в ходе обучения методом Франка-Вульфа.

**Актуальность исследования.** Вопрос о ёмкости нейронной сети является актуальным в наше время. Ограниченность ресурсов, стремление к увеличению скорости, эффективности работы сетей, увеличение объема обрабатываемых данных влечет за собой необходимость оптимизации используемых моделей нейронных сетей. Использование большого количества нейронов может не только оказаться бесполезной тратой ресурсов, но и даже навредить важному качеству нейронной сети - способности к обобщению. Оптимизация структуры нейронной сети, и, в частности, количества скрытых нейронов, положительно влияет на проблему „перетренированности“ сети.

**Степень изученности проблемы.** Исследование искусственных нейронных сетей в последние десятилетия получило большую популярность из-за широкого спектра возможных областей применения. Вопросу о применении нейронных сетей для аппроксимации функций в различных областях посвящали свои работы множество авторов, например, А. Пинкус, К. Хорник, А. Баррон и др.

Аппроксимационные свойства искусственных нейронных сетей могут зависеть от различных факторов, таких как структура сети (количество скрытых слоёв, нейронов, функции активации) или класса функции, которую приближает нейронная сеть. Различные функции активации нейронных рассматривали Г. Цибенко, Дж. Парк, С. Чуи. В работах Н. Ханма, Б. Лланаса, З. Б. Ксу изучается аппроксимация непрерывных функций с помощью нейронных сетей с сигмоидальными функциями активации. Известно, что нейронная сеть, не имеющая скрытых слоев, не способна приближать любую непрерывную функцию. Этот факт был доказан Б. Уидроу. В большинстве научных работ рассматриваются нейронные сети с одним скрытым слоем. Для таких сетей получены важные теоретические и практические результаты по изучению ёмкости, которые можно найти в работах Б. Малакути, Ф. Скарселли, Н. С. Узенцовой и др. Нейронные сети с двумя и более скрытыми слоями рассматриваются гораздо реже, поэтому теоретического материала по таким нейронным сетям находится немного.

**Цель исследования.** Целью работы является изучение на практике ёмкости нейронной сети прямого распространения сигнала.

**Основные задачи:**

1. Получить основные сведения о нейронных сетях, их построению, обучению и применению
2. Изучить литературу по вопросу о ёмкости нейронных сетей
3. Провести эксперименты в Matlab по приближению многочлена нейронными сетями с различным количеством нейронов
4. Ознакомиться с методом Франка-Вульфа, его особенностями
5. Изучить ёмкость нейронной сети, получив разреженную структуру с помощью обучения сети методом Франка-Вульфа
6. Применить построенную модель к задаче нахождения зависимости между двумя криптовалютами и оценить её эффективность

**Методы исследования.** В работе используются следующие методы изучения ёмкости нейронной сети как комплексно, так и самостоятельно:

— анализ литературных источников
— математическое моделирование (эксперимент)
— описание результатов
— графический анализ

**Научная новизна.** Новая идея, предложенная в данной работе, заключается в изучении ёмкости нейронной сети через получение её разреженной структуры методом Франка-Вульфа.

**Практическая и научная значимость.** Построенные модели однослойной нейронной сети, обученной методом Франка-Вульфа, могут быть адаптированы к моделям многослойных нейронных сетей и применены для изучения их аппроксимационных свойств. Помимо этого, модель может быть самостоятельно применена к реальным задачам для установления характера зависимостей между различными факторами по количеству активных скрытых нейронов.

**Апробация результатов.** Результаты работы были представлены на студенческой научной конференции механико-математического факультета, апрель 2018, Саратов, Саратовский государственный университет. Кроме того, результаты приняты в печать в журнал "Известия Саратовского университета" в рамках VIII международной научной конференции "Компьютерные науки и информационные технологии".

