

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«Тверской государственный университет»

Факультет прикладной математики и кибернетики

Направление 02.03.02 - Фундаментальная информатика и информационные  
технологии

Профиль «Информатика и компьютерные науки»

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА  
(ВЫПУСКНАЯ РАБОТА БАКАЛАВРА)**

Тема: «Эволюционные алгоритмы в нейронных сетях»

**Автор:**

Пурис Дмитрий Николаевич

**Научный руководитель:**

к.ф.-м.н., доцент

Дадеркин Дмитрий Ольгердович

Допущен к защите:

Руководитель ООП:

\_\_\_\_\_ /А.В. Язенин/

«\_\_\_\_\_» \_\_\_\_\_ 2019 г.

Тверь — 2019

# Оглавление

<b>Введение</b>	<b>4</b>
Актуальность . . . . .	4
Цели и задачи работы . . . . .	4
Основные результаты . . . . .	5
Структура работы . . . . .	5
<b>1 Нейронные сети</b>	<b>6</b>
1.1 Применение нейронных сетей . . . . .	6
1.2 Основные определения . . . . .	6
1.3 Виды нейронных сетей . . . . .	7
1.4 Обучающие алгоритмы . . . . .	9
1.5 Перцептрон . . . . .	10
1.5.1 Квадратичная целевая функция . . . . .	12
1.5.2 Обучение линейного нейрона . . . . .	13
1.6 Сигмоидальный нейрон . . . . .	14
1.6.1 Вывод правил обучения для логистического нейрона . . . . .	15
1.7 Стохастический градиентный спуск . . . . .	16
1.8 Многослойный перцептрон Румельхарта . . . . .	17
1.8.1 Алгоритм обратного распространения ошибки . . . . .	18
<b>2 Эволюционные алгоритмы</b>	<b>21</b>
2.1 Виды эволюционных алгоритмов . . . . .	21
2.2 Понятие естественного отбора . . . . .	22
2.3 Применение эволюционных алгоритмов в нейронных сетях . . . . .	25
<b>3 Разработанное ПО</b>	<b>27</b>
3.1 Требования к программному обеспечению . . . . .	27

3.2	Выбор средств для создания ПО . . . . .	27
3.3	Описание созданного ПО . . . . .	27
3.4	Результаты тестирования ПО . . . . .	31
	<b>Заключение</b>	<b>33</b>
	<b>Литература</b>	<b>34</b>
	<b>Приложение</b>	<b>36</b>

# Введение

## Актуальность

В настоящее время существует ряд задач, для решения которых сложно или невозможно создать алгоритм в традиционном виде. В таком случае используют нейронные сети. Главное преимущество нейронных сетей — это обучаемость. В процессе обучения искусственная нейронная сеть способна выявлять сложные зависимости между входными и выходными данными, а также выполнять обобщение. Это значит, что в случае успешного обучения сеть сможет вернуть правильный результат на основании данных, которые отсутствовали в обучающей выборке, а также неполных и/или «зашумленных», частично искажённых данных.

Тема искусственного интеллекта на сегодняшний день является очень актуальной. В наши дни создаются основополагающие концепции и алгоритмы, связанные с машинным обучением. Сегодня существует определенный класс актуальных задач, решение которых без применения искусственных нейронных сетей (ИНС) невозможно или трудноосуществимо. Зачастую сюда относятся такие задачи, как классификация, прогнозирование, автоматизация процесса распознавания и автоматизация процесса принятия решений.

## Цели и задачи работы

### Цель работы

Разработать алгоритм, реализующий работу эволюционного алгоритма в нейронных сетях.

## Задачи работы

- изучить обучающие алгоритмы;
- изучить виды нейронных сетей и методы их обучения;
- изучить применение эволюционных алгоритмов в нейронных сетях;
- разработать алгоритм, реализующий работу эволюционного алгоритма в нейронных сетях;

## Основные результаты

Во время создания выпускной квалификационной работы были достигнуты следующие результаты:

- изучены обучающие алгоритмы;
- изучены виды нейронных сетей и методы их обучения;
- изучено применение эволюционных алгоритмов в нейронных сетях;
- разработана нейронная сеть, обучающаяся с помощью эволюционного алгоритма, которая способна решать задачу классификации.

## Структура работы

В первой главе ВКР рассмотрены нейронные сети, их виды, применение и способы обучения. Даны основные понятия и определения. Приведены различные виды обучающих алгоритмов и их применение. Во второй главе рассмотрены эволюционные алгоритмы и применение их в нейронных сетях. В третьей главе рассказывается про разработанное программное обеспечение, про выбор средств разработки, а также про разработанный эволюционный алгоритм и алгоритм скрещивания.

# Глава 1

## Нейронные сети

### 1.1 Применение нейронных сетей

Нейронные сети используются для решения сложных задач, которые требуют аналитических вычислений подобных тем, что делает человеческий мозг. Ниже приведены самые основные задачи, которые решают нейронные сети.

**Классификация** — это распределение данных по параметрам. Задано конечное множество объектов, для которых известно, к каким классам они относятся. Это множество называется выборкой. Классовая принадлежность остальных объектов неизвестна. Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

**Предсказание** — возможность предсказывать следующий шаг. Например, рост или падение акций, основываясь на ситуации на фондовом рынке.

**Распознавание** — научное направление, связанное с разработкой принципов и построением систем, предназначенных для определения принадлежности данного объекта к одному из заранее выделенных классов объектов.

### 1.2 Основные определения

В данном разделе даны основные определения, связанные с нейронными сетями и эволюционными алгоритмами.

**Алгоритм** — Совокупность последовательных шагов, схема действий, приводящих к желаемому результату.

**Генетический алгоритм** — это эвристический алгоритм поиска, используемый для решения задач оптимизации и моделирования путём слу-

чайного подбора, комбинирования и вариации искомых параметров с использованием механизмов, аналогичных естественному отбору в природе.

**Популяция** — это конечное множество особей.

**Нейрон** — узел искусственной нейронной сети, являющийся упрощённой моделью естественного нейрона. Математически, искусственный нейрон обычно представляют как некоторую нелинейную функцию от единственного аргумента — линейной комбинации всех входных сигналов. Данную функцию называют функцией активации. Полученный результат посылается на единственный выход. Такие искусственные нейроны объединяют в сети — соединяют выходы одних нейронов с входами других.

**Искусственная нейронная сеть (ИНС)** — математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма. После разработки алгоритмов обучения получаемые модели стали использовать в практических целях: в задачах прогнозирования, для распознавания образов, в задачах управления и т.д.

**Эпоха** — одна итерация в процессе обучения, включающая предъявление всех примеров из обучающего множества и, возможно, проверку качества обучения на контрольном множестве.

**Алгоритм обратного распространения ошибки (Back propagation)** — один из методов обучения многослойных нейронных сетей прямого распространения, называемых также многослойными перцептронами.

## 1.3 Виды нейронных сетей

Нейронные сети прямого распространения (**feed forward neural networks, FF** или **FFNN**) и перцептроны (**perceptrons**) имеют прямолинейное распространение, они передают информацию от входа к выходу. Это означает, что в сети нет петель — информация всегда передается, никогда не возвращается. Нейроны одного слоя не связаны между собой, а соседние слои обычно полностью связаны. FFNN обычно обучается по методу обратного распространения ошибки, в котором сеть получает множества входных и выходных данных.

Однако есть и другие модели искусственных нейронных сетей, в которых возможны петли обратной связи. Эти модели называются **рекуррентными нейронными сетями**. Рекуррентная нейронная сеть (RNN) - это класс искусственной нейронной сети, где соединения между узлами образуют ориентированный граф вдоль временной последовательности. В отличие от нейронных сетей с прямой связью, RNN могут использовать свое внутреннее состояние (память) для обработки последовательностей входных данных. Это делает их применимыми к таким задачам, как несегментированное распознавание рукописного ввода с подключением или распознавание речи.

