

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
КАФЕДРА ТЕХНОЛОГИИ ПРОГРАММИРОВАНИЯ

Коротков Ян Глебович

Выпускная квалификационная работа бакалавра

**Исследование подходов к обучению
нейронных сетей**

Направление 010400

Прикладная математика, фундаментальная информатика и основы
программирования

Научный руководитель,
кандидат физ.-мат. наук,
доцент
Должиков В.В.

Санкт-Петербург
2016

Введение.....	3
Постановка задачи.....	5
Обзор используемой литературы	6
Глава 1. Нейронная сеть.	7
1.1. Основные сведения о нейронных сетях.....	7
1.2. Однослойный перцептрон.....	10
1.3. Многослойный перцептрон.....	11
1.4. Классификация нейронных сетей.....	13
Глава 2. Обучение нейронных сетей.	18
2.1. Общее описание подходов к обучению нейронных сетей.....	18
2.2. Алгоритмы обучения сетей.....	20
2.3. Концептуальные подходы к обучению сетей.....	22
2.4. Другие методики обучения	24
2.5. Методы увеличения скорости процесса обучения сети.....	27
2.6. Алгоритм обратного распространения ошибки	29
2.7. Обобщение информации о нейронных сетях.	31
Заключение.	33
Список используемой литературы	34

Введение

Искусственные нейронные сети, также как и биологические, представляют собой системы, состоящие из огромного количества функционирующих процессоров-нейронов, каждый из которых выполняет какой-либо небольшой объём работ, возложенный на него, при этом обладая большим числом связей с остальными, что и характеризует силу вычислений сети. Даже учитывая, что при построении искусственных сетей производится ряд упрощений и допущений, отличающий их от биологических, они всё же обладают теми же характерными чертами, что и мозг живого существа: обучение, основанное на опыте, поиск общих черт, нахождение важной информации в общем массиве поступающих данных [1].

Нейронные сети способны изменять ход своих действий под влиянием внешних факторов. В процессе обработки поступающих данных они могут подстраиваться и обучаться, дабы реакция на происходящие события была верной. Также сеть способна распознавать данные, имеющие некоторые отклонения, что даёт ей возможность выявить истинный образ, вне зависимости от сопровождающих его искажений или помех.

Нейронные сети могут выполнять задачи прогнозирования, оценки и классификации на основе имеющихся данных, что позволило применить их в совершенно разных областях человеческой жизни: физика, экономика, медицина, техника и даже творчество. Такой большой прорыв был возможен в связи со следующими особенностями:

- **Невероятные возможности.** Нейронные сети являются функционально богатым методом моделирования, способным воссоздавать исключительно сложные примеры взаимодействия.
- **Легкость в применении.** Обработывая поступающую информацию, подобранную пользователем, сети обучаются, находят отличительные черты и сходства во фрагментах получаемой информации. При этом поступающие для обучения

данные должны быть подготовлены пользователем, имеющим представление о принципах отбора подаваемой информации, способным выбрать подходящую архитектуру нейронной сети, а также умеющим интерпретировать получаемые от сети результаты. Тем не менее, для применения нейронных сетей не обязательно быть глубоко вовлеченным в эту область знаний, в отличие, например, от классических методов статистики.

- Надёжность. Сеть способна выдавать правильные результаты (правда, возможно, с меньшей точностью) даже в случае, если некоторые нейроны выйдут из строя или будут работать некорректно.

Нейронные сети привлекательны на интуитивном уровне тем, что они основаны на упрощенной модели биологических нервных систем. Развитие подобного рода сетей может в перспективе привести к созданию искусственного интеллекта, действительно способного мыслить.

Применение сетей возможно в совершенно различных областях: в прогнозировании курса валют, цифровой обработке изображений, распознавании и синтезе текста и речи, семантическом поиске и анализе текстов, системах поддержки принятия решений, системах безопасности и многих других.

В данной работе будут рассматриваться математические модели искусственных нейронных сетей, принципы классификации сетей, концептуальные подходы к их обучению, а также различные методы обучения каждого из подходов. Также будет исследован вопрос увеличения скорости обучения сети и рассмотрены более подробно области применения нейронных сетей.