**Структура работы.** Работа состоит из пяти разделов. В разделе 1 излагаются основные сведения о нейронных сетях, необходимые для дальнейшего изучения их свойств. В разделе 2 приводятся теоремы о ёмкости однослойной нейронной сети прямого распространения сигнала. В подразделе 2.1 данного раздела рассматриваются теоретические результаты для экспоненциальной сигмоиды, взятой в качестве функции активации скрытого слоя нейронной сети. Аналогичные результаты, но для рациональной сигмоиды, приведены в подразделе 2.2. В разделе 3 описывается вычислительные эксперименты, в ходе которых были получены результаты приближения алгебраического многочлена однослойной и двуслойной нейроными сетями с различным количеством скрытых нейронов и различными параметрами обучения нейронной сети. В разделе 4 описывается применение алгоритма Франка-Вульфа в качестве метода обучения нейронной сети. В подразделе 4.1 приводится общее описание алгоритма и его основные характеристики. В подразделе 4.2 описывается, как алгоритм Франка-Вульфа может быть адаптирован к обучению однослойной нейронной сети прямого распространения, и излагается общая суть к построению сети, её обучению и изучению ёмкости. В подразделе 4.3 приведены результаты работы описанного метода для приближения однослойной нейронной сетью многочленов различных степеней. Раздел 5 посвящен применению построенной модели однослойной нейронной сети, обученной методом Франка-Вульфа, к задаче нахождения зависимости между двумя криптовалютами. Подраздел 5.1 описывает особенности применения нейронных сетей в сфере экономики и, в частности, в аппроксимации зависимостей между экономическими факторами. Подраздел 5.2. содержит общую информацию об изучаемых криптовалютах - биткоин и биткоин кэш - и характере взаимосвязи между ними. В подразделе 5.3. приводится описание эксперимента по нахождению зависимости между данными криптовалютами нейронной сетью с разреженной структурой и оцениваются результаты проведенных исследований.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В разделе 1 излагаются основные сведения о нейронных сетях, необходимые для дальнейшего изучения их свойств. В подразделе 1.1 вводятся определения нейронной сети и её приложений, в том числе как аппроксиматора функций. В подразделе 1.2 представлена классификация нейронных сетей, определяется многослойная нейронная сеть прямого распространения - сеть, у которой сигнал распространяется по слоям, от входных слоев к скрытым и к выходным. В таких сетях информация с последующих слоев не передается на предыдущие.

Подраздел 1.3 описывает основные элементы искусственной нейронной сети и их математическую формализацию. Функционирование многослойного персептрона выполняется в соответствии с формулами:

$$\theta_j^{[k]} = \sum_{i=1}^{N_{k-1}} w_{ji}^{[k]} y_i^{[k-1]} + b_j^{[k]}, \quad j = 1, \ldots, N_k, \quad k = 1, \ldots, L;$$

$$y_j^{[k]} = \sigma(\theta_j^{[k]}), \quad j = 1, \ldots, N_k, \quad k = 1, \ldots, L-1,$$

$$y_j^{[L]} = \theta_j^{[L]},$$

где

— $y_i^{[k-1]}$ — выходной сигнал $i$–го нейрона $(k-1)$-го слоя;
— $w_{ji}^{[k]}$ — вес связи между $j$-м нейроном слоя $(k-1)$ и $i$-м нейроном $k$-го слоя;
— $b_j^{[k]}$ — значение смещения $j$-го нейрона $k$-го слоя;
— $y = \sigma(\theta)$ — функция активации;
— $y_j^{[k]}$ — выходной сигнал $j$-го нейрона $k$-го слоя;
— $N_k$ — число узлов слоя $k$;
— $L$ — общее число основных слоев;
— $n = N_0$ — размерность входного вектора;
— $m = N_L$ — размерность выходного вектора сети.

В подразделе 1.4 описывается процесс обучения нейронной сети с учителем. Обучение основано на обработке нейронной сетью примеров из обучающей выборки. В задаче аппроксимации функции обучающим примером является $(\tilde{x}, \tilde{y})$, $\tilde{x} \in R^n$, $\tilde{y} \in R^m$: $\tilde{x}$ — аргумент функции, $\tilde{y}$ — значение аппроксимируемой (приближаемой) функции в точке $\tilde{x}$.