**Нейронная сеть Хопфилда (Hopfield network, HN)** является формой рекуррентной искусственной нейронной сети, популяризированной Дж. Хопфилдом в 1982 году. Сети Хопфилда служат в качестве ассоциативных систем памяти с двоичными пороговыми узлами. Они гарантированно сходятся к локальному минимуму и, следовательно, могут сходиться к ложному шаблону (неправильный локальный минимум), а не к сохраненному шаблону (ожидаемый локальный минимум). Сети Хопфилда также предоставляют модель для понимания человеческой памяти.

**Цепи Маркова (Markov chains, MC или discrete time Markov Chains, DTMC)** — это стохастическая модель, описывающая последовательность возможных событий, в которой вероятность каждого события зависит только от состояния, достигнутого в предыдущем событии.

**Свёрточные нейронные сети (convolutional neural networks, CNN)** являются регуляризованными версиями многослойных персептронов. Многослойные персептроны обычно относятся к полностью соединенным сетям, то есть каждый нейрон в одном слое связан со всеми нейронами в следующем слое. «Полная связность» этих сетей делает их склонными к перегрузке данных. Типичные способы регуляризации включают добавление некоторой формы измерения весов к функции потерь. Однако CNN используют другой подход к регуляризации: они используют преимущества иерархического шаблона в данных и собирают более сложные шаблоны, используя меньшие и более простые шаблоны. Таким образом, по шкале связанности и сложности CNN находятся на более низком уровне.



## 1.4 Обучающие алгоритмы

### Линейная регрессия

Линейная регрессия — используемая в статистике регрессионная модель зависимости одной переменной  $y$  от другой или нескольких других переменных (факторов, регрессоров, независимых переменных)  $x$  с линейной функцией зависимости. Линейная регрессия бывает двух типов: простая линейная регрессия и множественная линейная регрессия. Простая линейная регрессия характеризуется одной независимой переменной. Множественная линейная регрессия характеризуется множеством независимых переменных.

### Логистическая регрессия

Логистическая регрессия — это статистическая модель, используемая для прогнозирования вероятности возникновения некоторого события путём подгонки данных к логистической кривой. Он используется для оценки дискретных значений (двоичные значения, такие как 0/1, да/нет, истина/ложь) на основе заданного набора независимых переменных. Этот алгоритм предсказывает вероятность возникновения события путем подгонки данных к функции  $\text{logit}$ . Следовательно, это также известно как регрессия логита. Поскольку он предсказывает вероятность, его выходные значения лежат между 0 и 1.

### Наивная байесовская классификация

Наивная байесовская классификация — это метод классификации, основанный на теореме Байеса с предположением независимости между предикторами. Наивный байесовский классификатор предполагает, что наличие определенной функции в классе не связано с наличием любой другой функции. Байесовский подход к классификации основан на теореме, утверждающей, что если плотности распределения каждого из классов известны, то искомый алгоритм можно выписать в явном аналитическом виде. Более того, этот алгоритм оптимален, то есть обладает минимальной вероятностью ошибок.

## 1.5 Перцептрон

Перцептроны передают информацию от входа к выходу. Нейронные сети часто описываются в виде слоев нейронов, где каждый слой состоит из входных, скрытых или выходных клеток. Клетки одного слоя не связаны между собой, а соседние слои обычно полностью связаны (рис. 1.1).

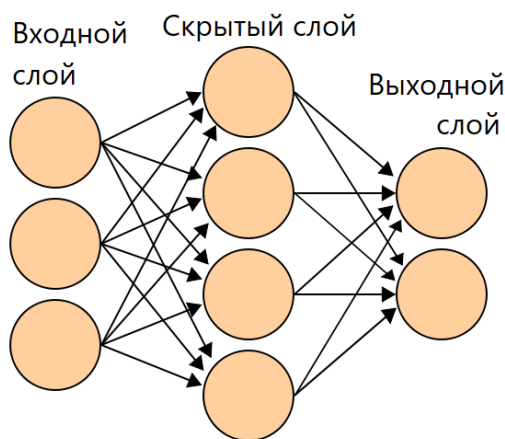


Рис. 1.1: Нейронная сеть

Самая простая нейронная сеть имеет две входных клетки и одну выходную, и может использоваться в качестве модели логических вентилях. Нейронная сеть прямого распространения обычно обучается по методу обратного распространения ошибки, в котором модель получает множества входных и выходных данных. Этот процесс называется обучением с учителем, и он отличается от обучения без учителя тем, что во втором случае множество выходных данных сеть составляет самостоятельно.

Модель перцептрона была предложена и популяризована Ф. Розенблаттом. В начале 1960-х Ф. Розенблатт верил в свою модель, и надеялся, что с помощью перцептрона человечество продвинется в создании искусственного интеллекта. В те времена была большая надежда на эту модель. Но в 1969 году М. Минский и С. Паперт выпустили книгу, в которой рассматривали вычислительные способности перцептронов – то, чему могут учиться и чему не могут учиться перцептроны. И в том числе показывали некоторые ограничения модели. В итоге критика, которая заключалась в доказательстве ограничения такой модели как перцептрон была обобщена на все нейросетевые модели и привела к застою в исследовании данной области на долгие годы.

Обучение перцептрона заключается в том, что ему «показывают» примеры и правильные ответы для этих примеров. Сперва веса перцептрона инициализируются случайно. И поэтому перцептрон, когда ему на вход поступает какой-нибудь пример, дает случайный ответ.

Перцептрон (линейный нейрон) – это оператор вида:

$$f(\mathbf{x}, \mathbf{w}, b) = \begin{cases} 1, & \text{если } \sum_{i=1}^n \mathbf{w}_i \mathbf{x}_i + b > 0 \\ 0, & \text{иначе.} \end{cases} \quad (1)$$

$f(\mathbf{x}, \mathbf{w}, b)$  - активационная функция;

$\sum_{i=1}^n \mathbf{w}_i \mathbf{x}_i + b > 0$  - сумматорная функция;

$\mathbf{x}$  - вектор входных активаций;

$\mathbf{w}$  - вектор весов;

$b$  - смещение.

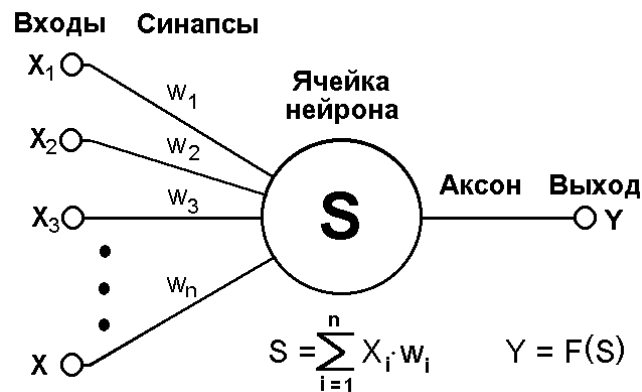


Рис. 1.2: Нейрон

Модель перцептрона принимает несколько вещественных входных данных и дает один двоичный выход. В модели перцептрона каждый вход  $x_i$  имеет вес,  $w_i$  связанный с ним. Веса указывают на важность ввода в процессе принятия решений. Выход модели определяется сумматорной функцией, если сумматорная функция больше нуля, то выход будет равен 1, иначе выход будет равен 0. Другими словами, модель будет срабатывать, если взвешенная сумма больше порога. Такой линейный нейрон способен решать лишь ряд простых задач, для которых хватает линейной функции, например, предсказание стоимости квартиры от ее площади.