Постановка задачи

В ВКР ставилась задача рассмотрения различных математических моделей нейронных сетей и парадигм их обучения, обзора методов, применяемых для обучения, исследование информации об областях их применения, а также изучения соответствующей научной литературы, описания алгоритмов обучения сети. Задача свелась к исследованию теоретических материалов об искусственных нейронных сетях, описанию некоторых математических моделей нейронных сетей, проведению соответствующего анализа полученных данных и подведению итогов относительно полученной информации.

Обзор используемой литературы

При написании данной научной работы использовались научные источники, учебная литература, интернет-публикации, а также статьи из научных изданий.

Основные сведения о нейронных сетях были получены из текстов [1],[2],[7],[9] в которых также давалась подробная информация и определения из теории нейронных сетей.

Сведения о подходах к обучению нейронных сетей взяты из ресурсов [6],[10],[14],[16]. Помимо этой информации из них также получены сведения о преимуществах и недостатках алгоритмов и принципов обучения сетей.

Данные о применении сетей в жизни основаны на информации из [3],[6],[8],[12],[15].

Стоит отметить, что существует множество электронных ресурсов, в которых описываются подробно алгоритмы обучения нейронных сетей. Особенно следует выделить [2],[8],[13].

Необычным с точки зрения стиля повествования, но при этом весьма полезным в вопросе предоставляемой информации оказался источник [16].

Также о нейронных сетях будущего выдвигаются прогнозы в статье [15].

Глава 1. Нейронная сеть.

В данной главе будет даваться определение и общие сведения о нейронной сети в целом, математическая модель нейрона и описание моделей однослойного и многослойного перцептронов. Также будет рассмотрен вопрос классификации сетей в зависимости от различных характеристик.

1.1. Основные сведения о нейронных сетях

Искусственные нейронные сети появились в процессе исследования возможности создания искусственного интеллекта. Если говорить конкретнее, из попыток воспроизвести умение биологической нейронной сети обучаться и находить ошибки, основываясь на предыдущих данных, моделируя, таким образом, на низком уровне мозг живого существа. Мозг состоит из огромного числа нейронов (приблизительно 10 миллиардов), на каждый из которых приходится несколько тысяч связей. Нейроны – специальные клетки, умеющие распознавать электрохимические сигналы. У каждого нейрона имеются отростки двух типов, выполняющие разные функции: дендриты, отвечающие за получение сигналов и один аксон, выполняющий функцию передачи импульса другим нейронам (рис. 1). Места соединения аксона одной нервной клетки с дендритами другой называются синапсами. Сила нервного импульса, проходящего через синапс, зависит от количества нейромедиатора – биологически активного химического вещества, вырабатываемого в синапсе. Также при прохождении через синапс меняется и сила импульса в некоторое количество раз. Это число в теории искусственных нейронных сетей называют весом синапса. Поступившие по нескольким дендритам одновременно импульсы суммируются.

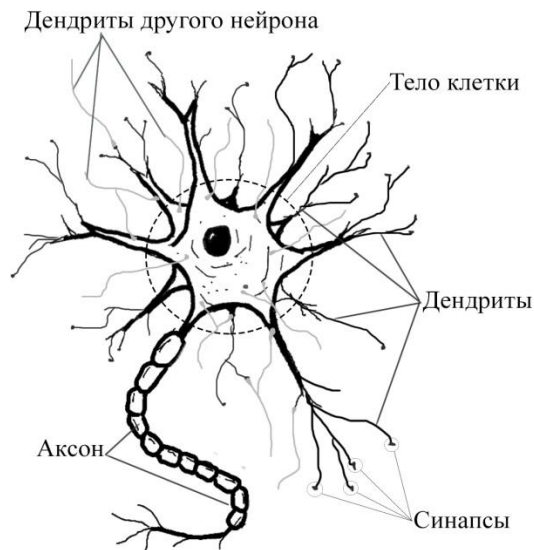


Рис. 1 Схематическое изображение нейрона

Когда суммарный импульс становится выше некоторого порога, нейрон переходит в состояние возбуждения, в нём формируется свой импульс, который далее передаётся по аксону.