Обучающую выборку рекомендуется разбивать на данные обучения, тестирования и валидации. Для сходимости модели к целевой функции алгоритм обучения нейронной сети использует подмножество данных обучения. Подмножество данных тестирования используется для предотвращения ее перетренированности, что, в конечном итоге, может повлиять на надежность модели. После того как модель нейронной сети обучена, она должна пройти через жесткий процесс валидации, что даст уверенность в надежной работе в реальной среде.

В разделе 2 приводятся теоремы о ёмкости однослойной нейронной сети прямого распространения сигнала.

Подраздел 2.1 содержит несколько вспомогательных утверждений, а также теорему о ёмкости для нейронной сети с экспоненциальной сигмоидой и следствие из нее. Рассматривается нейронная сеть $F : \mathbb{R} \to \mathbb{R}$ прямого распространения с одним входом и одним выходом:

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

где $c_j, w_j, b_j \in \mathbb{R}$, $j = 1, \ldots, N$, есть параметры сети. Обозначим множество всех таких отображений $\mathcal{F}_N(\sigma)$.

Полиномиальная функция $f(x) : \mathbb{R} \to \mathbb{R}$ степени $r$ имеет вид:

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_r x^r,$$

где $a_i \in \mathbb{R}$, $i = 0, 1, \ldots, r$.

**Теорема 1.** Пусть $x_{\max} > 0$ – произвольное число. Тогда однослойная нейронная сеть прямого распространения сигнала $F(x)$ с $N$ скрытыми нейронами может приближать любой алгебраический многочлен $f(x)$ степени $N-1$ с любой заданной точностью, где $|x| < x_{max}$. Другими словами, можно построить нейронную сеть $F(x)$ такую, что:

$$|F(x) - f(x)| < \varepsilon,$$

где $\varepsilon$ является максимальным значением для ошибки приближения на $[-x_{\max}, x_{\max}]$.

**Следствие 1.** Если количество скрытых узлов $N \geq 8$, то ёмкость многослойной нейронной сети прямого распространения сигнала больше ёмкости однослойной нейронной сети.

В подразделе 2.2 приведена аналогичная теорема для нейронной сети с рациональной сигмоидой.

Раздел 3 описывает вычислительные эксперименты, в ходе которых были получены результаты приближения алгебраического многочлена 9-той степени однослойной и двуслойной нейроными сетями с различным количеством скрытых нейронов и различными параметрами обучения нейронной сети. В качестве среды для программирования использовалась система Matlab и её библиотека Neural Network Toolbox.

Входные данные обучающей выборки генерировались рандомным образом в рассматриваемом диапазоне. Выходными данными являлись значения указанного выше многочлена в данных точках. В качестве функции активации скрытых слоев рассматривалась рациональная сигмоида. В качестве функции ошибки использовалась среднеквадратическая ошибка по всей выборке.

В первой части эксперимента количество скрытых нейронов однослойной нейронной сети варьировалось от 8 до 11. Количество итераций обучения менялось 300 до 3000 с шагом 300 итераций

Результаты работы нейронных сетей показали, что, во-первых, нейронная сеть с 8 нейронами, в отличие от остальных сетей, не смогла достигнуть требуемой точности. Во-вторых, с увеличением количества скрытых нейронов количество итераций обучения, необходимых для достижения требуемой точности, преимущественно уменьшалось. В целом, можно утверждать, что полученные практические результаты не противоречат теореме о ёмкости однослойной нейронной сети.