К неприятным особенностям линейного нейрона можно отнести то, что небольшое изменение весов или смещений любого отдельного перцептрона в сети может иногда приводить к тому, что выход этого перцептрона полностью переворачивается, например, от 0 до 1.

### 1.5.1 Квадратичная целевая функция

Квадратичная целевая функция — это квадратичная функция нескольких переменных, подлежащая оптимизации (минимизации или максимизации) в целях решения некоторой оптимизационной задачи. Данная функция выглядит следующим образом:

$$J = \frac{1}{2} \sum_{i=1}^n (\hat{y}^{(i)} - y^{(i)})^2, \quad (2)$$

где  $n$  - количество примеров на обучающей выборке;

$\hat{y}^{(i)} = \sigma(w^T x^{(i)})$  - предсказанное моделью значение;

$y^{(i)}$  - реальное значение целевой переменной на конкретном примере.

Функцией потерь в данном случае будет выражение  $(\hat{y}^{(i)} - y^{(i)})^2$ . В зарубежной литературе функция потерь пишется как *loss function* или *cost function*. В дальнейшем удобство функции будет заметно при нахождении частных производных. Можно заметить, что функция зависит от двух аргументов: с одной стороны от данных  $(\hat{y}^{(i)})$ , с другой стороны от параметров модели  $(y^{(i)})$ . Данные - это то, что фиксировано, а параметры - то, что будет меняться. Задача заключается в подборке таких параметров, на которых модель будет хорошо работать.

Целевая функция показывает на сколько сильно модель ошибалась. Чем меньше модель ошибается, тем лучше она справляется со своей задачей. Значит необходимо уменьшить эту ошибку. Для уменьшения ошибки хорошо подходит градиент, показывающий направление наибольшего роста функции. А значит отрицание градиента будет показывать нискорейшее убывание функции.

Данное понимание градиента можно применить к обучению линейного нейрона. Один шаг алгоритма будет уменьшать веса модели:

$$w_{j+1} = w_j - \alpha \nabla J \quad (3)$$

В данной формуле  $w_j$  - это текущий вес,  $\alpha$  - коэффициент обучения (*learning rate*), а  $\nabla J$  - градиент целевой функции. Этот шаг продолжается до тех пор, пока веса не перестанут меняться (с какой-то точностью), либо пока количество итераций не достигнет заданного максимума.

### 1.5.2 Обучение линейного нейрона

Чтобы обучать нейрон, надо знать градиент целевой функции ( $\nabla J$ ). В данном разделе будет выведена формула для  $\nabla J$ . Значение целевой переменной будет находиться по формуле  $\hat{y}^{(i)} = w^T x^{(i)}$ . Где  $x^{(i)}$  — это  $i$ -ый пример из множества обучающей выборки.

$$\frac{\partial J}{\partial w_j} = \frac{\partial \left( \frac{1}{2} \sum_{i=1}^n (\hat{y}^{(i)} - y^{(i)})^2 \right)}{\partial w_j} = \sum_{i=1}^n \frac{\partial \left( \frac{1}{2} (\hat{y}^{(i)} - y^{(i)})^2 \right)}{\partial w_j}.$$

Выполним следующий шаг:

$$\sum_{i=1}^n \frac{\partial \left( \frac{1}{2} (\hat{y}^{(i)} - y^{(i)})^2 \right)}{\partial w_j} = \sum_{i=1}^n \frac{\partial \left( \frac{1}{2} (\hat{y}^{(i)} - y^{(i)})^2 \right)}{\partial \hat{y}^{(i)}} \cdot \frac{\partial \hat{y}^{(i)}}{\partial w_j}. \quad (4)$$

Вычисляем частную производную:

$$\frac{\partial \left( \frac{1}{2} (\hat{y}^{(i)} - y^{(i)})^2 \right)}{\partial \hat{y}^{(i)}} = \frac{1}{2} \cdot 2 \cdot (\hat{y}^{(i)} - y^{(i)}) = \hat{y}^{(i)} - y^{(i)}. \quad (5)$$

Так как  $\hat{y}^{(i)} = w^T x^{(i)}$ , то получаем:

$$\frac{\partial \hat{y}^{(i)}}{\partial w_j} = \frac{\partial (w_1 x_1^{(i)} + w_2 x_2^{(i)} + \dots + w_j x_j^{(i)} + \dots + w_m x_m^{(i)})}{\partial w_j} = \frac{\partial w_j x_j^{(i)}}{w_j} = x_j^{(i)}. \quad (6)$$

Объединяем результаты (5) и (6) и получаем:

$$\frac{\partial J}{\partial w_j} = \sum_{i=1}^n (\hat{y}^{(i)} - y^{(i)}) \cdot x_j^{(i)}. \quad (7)$$

Градиент целевой функции будет выглядеть следующим образом:

$$\nabla J = \begin{bmatrix} \dots \\ (\hat{y}^{(i)} - y^{(i)}) \cdot x_j^{(i)} \\ (\hat{y}^{(i)} - y^{(i)}) \cdot x_{j+1}^{(i)} \\ \dots \end{bmatrix} \quad (8)$$

Суммарная ошибка вычисляется по формуле (2). Средняя ошибка имеет следующий вид:

$$J = \frac{1}{2n} \sum_{i=1}^n (\hat{y}^{(i)} - y^{(i)})^2. \quad (9)$$

## 1.6 Сигмоидальный нейрон

Для решения более сложных задач используются нейроны с непрерывными дифференцируемыми активационными функциями, которые не являются линейными. Один из таких нейронов – сигмоидальный нейрон.

Введение функций сигмоидального типа было обусловлено ограниченностью нейронных сетей с пороговой функцией активации нейронов — при такой функции активации любой из выходов сети равен либо нулю, либо единице, что ограничивает использование сетей в задачах, не связанных с классификацией. Использование сигмоидальных функций позволило перейти от бинарных выходов нейрона к непрерывным. Функции передачи такого типа, как правило, присущи нейронам, находящимся во внутренних слоях нейронной сети.

Сигмоидальный нейрон описывается математически следующим образом:

$$f(\mathbf{x}, \mathbf{w}, b) = \sigma(\mathbf{w}\mathbf{x} + b),$$

где  $\sigma(x) = \frac{1}{1 + e^{-x}}$  — логистическая функция (рис 1.3)

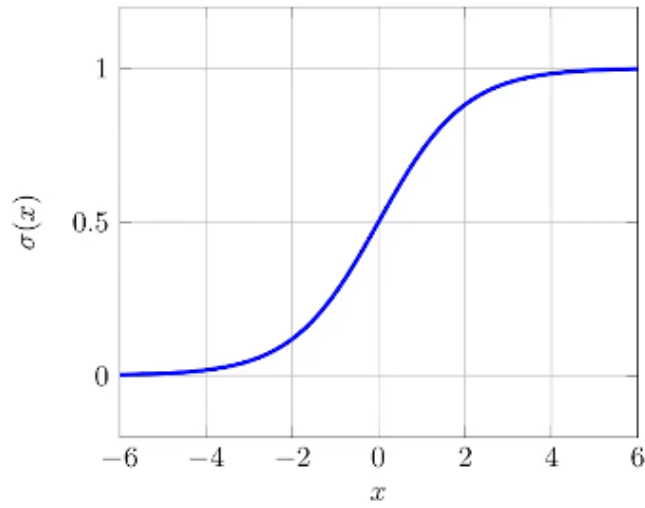


Рис. 1.3: График логистической функции

Особенностью нейронов с логистической функцией является то, что они усиливают сильные сигналы существенно меньше, чем слабые, поскольку области сильных сигналов соответствуют пологим участкам характеристики. Это позволяет предотвратить насыщение от больших сигналов.