Во второй части проводились аналогичные эксперименты для двуслойных нейронных сетей. Целью данного эксперимента являлась проверка гипотезы о том, что при приближении алгебраического многочлена степени $r$ двуслойной нейронной сетью достаточно $N \leq r$ нейронов для достижения любой наперед заданной точности. Для эксперимента использовались следующие варианты количества нейронов:

$$NN_{2l} = \{[5,5],[5,4],[4,4],[5,3],[4,3],[3,3]\}.$$

Эксперименты показали, что двуслойные сети с числом скрытых нейронов $N \geq 6$ могут достигать точности $\epsilon = 0.001$ при приближении многочлена 9 степени. При требуемой точности в 0.0001 нейронные сети с количеством нейронов $[5,5]$, $[5,4]$ и $[5,3]$ смогли уменьшить максимальную ошибку до данного уровня при большом количестве итераций. В отличие от однослойной сети двуслойная сеть с 8 скрытыми нейронами достигла требуемой точности. При $N < 8$ точность приближения двуслойной нейронной сетью не была достигнута, даже при попытке увеличить максимальное число итераций. При распределении нейронов $[4,4]$ ошибку также не удалось уменьшить до значения, меньшего 0.0001. Из полученных результатов можно сделать вывод, что двуслойная нейронная сеть с числом нейронов, меньших 10, дает такую же ошибку приближения, что и однослойная нейронная сеть с 10 нейронами. Таким образом, многослойные нейронные сети, возможно, обладают большей емкостью по сравнению с однослойными. Вопрос о конструктивном алгоритме поиска структуры сети и оптимальных значений весов сети оставлен за рамками настоящей работы.

В разделе 4 описывается подход к изучению ёмкости искусственной нейронной сети, заключающийся в получении и анализе разреженной структуры нейронной сети. Термин "разреженная структура" означает, что большое количество скрытых нейронов имеет нулевые или близкие к нулю параметры. Будем называть такие нейроны *неактивными*. Нейроны, которые имеют хотя бы один отличный от нуля параметр, будет называть *активными*. По разреженной структуре нейронной сети можно сделать вывод о том, сколько нейронов достаточно для приближения многочлена определенной степени.

Как было описано ранее, параметры искусственной нейронной сети меняются в процессе обучения, а обучение нейронной сети, как известно, сводится к задаче выпуклого программирования. Одним из методов оптимизации, применяющихся для решения задач выпуклого программирования и позволяющих получить разреженные решения, является метод Франка-Вульфа. Метод Франка-Вульфа, или метод условного градиента - один из основных методов условной оптимизации. Особенность данного метода состоит в возможности получения разреженных решений, поскольку на каждом шаге метода происходит изменение лишь одного параметра.

В подразделе 4.1 приводится общее описание алгоритма и сходимости метода Франка-Вульфа.

Пусть $A \subset R^n$ - замкнутое выпуклое множество; функция $f : A \to R$ выпуклая и непрерывно дифференциируемая на $A$. Требуется минимизировать функцию $f(x)$:

$$f(x) \to \min_{x \in A} . \tag{1}$$

Краткая последовательность шагов алгоритма Франка – Вульфа:

Шаг 0. Определяют исходное допустимое решение задачи.

Шаг 1. Вычисляют градиент функции $f(x)$.

Шаг 2. Формируют дополнительную линейную целевую функцию $\phi$ по формуле

$$\phi_{k+1} := \left\langle s^k, \nabla f(x^k) \right\rangle , \ s \in A \tag{2}$$

и находят ее максимальное значение на множестве допустимых рещений $A$.

Шаг 3. Определяют шаг вычислений $\gamma$.

Шаг 4. Находят новое допустимое решение:

$$x^{k+1} = x^k + \gamma_k (s^k - x^k), \tag{3}$$

Шаг 5. Проверяют необходимость перехода к следующему допустимому решению и при таковой переходят к шагу 1. В противном случае найдено оптимальное решение задачи.

Возможность получить разреженные решения отражается на скорости сходимости метода. Чтобы достигнуть заданной точности $\epsilon$, требуется совершить следующее количество итераций алгоритма:

$$K > \frac{(diam(A))^2}{\epsilon}.$$

В подразделе 4.2 описывается, как алгоритм Франка-Вульфа может быть применен в качестве метода обучения нейронной сети. Рассматривается однослойная нейронная сеть прямого распространения сигнала, у которой функцией активации является экспоненциальная сигмоида. На вход сети подается одно значение - $x$. В скрытом слое содержится $m$ нейронов - данный параметр варьируется в процессе проведения экспериментов. Выход сети $F(x, V, W, B)$ вычисляется по формуле:

$$F(x, V, W, B) = \sum_{j=1}^{m} v_j \sigma(w_j x + b_j)$$

Сеть имеет $3m$ параметров, оптимальные значения которых должны быть найдены в процессе обучения.

Основываясь на сгенерированных тестовых примерах, при обучении вычисляется среднеквадратическая ошибка:

$$E(V, W, B, M) = \frac{1}{N} \sum_{i=1}^{N} (F(x_i, V, W, B) - y_i)^2. \tag{4}$$

Данная функция ошибки рассматривается как целевая функция в задаче выпуклого программирования. В качестве множества допустимых решений данной задачи берется n-мерный шар радиуса $a$.

Обучение сети проводится согласно алгоритму из подраздела 3.1 при применением формул разностных производных для вычисления градиента. В результате получаем обученную сеть с оптимальными параметрами $(V^*, W^*, B^*)$. Для получения ответа на вопрос о том, сколько нейронов задействовано в полученной сети, достаточно проанализировать значения параметров $(V^*, W^*, B^*)$ и подсчитать количество активных нейронов.

Подраздел 4.3 описывает результаты работы однослойной нейронной сети, обученной методом Франка-Вульфа и аппроксимирующей алгебраические многочлены различных степеней.

Для алгебраического многочлена $x^2 - x$ при 5 скрытых нейронах в сети, точности 0.001, 30 тестовых примерах и шаге $\Delta = 0.0001$, после обучения нейронной сети осталось 3 активных нейрона и следующие параметры $V, W$ и $B$:

V = {-4.88873e-007, 3.87485, -0.724698, 0.117221, 1.29408e-007} ,
W = {0, -1.29685, -1.04871, 0, 0} ,
B = {-3.30708e-007, -0.952678, 0.686039, -1.15029e-007, -1.43786e-007} .

Полученный результат не противоречит теореме о емкости однослойной сети, для приближения многочлена второй степени необходимо построить однослойную нейронную сеть с тремя скрытыми нейронами. Данное утверждение подтвердилось и работой сети с 3 скрытыми нейронами.

Результаты работы программы для многочленов других степеней при приближении их сетью с 10 скрытыми нейронами приведены в таблице 1.

Таблица 1 – Результат применения метода Франка-Вульфа к аппроксимации многочленов однослойной нейронной сетью

| Функция | Кол-во активных нейронов | Кол-во итераций | Точность |
|---|---|---|---|
| $x^3 - x^2 + x$ | 4 | 120000 | 0.00049 |
| $x^4 - x^3$ | 5 | 120000 | 0.00098 |
| $x^5 + x^2 - x$ | 6 | 120000 | 0.00049 |

По полученным для многочленов выше результатам можно сделать выводы, аналогичные рассуждениям выше. При достаточном количестве итераций и не слишком малом $\epsilon$ однослойной нейронной сети для приближения алгебраического многочлена степени $r$ требуется $r + 1$ скрытый нейрон. Остальные нейроны имеют либо нулевые, либо пренебрежительно малые значения всех параметров.

Раздел 5 посвящен вопросу об использовании нейронной сети разреженной структуры как средства для нахождения зависимости между двумя криптовалютами - биткоин и биткоин кэш.

В подразделе 5.1 кратко описывается применение нейронной сети как универсального аппроксиматора в экономике. Нейронная сеть является универсальным аппроксиматором при правильно подобранной структуре, в т. ч. количестве скрытых нейронов. Сеть не должна быть слишком сложной и не должна слишком точно следовать данным из обучающего множества, но при этом она должна достаточно точно улавливать действительную структуру аппроксимируемой функции. Еще одно важное условие применения нейронных сетей заключается в том, что должно быть серьезное предположение о наличии связи между известными входными значениями и неизвестными выходами. Далее рассмотрено применение нейронной сети для определения зависимостей между экономическими показателями – курсами криптовалют, относительно которых имеется предположение касательно характера зависимости.