Сигмовидные нейроны похожи на перцептроны, но модифицированы так, что небольшие изменения их веса и смещения вызывают только небольшое изменение их выхода. Это решающий факт, который позволяет сети сигмовидных нейронов учиться.

### 1.6.1 Вывод правил обучения для логистического нейрона

Используя данные полученные в главе 1.5.2, выведем правила обучения для сигмоидального нейрона. Целевая функция остается такой же как и для линейного нейрона. Меняется только активационная функция, она будет иметь вид:

$$\hat{y}^{(i)} = \sigma(w^T x^{(i)}). \quad (10)$$

Следовательно меняется и частная производная:

$$\frac{\partial J^{(i)}}{\partial w_j} = \frac{\partial J^{(i)}}{\partial \hat{y}^{(i)}} \cdot \frac{\partial \hat{y}^{(i)}}{\partial S} \cdot \frac{\partial S}{\partial w_j}, \quad (11)$$

где  $S = w^T x^{(i)}$  — сумматорная функция. Находим частные производные:

$$\frac{\partial J^{(i)}}{\partial \hat{y}^{(i)}} = \sigma(w^T x^{(i)}) - y^{(i)}; \quad (12)$$

$$\frac{\partial \hat{y}^{(i)}}{\partial S} = \frac{\partial \sigma(w^T x^{(i)})}{\partial (w^T x^{(i)})} = \sigma(w^T x^{(i)}) \cdot (1 - \sigma(w^T x^{(i)})); \quad (13)$$

$$\frac{\partial S}{\partial w_j} = x_j^{(i)}. \quad (14)$$

Используя результаты (12), (13) и (14), находим градиент функции потерь:

$$\nabla J = \frac{1}{n} \cdot \sum_{i=1}^n (\sigma(w^T x^{(i)}) - y^{(i)}) \cdot \sigma(w^T x^{(i)}) \cdot (1 - \sigma(w^T x^{(i)})) \cdot x_j^{(i)}. \quad (15)$$

## 1.7 Стохастический градиентный спуск

Стохастический градиентный спуск (stochastic gradient descent, SGD) - оптимизационный алгоритм, отличающийся от обычного градиентного спуска тем, что градиент оптимизируемой функции считается на каждом шаге не как сумма градиентов от каждого элемента выборки, а как градиент от нескольких случайно выбранных элементов.

Алгоритм стохастического градиентного спуска:

- Случайным образом выбираем часть примеров из всех имеющихся;
- Считаем  $\hat{y}$  для каждого из них;
- Вычисляем градиент целевой функции по весам для каждого из них;
- Суммируем то, что получилось;
- Обновляем веса;
- Проверяем критерии остановки алгоритма. Если хотя бы один из них отработал - выходим из цикла;

Реализованный алгоритм стохастического градиентного спуска на языке Python:



```

import numpy as np
import math

def SGD(self, X, y, batch_size, learning_rate, eps, max_steps):
    indexes = list(range(len(X)))
    for s in range(max_steps):
        sample_ind = np.random.choice(indexes, batch_size, replace = False)
        X_example = X[sample_ind]
        y_answer = y[sample_ind]

        res = self.update_mini_batch(X_example, y_answer, learning_rate, eps)
        if res:
            return 1
    return 0

def update_mini_batch(self, X, y, learning_rate, eps):
    nabla_j = compute_grad_analytically(self, X, y)
    j_old = J_quadratic(self, X, y)
    delta_w = -1 * learning_rate * nabla_j
    self.w += delta_w
    j_new = J_quadratic(self, X, y)

    if math.fabs(j_old - j_new) < eps:
        return 1
    else:
        return 0

```

Листинг 1.1: Стохастический градиентный спуск

## 1.8 Многослойный перцептрон Румельхарта

Многослойный перцептрон — частный случай перцептрона Розенблатта, в котором один алгоритм обратного распространения ошибки обучает все слои. Название по историческим причинам не отражает особенности данного вида перцептрона, то есть не связано с тем, что в нём имеется несколько слоёв (так как несколько слоёв было и у перцептрона Розенблатта). Особенностью является наличие более чем одного обучаемого слоя (как правило — два или три).

Введем следующие обозначения:

- 1)  $w_{jk}^l$  - вес, который соединяет нейрон с номером  $j$  из слоя с номером  $l$  с нейроном с номером  $k$  из слоя  $l - 1$ ;

- 2)  $z_j^l$  - результат сумматорной функции нейрона  $j$  из слоя  $l$ ;
- 3)  $a_j^l$  - результат активационной функции нейрона  $j$  из слоя  $l$ ;
- 4)  $b_j^l$  - смещение нейрона  $j$  из слоя  $l$ ;
- 5)  $W^l$  - матрица весов входящих в нейроны в с номером слоя  $l$ .

### 1.8.1 Алгоритм обратного распространения ошибки

Определим ошибку в нейроне с номером  $j$  в слое  $l$  следующим образом:

$$\delta_j^l = \frac{\delta J}{\delta z_j^l}. \quad (16)$$

Тогда частная производная целевой функции по сумматорной будет иметь вид:

$$\frac{\delta J}{\delta z_j^l} = (\sigma(z_j) - \sigma^2(z_j)) \cdot (\sigma(z_j) - y_j). \quad (17)$$

Будем обозначать  $\nabla_a J$  — градиент целевой функции по активациям входного слоя, а  $\odot$  — произведение Адамара (поэлементное умножение векторов).

Следовательно из формул (16) и (17) можно получить вектор ошибок выходного слоя:

$$\delta_L = \nabla_a J \odot \sigma'(z^L). \quad (18)$$

Суть алгоритма обратного распространения ошибки - зная ошибку на слое с номером  $l$ , узнать ошибку на  $l-1$  слое. Рассмотрим любой нейрон с номером  $j$  в слое  $l$ . Этот нейрон связан со всеми нейронами из слоя  $l+1$ . Соответственно, если увеличить значение его активационной функции, то это изменение домноженное на соответствующие веса пойдет в нейроны следующего слоя. Допустим значение активационной функции было  $a_j^l$  и стало  $a_j^l + \Delta$ . Можно узнать как изменится значение сумматорной функции каждого из нейронов следующего слоя. До изменения активация нейрона равнялась  $a_j^l w_{jk}^l + C$ , где  $C$  - константа, а после добавления  $\Delta$  получаем, что активация изменилась на

$$\Delta \cdot w_{jk}^l. \quad (19)$$

Известно, что изменение функции равно произведению градиента на разницу вектора аргументов  $\Delta J \approx \nabla J \cdot \Delta x$ . Если вектор изменения входных активаций слоя  $l + 1$  равен  $(\Delta \cdot w_{1k}^{l+1}, \dots, \Delta \cdot w_{n_{l+1}k}^{l+1})$  и  $\Delta$  мало, то тогда получаем:

$$\Delta J = \Delta \cdot (W^{l+1})_k^T \cdot \delta^{l+1},$$

где  $(W^{l+1})_k^T$  —  $k$ -й столбец матрицы весов между слоем  $l$  и  $l + 1$ .

Выразим ошибку  $\delta^l$  через ошибку  $\delta^{l+1}$ . Из формулы (18) известно, что

$$\delta_l = \nabla_{a^l} J \odot \sigma'(z^l). \quad (20)$$

Осталось выразить  $\nabla_{a^l} J$  через  $\delta^{l+1}$ .  $\nabla_{a^l} J$  состоит из частной производной  $\frac{\partial J}{\partial a_j^l}$ .

$$\Delta J = \Delta \cdot (W^{l+1})_k^T \cdot \delta^{l+1}, \quad (21)$$

где  $(W^{l+1})_k^T$  — это  $k$ -й столбец матрицы  $W^{l+1}$ . Следовательно,  $\frac{\partial J}{\partial a_i^l} = (W^{l+1})_k^T \cdot \delta^{l+1}$ . Из формулы (21) получаем, что  $\nabla_{a^l} J = (W^{l+1})_k^T \cdot \delta^{l+1}$ . Итого:

$$\delta^l = ((W^{l+1})^T \cdot \delta^{l+1}) \odot \sigma'(z^l). \quad (22)$$

Сумматорную функцию нейрона  $j$  в слое  $l$  можно выразить через активационную функцию нейронов из слоя  $l - 1$ .  $z_j^l = W_j^l \cdot a_{l-1} + b_j^l$

Тогда частная производная целевой функции по весу  $\frac{\partial J}{\partial W_{jk}^l}$  будет равна  $a_k^{l-1} \cdot \delta_j^l$ , а по смещению:  $\frac{\partial J}{\partial w_{jk}^l} = \delta_j^l$ .

Все нужные формулы получены. Теперь можно выделить 4 основных, которые будут участвовать в алгоритме обратного распространения ошибки:

- 1)  $\delta_L = \nabla_a J \odot \sigma'(z^L)$
- 2)  $\delta^l = ((W^{l+1})^T \cdot \delta^{l+1}) \odot \sigma'(z^l)$
- 3)  $\frac{\partial J}{\partial w_{jk}^l} = \delta_j^l$
- 4)  $\frac{\partial J}{\partial W_{jk}^l} = a_k^{l-1} \cdot \delta_j^l$

Первая формула из списка помогает получить вектор ошибок выходного слоя, зная градиент целевой функции по активациям нейронов выходного слоя  $\nabla_a J$ . По второй формуле можно вычислить ошибки в слое  $l - 1$ . Это делает возможным осуществить шаг алгоритма с дальнего слоя к ближнему.

Третья и четвертая формула позволяют вычислить нужные производные (по смещению и по любому из весов).

## Шаги алгоритма обратного распространения ошибки

Используя формулы из предыдущего пункта, получаем алгоритм обратного распространения ошибки.

- Выбрать  $m$  примеров, на основании которых будет производиться изменение весов;
- осуществить прямое распространение активации для каждого примера;
- использовать первую формулу, чтобы посчитать ошибки нейронов в выходном слое по каждому примеру;
- использовать вторую формулу, для нахождения ошибок нейронов во всех слоях по каждому примеру;
- использовать третью и четвертую формулу, чтобы получить градиенты по параметрам и смещениям;
- обновить параметры;
- проверить критерии остановки алгоритма.

## Нормализация

Для правильной работы алгоритма обратного распространения ошибки важно, чтобы веса были разные. Действительно, если все веса одинаковые, то все нейроны из первого скрытого слоя получают одинаковую активацию. Потому что каждый нейрон связан с каждым входом скрытого слоя. Соответственно, никакой разницы между нейронами первого скрытого слоя не существует. Эту проблему можно решить, инициализируя веса случайными значениями. Например, нормальный шум или равномерный шум.

Большие векторы весов, также как и большие векторы входных активаций могут привести к тому, что производная будет близка к нулю. Соответственно, можно получить состояние *ступора сети*. Если активации имеют разный масштаб, то и градиенты будут иметь разный масштаб. Поэтому очень важно правильно нормализовать входы.

## Глава 2

# Эволюционные алгоритмы

### 2.1 Виды эволюционных алгоритмов

**Эволюционные алгоритмы** — это направление в компьютерных науках, использующее принципы биологической эволюции для решения задач искусственного интеллекта. Основным принцип биологической эволюции — это сочетание естественного отбора, мутаций и воспроизводства. Хотя эволюционные алгоритмы и пытаются имитировать биологическую эволюцию, они более схожи с искусственным разведением животных, то есть скрещиванием самых лучших представителей, отбором их лучших потомков и повторным скрещиванием уже этих потомков.

#### Виды алгоритмов

- Генетические алгоритмы — эвристический алгоритм поиска, используемый для решения задач оптимизации и моделирования путём случайного подбора, комбинирования и вариации искомых параметров;
- генетическое программирование — автоматическое создание или изменение программ с помощью генетических алгоритмов;
- эволюционное программирование — аналогично генетическому программированию, но структура программы постоянна, изменяются только числовые значения;
- эволюционные стратегии — похожи на генетические алгоритмы, но в следующее поколение передаются только положительные мутации;

- нейроеволюция — аналогично генетическому программированию, но геномы представляют собой искусственные нейронные сети, в которых происходит эволюция весов при заданной топологии сети, или помимо эволюции весов также производится эволюция топологии.

## 2.2 Понятие естественного отбора

Процесс естественного отбора начинается с выбора наиболее приспособленных особей из популяции. Они производят потомство, которое наследует характеристики родителей и будет добавлено к следующему поколению. Если родители подготовлены лучше, то их дети имеют больше шансов на выживание. Этот процесс продолжает повторяться, и в конце, поколение с самыми подходящими особями будет найдено.

Это понятие может быть применено к проблеме обучения нейронной сети. В работе рассматривается набор решений для проблемы и выбирается из них набор лучших.

### Пять фаз эволюционного алгоритма

- Начальная популяция;
- фитнес-функция;
- выбор;
- скрещивание;
- мутация.

### Начальная популяция

Процесс начинается с набора нейронных сетей. Каждая нейронная сеть (индивид) умеет решать поставленную задачу с каким-то уровнем успеха. Индивид характеризуется набором параметров (переменных), известных как гены. Гены объединяются в цепочку, образуя хромосому.

В эволюционном алгоритме набор генов индивида представлен строкой в алфавитном порядке. Обычно используются двоичные значения (строка из 1 и 0). Часто говорят, что таким образом кодируются гены в хромосоме.

## Фитнес-функция

Функция пригодности (фитнес-функция) определяет, насколько подходит индивид (способность нейросети конкурировать с другими нейронными сетями) для решения конкретной задачи. Это дает оценку пригодности для каждой сети. Вероятность того, что конкретная нейронная сеть будет выбрана для скрещивания, основана на ее оценке пригодности.

Идея фазы отбора состоит в том, чтобы отобрать наиболее подходящих особей и позволить им передать свои гены следующему поколению. Две пары индивидов (родителей) выбираются на основе их показателей пригодности. Нейронные сети с высоким процентом ответов имеют больше шансов быть отобранными для размножения.

## Скрещивание

Скрещивание является наиболее значимой фазой в генетическом алгоритме. Для каждой пары родителей, которые должны быть спарены, точка пересечения (crossover point) выбирается случайным образом из генов. Например, рассмотрим точку пересечения 3, как показано ниже (рис 2.1).



Рис. 2.1: Точка пересечения (crossover point)

Потомки создаются путем обмена генами родителей между собой, пока не будет достигнута точка пересечения.

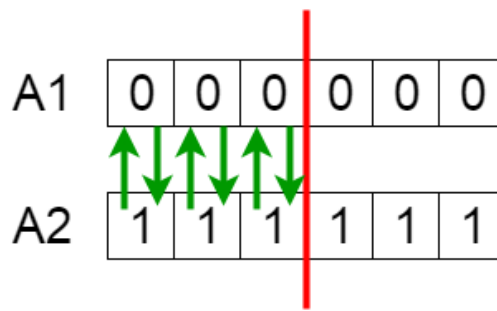


Рис. 2.2: Обмен генами

В итоге получается новое потомство.