В подразделе 5.2 даны основные понятия о криптовалютах биткоин, биткоин кэш, их краткая история и наблюдаемые взаимосвязи. Криптовалюта – цифровая или виртуальная валюта, одной единицей которой принято считать coin. Первостепенной криптовалютой стал Bitcoin (BTC), на сегодняшний день эта валюта является самой высокой по стоимости за одну единицу. Котировка биткоин зависит от доверия к нему и формируется балансом спроса и предложения, также она не имеет привязки ни к материальным ценностям, ни к валюте. Кроме основного игрока криптовалютного рынка – биткоина, с успехом функционируют сотни других валют. На основании волатильности BTC, как флагмана рынка можно сделать предположение об изменении курса остальных криптовалют.

Bitcoin Cash (BCH) – независимая децентрализованная криптовалюта, которая существует по тем же принципам, что и классический Биткоин, но при этом обладает меньшей комиссией при расчетах и более высокой скоростью платежей. В целом BCH в технических аспектах практически повторяет BTC. Если внимательно изучить поведение курса валют BTC и BCH, можно заметить определённую зависимость между ними на относительно небольших временных промежутках. С конца декабря 2017 года наблюдается зависимость, близкая к линейной.

В подразделе 5.3. приводится описание эксперимента по нахождению зависимости между BTC и BCH нейронной сетью с разреженной структурой и оцениваются результаты проведенных исследований.

Если зависимость между курсом биткоина (BTC) в USD и курсом биткоин кэш (BCH) выражается многочленом первой степени, то по теореме о ёмкости однослойной нейронной сети достаточно двух скрытых нейронов для приближения этого многочлена с любой заданной точностью. Целью экспериментов являлось не только нахождение минимально допустимого числа скрытых нейронов в однослойной сети, но и анализ поведения сети на большем количестве нейронов. Для достижения данной цели применялся метод Франка-Вульфа как метод обучения сети, позволяющий получить разреженную структуру сети, благодаря которой можно выдвинуть предположение о характере зависимости между величинами по количеству активных нейронов.

Для проведения экспериментов взяты данные о курсах BTC/USD и

BCH/USD за 3 месяца - с 1 января по 31 марта 2018 года с сайта finam.ru. Получены 91 наблюдаемых значений курса для каждой из валют. Все наблюдаемые значения были разделены на данные для обучения сети и данные для валидации.

Созданная ранее программа по построению и обучению однослойной нейронной сети методом Франка-Вульфа была адаптирована под применение к нахождению зависимости между BTC и BCH. Значения курса BTC рассматривались в качестве входных данных, а значения курса BCH - как выходные (зависимые) данные. Затем анализировалась обратная зависимость тем же подходом. Целевой функцией по-прежнему считалась среднеквадратическая ошибка приближения.

Первый эксперимент был проведен для количества скрытых нейронов, равного 10. С указанными параметрами сети алгоритм провел максимальное количество итераций, при этом ошибка сети сократилась с 0.148618 до 0.0122699 на данных обучения. Ошибка сети на валидационных примерах составила 0.068389, то есть, подтвердились предположения о том, что на данных валидации сеть с большим количеством нейронов не будет столь же точна, как на данных обучения. Из десяти скрытых нейронов в нейронной сети осталось 2 активных нейрона. Это говорит в пользу нашего предположения о линейном характере зависимости между курсами BTC и BCH.

Далее были проведены экспернименты для двух скрытых нейронов. По результатам работы программы оба нейрона оказались активными, и ошибка нейронной сети на данных валидации составила 0.00886493, т.е. была достигнута требуемая точность $\epsilon = 0.01$. По всему множеству наблюдаемых данных среднеквадратическая ошибка оказалась чуть больше (0.0102584) за счет ошибки на данных обучения (0.0107154).