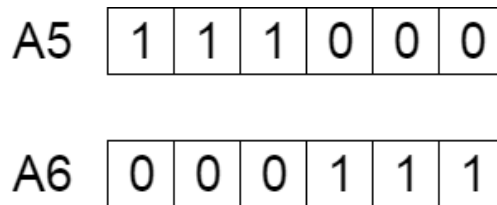
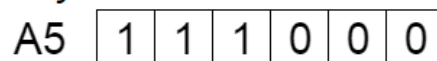


Рис. 2.3: Потомки

## Мутация

У некоторых новых потомков отдельные гены могут подвергаться мутации с низкой случайной вероятностью. Это подразумевает, что некоторые биты в битовой строке могут быть перевернуты.

До мутации



После мутации

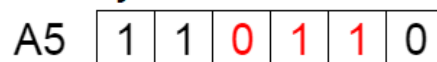


Рис. 2.4: Мутация

Мутация происходит для поддержания разнообразия в популяции и предотвращения преждевременной конвергенции. Алгоритм завершается, если популяция не производит потомство, которое значительно отличается от предыдущего поколения.



## 2.3 Применение эволюционных алгоритмов в нейронных сетях

Разные исследователи высказывали мысль о том, что нейронные сети могут обучаться с помощью генетического алгоритма. В первых работах на эту тему генетический алгоритм рассматривали в качестве метода обучения небольших однонаправленных нейронных сетей, но в последующем было реализовано применение этого алгоритма для сетей с большей размерностью.

Часто задача заключается в оптимизации весов нейронной сети, имеющей априори заданную топологию. Веса кодируются в виде двоичных последовательностей (хромосом). Каждая особь популяции характеризуется полным множеством весов нейронной сети. Оценка приспособленности особей определяется фитнес функцией, задаваемой в виде суммы квадратов погрешностей.

Существует как минимум два важнейших аргумента в пользу применения генетических алгоритмов для оптимизации весов нейронной сети. Прежде всего, генетические алгоритмы обеспечивают глобальный просмотр пространства весов и позволяют избегать локальные минимумы. Также они могут использоваться в задачах, для которых очень сложно получить информацию о градиентах либо она оказывается слишком дорогостоящей.

### Два этапа обучения нейронных сетей

Эволюционный подход к обучению нейронных сетей состоит из двух основных этапов. Первый этап - это выбор соответствующей схемы представления весов связей. Он заключается в принятии решения - можно ли кодировать эти веса двоичными последовательностями или требуется какая-то другая форма. На втором этапе уже осуществляется сам процесс эволюции, основанный на генетическом алгоритме. В данной работе был выбран метод кодирования матрицы весами нейронной сети.

После выбора схемы хромосомного представления генетический алгоритм применяется к популяции особей с реализацией типового цикла эволюции, состоящего из четырех шагов.

- 1) Декодирование хромосомы текущего поколения для восстановления множества весов и конструирование соответствующей этому множеству нейронной сети с априорно заданной архитектурой и правилом обучения.

- 2) Расчет общей среднеквадратичной погрешности между фактическими и заданными значениями на всех выходах сети при подаче на ее входы обучающих образов. Погрешность определяет приспособленность сети. Функция приспособленности может быть выбрана в зависимости от задачи.
- 3) Репродукция особей согласно их фитнес функции. Чем выше приспособленность, тем больше шансов имеет особь на скрещивание.
- 4) Применение генетических операторов - таких как скрещивание, мутация и/или инверсия для получения нового поколения.

Достоинством эволюционного подхода считается тот факт, что функцию приспособленности можно легко определить специально для эволюции сети со строго определенными свойствами. Например, если для оценивания приспособленности использовать результаты тестирования вместо результатов обучения, то будет получена сеть с лучшей способностью к обобщению.

## Глава 3

# Разработанное ПО

### 3.1 Требования к программному обеспечению

Разработать программное обеспечение для исследования работы эволюционных алгоритмов в нейронных сетях. Провести тестирование его эффективности.

### 3.2 Выбор средств для создания ПО

Для написания программы использовался язык программирования Python в связке со средой разработки Jupyter Notebook. Этот язык был выбран для написания выпускной работы бакалавра в связи с большим количеством библиотек, в которых реализованы основные методы для работы с матрицами, строками и файлами. Среда разработки Jupyter Notebook позволяет удобно работать с кодом, подсвечивая синтаксические конструкции языка. Главной интересной функцией среды разработки является возможность компилировать код построчно, что позволяет увидеть ошибки сразу.

### 3.3 Описание созданного ПО

Нужно научиться решать такую задачу: Зная начальное направление (угол от поверхности земли) и скорость мяча, определить, где этот мяч упадёт. Или иначе: попадёт ли мяч в мишень, находящуюся на земле, при заданных начальных условиях.

Данные представляют набор четырех параметров:

- 1) скорость удара.
- 2) Угол удара.
- 3) Расстояние до мишени.
- 4) Результат(недолетел, попал, перелетел).

Скорость броска принимает значения от 0 и до 50. Угол принимает значения из промежутка от 0 до  $\frac{\pi}{2}$ .

Файл с данным загружается следующей командой:

```
data = np.loadtxt("data.csv", delimiter=",")
```

Разница в масштабах настолько существенна, что имеет смысл нормализовать наши данные, чтобы при обучении нейронная сеть не отвлекалась на попытки скомпенсировать масштаб данных масштабом весов.

```
means = data.mean(axis=0)
2 means[-1] = 0
stds = data.std(axis=0)
4 stds[-1] = 1
data = (data - means) / stds
```

Для проверки точности работы программы множество всех примеров было разделено на обучающую выборку и тестовую выборку.

```
np.random.seed(42)
2 test_index = np.random.choice([True, False], len(data), replace=True, p=[0.25,
    0.75])
test = data[test_index]
4 train = data[np.logical_not(test_index)]
```

Данные приводятся в тот вид, в котором они понимаются нейросетью. Для обучения нейронной сети нужно, чтобы ответ был в формате one-hot: вектор длины 3 (общее количество классов), состоящий из нулей и одной единицы на месте правильного класса наблюдения. Это сделано с помощью `np.eye`: для единицы, стоящей на `i`-м месте, был сделан вектор `np.eye(3, 1, k=-i)`.

```
train = [(d[:3][:, np.newaxis], np.eye(3, 1, k=-int(d[-1]))) for d in train]
2 test = [(d[:3][:, np.newaxis], d[-1]) for d in test]
```

---

Параметры нейронной сети:

- `inputCount` — количество нейронов входного слоя;
- `hiddenCount` — количество нейронов скрытого слоя;
- `outputCount` — количество нейронов выходного слоя;
- `SIZE` — размер популяции нейронных сетей.

```
input_count = 3
hidden_count = 7
output_count = 3
SIZE = 10
```

Следующим этапом идет создание популяции из 10 особей нейронных сетей, а затем применение эволюционного алгоритма к этой популяции.