В последнем эксперименте в качестве входных данных значения курсов криптовалюты BCH, а в качестве зависимых выходных данных - значения BTC. В скрытом слое нейронной сети оставлены пять нейронов. На рисунке 1 отражены результаты работы нейронной сети для обратной зависимости BTC-BCH. Судя по графику, при перемене местами входных и выходных данных нейронная сеть определяет практически четкую линейную зависимость между этими величинами. При этом в нейронной сети по-прежнему осталось лишь два активных скрытых нейрона. Ошибка сети на всем множестве примеров составила 0.0146293, на множестве валидационных примеров - 0.0156788.

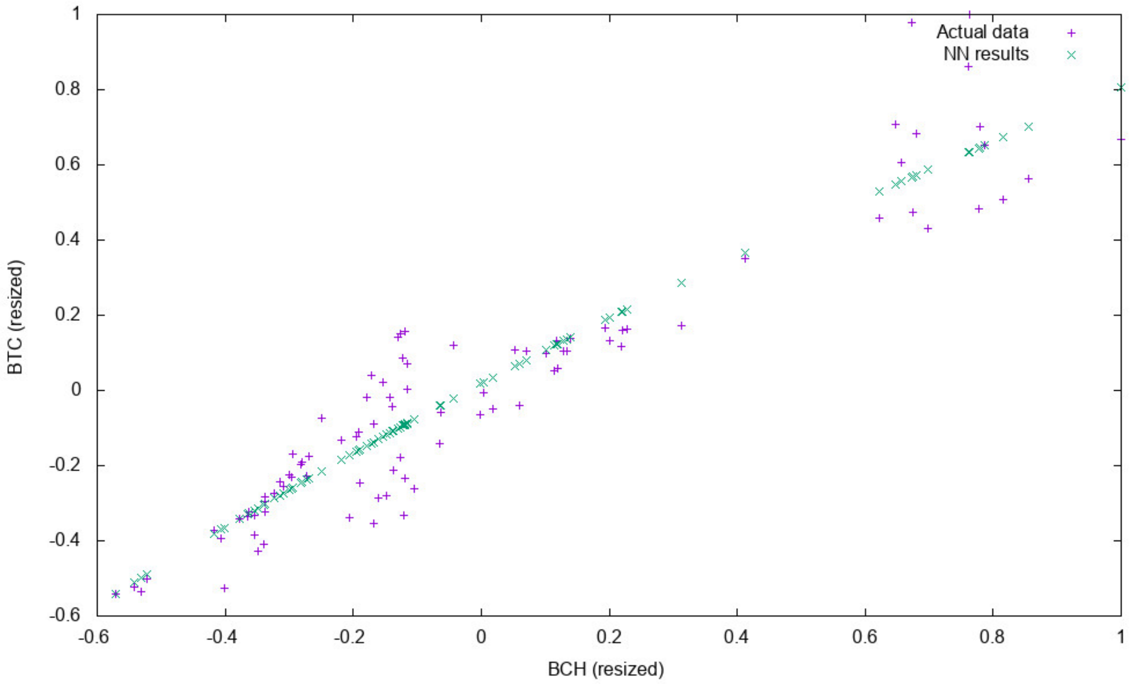


Рисунок 1 – Нахождение обратной зависимости между BCH и BTC нейронной сетью с 5 скрытыми нейронами

Подведем итоги проведенных экспериментов по определению зависимости между курсами BTC и BCH. Во-первых, построенная нами модель нейронной сети показала себя эффективной в применении к реальным экономическим данным. Во-вторых, подтвердилось предположение о линейном характере зависимости между значениями курсов криптовалют BTC и BCH (по крайней мере в рамках рассмотренного диапазона в 3 месяца для последних актуальных данных). В-третьих, удалось на практике продемонстрировать достоинство применения метода Франка-Вульфа в обучении нейронной сети. В результате экспериментов была получена разреженная структура сети, благодаря которой выяснилось минимальное достаточное количество скрытых нейронов для решения конкретно данной задачи. Таким образом, была выявлена оптимальная структура нейронной сети, которая, с одной стороны, способная определять неизвестную зависимость с заданной точностью, а с другой стороны, не теряет обобщающей способности и довольно хорошо приближает данные, находящиеся вне обучающего множества.

# ЗАКЛЮЧЕНИЕ

Аппроксимация алгебраических многочленов, имеющих важное значение и активно применяющихся во многих областях науки, является актуальной задачей, которую можно решать различными методами. Один из способов приближения таких функций - обработка данных с помощью нейронных сетей - представляет большой интерес на сегодняшний момент, поскольку позволяет воспроизвести сложные зависимости будучи достаточно простым в применении по сравнению с другими методами. Для оптимизации работы такого инструмента, как нейронная сеть, можно прибегать к различным методам и модификациям.

В данной работе вопрос об оптимизации нейронной сети был изучен со стороны архитектуры сети касательно количества слоёв и скрытых нейронов. Рассматривались искусственные однослойные и двуслойные нейронные сети прямого распространения сигнала. Теоретические результаты, приведенные в первом разделе, показывают, что для аппроксимации алгебраического многочлена степени $r$ однойслойной нейронной сетью с любой наперед заданной точностью необходимо $N \geq r+1$ нейронов в скрытом слое. Основываясь на следствии из теоремы о ёмкости однослойной нейронной сети было введено предположение о том, что для приближения алгебраического многочлена степени $r$ может быть достаточно $N \leq r$ скрытых нейронов в многослойной нейронной сети.

Проведенные вычислительные эксперименты по изучению ошибок приближения многочлена однослойной нейронной сетью с различным количеством скрытых нейронов подтвердили существующую теорию о ёмкости однослойной нейронной сети прямого распространения сигнала. Аналогичные эксперименты для двуслойной нейронной сети показали, что двуслойная нейронная сеть достигает той же точности приближения алгебраического многочлена, что и однослойная нейронная сеть, но при меньшем количестве нейронов. Это в свою очередь косвенно подтверждает теоретические выводы о ёмкости многослойной нейронной сети. Можно утверждать, что результаты проведенного вычислительного эксперимента не противоречат выдвинутой нами гипотезе о количестве нейронов, необходимых для достижения заданной точности.

Вопрос об емкости однослойной нейронной сети также был изучен с помощью другого подхода. Обучение нейронной сети было рассмотрено как задача выпуклого программирования, и решалась эта задача с использованием метода Франка-Вульфа. Несмотря на явный недостаток данного метода - медленную сходимость, при проведении эксперимента удалось продемонстрировать одно из его достоинств. Результаты экспериментов показали, что применение метода Франка-Вульфа позволило получить разреженную структуру нейронной сети, при которой часть скрытых нейронов имеют нулевые или пренебрежительно малые коэффициенты. Это означает, что такие нейроны не влияют значительно на результат работы нейронной сети и их можно отбросить. После такого сокращения было получено наименьшее количество скрытых нейронов, необходимых для приближения алгебраических многочленов с заданной точностью. Согласно полученным результатам, для приближения алгебраического многочлена $r$-той степени однослойной нейронной сетью требуется $r + 1$ скрытый нейрон. Данные результаты не противоречат имеющейся теории о емкости однослойных нейронных сетей, а значит, построенные модели могут применяться к изучению емкости и многослойных нейронных сетей.

Построенную модель однослойной нейронной сети удалось применить на практике для определения зависимости курсов двух криптовалют - биткоин и биткоин кэш. В работе экспериментально установлено, что нейронная сеть с избыточным количеством нейрона теряет часть своих обобщающих свойств. Применение метода Франка-Вульфа, как и в случае приближения многочлена, позволило получить разреженную структуру сети, благодаря чему было выяснено минимальное достаточное количество скрытых нейронов и подтвержден линейный характер изучаемой зависимости.

Подводя итоги всей работы, можно говорить об успешности проведенных исследований по изучению ёмкости нейронной сети и эффективности применения метода Франка-Вульфа к получению разреженной структуры сети. Результаты работы могут быть использованы как при изучении ёмкости многослойных нейронных сетей, так и при оптимизации нейронных сетей для нахождения зависимостей между различными факторами в реальном мире.