Эволюционный алгоритм:

- Создается популяция из 10 нейронных сетей со случайными весами;
- каждая особь из популяции обучается на тренировочной выборке;
- выбирается 5 лучших особей по проценту правильных ответов;
- происходит скрещивание двух самых лучших особей и 4 пары из 5 лучших (получается 10 новых нейронных сетей);
- случайным образом выбирается особь из нового поколения, и мутируется один из ее генов;
- повторяются пункты 2-5 по количеству эпох, в данной работе будет 100 эпох (поколений нейросетей).

```
nn = []
for i in range(SIZE):
    network = Network([input_count, hidden_count, output_count]) percent =
    network.SGD(training_data=train, epochs=1, mini_batch_size=5,
```

```

6  eta=1, test_data=test)
   nn.append((network, percent))
8
for epoch in range(100):
10  nn.sort(key=lambda x: x[1])
   nn = nn[5:]
12  nnChild = []
   indexes = np.random.randint(0, 5, 10)
14  np.random.shuffle(indexes)

16  nn[indexes[4]][0].genetic(nn[indexes[3]][0], nnChild)

18  for x in range(4):
       nn[indexes[x]][0].genetic(nn[indexes[x + 5]][0], nnChild)
20  nn = nnChild

```

Каждая нейронная сеть имеет 3 слоя — входной, скрытый и выходной. Это означает, что веса нейронной сети будут храниться в 2-х массивах:

- первый массив — веса между входным и скрытым слоем, размерности  $7 \times 3$ ,
- второй массив — веса между скрытым и выходным слоем, размерности  $3 \times 7$ .

Алгоритм скрещивания работает следующим образом:

- На вход подается 2 особи (нейронных сети);
- выбирается случайным образом индекс из матрицы весов;
- создается 2 новых потомка, таких, что в первом потомке будут веса из первой нейронной сети до номера случайного индекса, остальная часть из матрицы весов второй нейронной сети, а у второго потомка наоборот;
- обучается алгоритмом обратного распространения ошибки и проверяется результат на тестовой выборке;
- возвращается 2 новых нейронных сети с их результатами.

После обучения выбирается нейронная сеть с самым лучшим результатом.

### 3.4 Результаты тестирования ПО

В ходе работы программы были выведены результаты обучения нейронной сети с использованием эволюционного алгоритма. Нейронная сеть обучалась на тренировочной выборке 100 раз. Ниже представлены результаты начальной эпохи, 51-й эпохи и последней эпохи.

После первого обучения нейронные сети дают в среднем 40% правильных ответов (рис 3.1).

Эпоха	0
Процент верных ответов:	39 %
Процент верных ответов:	42 %
Процент верных ответов:	41 %
Процент верных ответов:	37 %
Процент верных ответов:	35 %
Процент верных ответов:	38 %
Процент верных ответов:	40 %
Процент верных ответов:	39 %
Процент верных ответов:	42 %
Процент верных ответов:	38 %

Рис. 3.1: Начальное поколение нейронных сетей

Поколение с номером 51 дает уже 73% правильных ответов (рис 3.2). Заметен существенный прирост в количестве верных ответов.

Эпоха	51
Процент верных ответов:	72 %
Процент верных ответов:	71 %
Процент верных ответов:	77 %
Процент верных ответов:	76 %
Процент верных ответов:	74 %
Процент верных ответов:	72 %
Процент верных ответов:	73 %
Процент верных ответов:	76 %
Процент верных ответов:	76 %
Процент верных ответов:	74 %

Рис. 3.2: 51 поколение нейронных сетей

Последнее поколение будет предугадывать правильный ответ в среднем в 93% случаев (рис 3.3).

Эпоха	99
Процент верных ответов:	91 %
Процент верных ответов:	95 %
Процент верных ответов:	92 %
Процент верных ответов:	89 %
Процент верных ответов:	94 %
Процент верных ответов:	94 %
Процент верных ответов:	96 %
Процент верных ответов:	93 %
Процент верных ответов:	94 %
Процент верных ответов:	92 %

Рис. 3.3: 99 поколение нейронных сетей

После прохождения 100 эпох получилась рабочая нейронная сеть, способная отвечать правильно на поставленную задачу в 96% случаев. Это и будет итоговая нейронная сеть.



# Заключение

В настоящее время приобретает популярность такая технология, как нейронные сети. Это происходит благодаря их обучаемости и решению проблем, с которыми не справляются обычные алгоритмы.

В данной выпускной квалификационной работе были рассмотрены такие виды нейронов, как перцептрон, сигмоидальный нейрон и многослойный перцептрон Румельхарта. Были выведены основные формулы для их обучения. Также было рассмотрено применение эволюционных алгоритмов в нейронных сетях. В ходе исследования был разработан алгоритм скрещивания и эволюционный алгоритм.

В данной работе была разработана программа, решающая задачу классификации. Программа определяет, попадет ли футболист в мишень только по исходным данным. Были проведены опыты, которые показали наилучшую архитектуру для сети. Проведенные тесты показали хорошую обучаемость нейронной сети с применением эволюционного алгоритма.

# Литература

- [1] Барский А. Нейронные сети: распознавание, управление, принятие решений. — М.: Финансы и статистика, 2004. — 176 с: ил. — (Прикладные информационные технологии). ISBN 5-279-02757-X.
- [2] Вандер Дж. Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 576 с.: ил. — (Серия «Бестселлеры O'Reilly»). ISBN 978-5-496-03068-7.
- [3] Гладков Л., Курейчик В. Генетические алгоритмы — М.: Издательская фирма «Физико-математическая литература», 2010. — 367 с: ил. ISBN: 978-5-9221-0510-1.
- [4] Голованов В. Нейросеть в 11 строчек на Python. — 2015. — Электронные данные. — Режим доступа: <https://habrahabr.ru/post/271563/>.
- [5] Голованов В. Нейросеть на Python, часть 2: градиентный спуск. — 2015. — Электронные данные. — Режим доступа: <https://habrahabr.ru/post/272679/>.
- [6] Дорогов А. Теория и проектирование быстрых перестраиваемых преобразований и слабосвязанных нейронных сетей. — Санкт-Петербург: Политехника, 2014. — 328 с ISBN 978-5-7325-1036-2.
- [7] Дрейпер Н., Смит Г. Прикладной регрессионный анализ / Пер. с англ. — М.: Издательский дом «Вильямс», 2007. — 912 с.: ил. ISBN: 978-5-8459-0963-3.
- [8] Кочуков А. Моделирование нейросети Машина Больцмана. — 2013. — Электронные данные. — Режим доступа: <https://geektimes.ru/post/188368/>.

- [9] Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы: Пер. с польск. И. Д. Рудинского. — 2-е изд., стереотип. — М.: Горячая линия — Телеком, 2013. — 384 с.: ил. ISBN 978-5-9912-0320-3.
- [10] Терехов С. Лекции по теории и приложениям искусственных нейронных сетей. — 1999. — Электронные данные. — Режим доступа: <http://alife.narod.ru/lectures/neural/Neuch08.html>.
- [11] Хайкин С. Нейронные сети. Полный курс. 2-е изд. Пер. с англ. — М.: Издательский дом "Вильямс 2006. — 1104 с.: ил. ISBN: 978-5-8459-0890-2.
- [12] Хуршудов А. Байесовская нейронная сеть. — 2016. — Электронные данные. — Режим доступа: [https:// habrahabr.ru/post/276355/](https://habrahabr.ru/post/276355/).
- [13] Шолле Ф. Глубокое обучение на Python СПб.: Питер, 2018. — 400 с.: ил. (Серия «Библиотека программиста»). ISBN 978-5-4461-0770-4.
- [14] Nielsen M. Neural Networks and Deep Learning, 2017. — Электронные данные. — Режим доступа: <http://neuralnetworksanddeeplearning.com/>.

# Приложение

```
import math
2 import doctest
import random
4 import matplotlib.pyplot as plt
import time
6 import numpy as np

8 def sigma(x):
    return 1 / (1 + np.exp(-x))
10
11 def sigma_prime(x):
12     return sigma(x) * (1 - sigma(x))

14 def J_quadro(neuron, X, y):
    return 0.5 * np.mean((neuron.vector_forward_pass(X) - y) ** 2)
16
17 def J_quadro_derivative(y, y_hat):
18     return (y_hat - y) / len(y)

20 def compute_gradient(neuron, X, y, J_prime=J_quadro_derivative):
    # Compute activation vector
22
    z = neuron.summatory_function(X)
24     y_hat = neuron.activation(z)

26     dy_dyhat = J_prime(y, y_hat)
    dyhat_dz = neuron.activation_derivative(z)
28
    dz_dw = X
30     grad = ((dy_dyhat * dyhat_dz).T).dot(dz_dw)
    grad = grad.T
32     return grad

34 class Network:

36     def __init__(self, sizes, output=True):
        self.num_layers = len(sizes)
38         self.sizes = sizes
```

```

self.biases = [np.random.randn(y,1) for y in sizes[1:]]
self.weights = [np.random.randn(y, x) for x, y in zip(sizes[:-1], sizes
[1:])]
self.output = output

def feedforward(self, a):
    for b, w in zip(self.biases, self.weights):
        a = sigma(np.dot(w, a) + b)
    return a

def SGD(self, training_set, epochs, mini_batch_size, eta,
        test_set=None):

    if test_set is not None: n_test = len(test_set)
    n = len(training_set)
    success_tests = 0

    for j in range(epochs):
        random.shuffle(training_set)
        mini_batches = [
            training_set[k:k+mini_batch_size]
            for k in range(0, n, mini_batch_size)]

        for mini_batch in mini_batches:
            self.update_mini_batch(mini_batch, eta)

        if test_set is not None and self.output:
            success_tests = self.evaluate(test_set)
            print("Percent of correct answers: ", int(success_tests /
n_test * 100), "%")

        if test_set is not None:
            return success_tests / n_test

def update_mini_batch(self, mini_batch, eta):
    #Update biases and weights neural network, make one step of SGD with
mini batch

    nabla_biases = [np.zeros(b.shape) for b in self.biases]
    nabla_weights = [np.zeros(w.shape) for w in self.weights]
    for x, y in mini_batch:
        delta_nabla_biases, delta_nabla_weights = self.back_propagation(x, y
)

    nabla_biases = [nb+dnb for nb, dnb in zip(nabla_biases,
delta_nabla_biases)]
    nabla_weights = [nw+dnw for nw, dnw in zip(nabla_weights,
delta_nabla_weights)]

```

```

82         epsilon = eta / len(mini_batch)
        self.weights = [w - epsilon * nw for w, nw in zip(self.weights,
nabla_weights)]
        self.biases = [b - epsilon * nb for b, nb in zip(self.biases,
nabla_biases)]

84
def back_propagation(self, x, y):
86     nabla_biases = [np.zeros(b.shape) for b in self.biases]
    nabla_weights = [np.zeros(w.shape) for w in self.weights]

88
    activations = [x]
90    activation = x
    zs = []
92    for b, w in zip(self.biases, self.weights):
        # Compute activations
94        z = w.dot(activation) + b
        zs.append(z)
96        activation = sigma(z)
        activations.append(activation)
98    pass

100    # Back propagation
    delta = (activations[-1] - y) * activations[-1] * (1 - activations[-1])
102    nabla_biases[-1] = delta
    nabla_weights[-1] = delta.dot(activations[-2].T)

104
    for l in range(2, self.num_layers):
106        delta = (self.weights[-l+1].T.dot(delta) * activations[-1] * (1 -
activations[-1])) # error on layer L-l
        nabla_biases[-l] = delta
            # derivative J by L-l layer
108        nabla_weights[-l] = delta.dot(activations[-l-1].T)
    return nabla_biases, nabla_weights

110
def evaluate(self, test_set):
112    #Return number of tests with correct answers
    test_results = [(np.argmax(self.feedforward(x)), y)
114                    for (x, y) in test_set]
    return sum(int(x == y) for (x, y) in test_results)

116
def evolution_algorithm(self, network, neural_network_child):
118    input_count = 3
    hidden_count = 7
120    output_count = 3

122
    first = Network([input_count, hidden_count, output_count])
    second = Network([input_count, hidden_count, output_count])
124

```

```

126     a = self.weights[0]
127     b = self.weights[1]
128     c = network.weights[0]
129     d = network.weights[1]

130     a = a.reshape(1, 21)
131     b = b.reshape(1, 21)
132     c = c.reshape(1, 21)
133     d = d.reshape(1, 21)

134

135     first_arr = np.empty((0,1), int)
136     second_arr = np.empty((0,1), int)

137
138     for i in range(len(a)):
139         first_arr = np.append(first_arr, a[i])
140         second_arr = np.append(second_arr, c[i])
141     for i in range(len(b)):
142         first_arr = np.append(first_arr, b[i])
143         second_arr = np.append(second_arr, d[i])
144
145
146     point = random.randint(0, len(first_arr))

147
148     for i in range(point):
149         first_arr[i], second_arr[i] = second_arr[i], first_arr[i]

150
151     a = first_arr[:a.shape[1]].reshape(7,3)
152     b = first_arr[b.shape[1]:].reshape(3,7)
153     c = second_arr[:c.shape[1]].reshape(7,3)
154     d = second_arr[d.shape[1]:].reshape(3,7)

155
156     first.weights[0] = a
157     first.weights[1] = b
158     second.weights[0] = c
159     second.weights[1] = d

160
161     random_point = random.randint(0, 1)
162     random_child = random.randint(0, 1)

163
164     if random_child == 0:
165         first.weights[random_point] = -first.weights[random_point]
166     else:
167         second.weights[random_point] = -second.weights[random_point]

168
169     first_result = first.SGD(training_set=train, epochs=1, mini_batch_size
=5, eta=1, test_set=test)
170     second_result = second.SGD(training_set=train, epochs=1, mini_batch_size
=5, eta=1, test_set=test)

```

```

172         neural_network_child.append((first , first_result))
        neural_network_child.append((second , second_result))

174 data = np.loadtxt("data.csv", delimiter=",")

176 means = data.mean(axis=0)
    means[-1] = 0
178 stds = data.std(axis=0)
    stds[-1] = 1
180 data = (data - means) / stds

182 np.random.seed(42)
    test_index = np.random.choice([True, False], len(data), replace=True, p=[0.25,
        0.75])
184 test = data[test_index]
    train = data[np.logical_not(test_index)]

186
    train = [(d[:3][:, np.newaxis], np.eye(3, 1, k=int(d[-1]))) for d in train]
188 test = [(d[:3][:, np.newaxis], d[-1]) for d in test]

190 input_count = 3
    hidden_count = 7
192 output_count = 3
    SIZE = 10
194 epoch = 1

196 random.seed(1)
    np.random.seed(1)
198 neural_network = [] #array of neural network
    for i in range(SIZE):
200         #create neural network, learn and save results
        network = Network([input_count, hidden_count, output_count])
202         percent = network.SGD(training_set=train, epochs=1, mini_batch_size=5, eta
            =1, test_set=test)

204         #add pair of neural network in array
        neural_network.append((network, percent))

206
    # start evolution algorithm
208 for epoch in range(epoch):
        print("\nEpoch ", epoch, )

210
        neural_network.sort(key=lambda x: x[1]) # sorting by number of correct
        answers
212 neural_network = neural_network[5:] # pick top 5
        neural_network_child = [] # array of childrens
214 indexes = np.random.randint(0, 5, 10) # array of random indexes
        np.random.shuffle(indexes)

```



```
216     #pick top 2 neural network
218     neural_network[indexes[4]][0].evolution_algorithm(neural_network[indexes
219     [3]][0], neural_network_child)
220
221     for x in range(4):
222         neural_network[indexes[x]][0].evolution_algorithm(neural_network[indexes
223         [x + 5]][0], neural_network_child)
224
225     neural_network = neural_network_child
```