С течением времени веса синапса могут изменяться, следовательно, изменяется и реакция соответствующего нейрона.

Несложно изобразить математическую модель описанного выше процесса (Рис. 2):

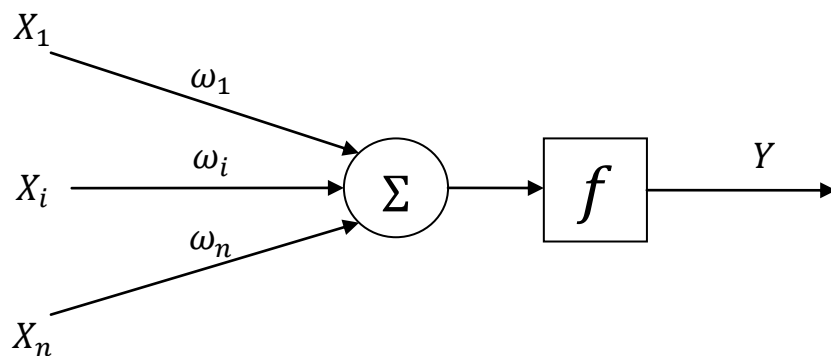


Рис. 2 Математическая модель нейрона

Имеются входы-дендриты, при этом синапсам этих дендритов соответствуют веса $\omega_1, \dots, \omega_i, \dots, \omega_n$. На синапсы поступают импульсы с силами $X_1, \dots, X_i, \dots, X_n$ соответственно. Следовательно, после прохождения синапсов и дендритов к нейрону поступают веса $\omega_1 X_1, \dots, \omega_i X_i, \dots, \omega_n X_n$.

Нейрон суммирует полученный импульс

$$X = \omega_1 X_1 + \dots + \omega_i X_i + \dots + \omega_n X_n = \sum_{i=1}^n \omega_i X_i$$

и меняет его с помощью некоторого нелинейного преобразования $f(X)$ – так называемая «функция активации». Помимо описанных выше параметров, у модели нейрона есть такая характеристика, как пороговый уровень – θ , имеющая то же значение, что и в живом нейроне. Совокупность параметров всех нейронов являет собой параметры нейронной сети.

Таким образом, выходной импульс будет иметь силу:

$$Y = f(X - \theta) = f\left(\sum_{i=1}^n \omega_i X_i - \theta\right)$$

То есть можно сказать, что нейрон полностью описывают веса $\omega_1, \dots, \omega_i, \dots, \omega_n$, пороговый уровень θ и передаточная функция нелинейного преобразования $f(X)$. В качестве входов нейрон получает некоторый вектор чисел X_n , при этом на выходе выдавая число Y .

Существует также модификация модели без порогового уровня. В ней к нейрону добавляется дополнительный вход $X_j = 1$ для любого входящего через него сигнала, а вес $\omega_j = -\theta$.

Разумеется, такие модели являются эквивалентными:

$$f\left(\sum_{i=1}^n \omega_i X_i - \theta\right) = f\left(\sum_{i \neq j} \omega_i X_i - \omega_j X_j\right)$$

Описанная выше модель нейрона является классической, и носит название «формальный нейрон». Помимо неё существует множество других моделей нейрона, отличающихся сложностью вычислений и степенью подобия живому нейрону.

Недостатки математической модели формального нейрона.

При построении данной модели предполагается, что выход вычисляется нейроном мгновенно, поэтому с её помощью невозможно создавать модели систем, в которых существенно внутреннее состояние.

- Формальные нейроны отличаются от живых тем, что не могут синхронно обрабатывать входящие данные.
- Отсутствуют чёткие алгоритмы для выбора активирующей функции.
- Невозможно регулирование процесса работы всей нейронной сети в целом.
- Понятия «весовые коэффициенты» и «порог» в модели являются излишне формализованными. У нейронов биологических сетей порог является величиной динамической, меняющейся в зависимости от состояния сети в целом и активности каждого нейрона в частности, кроме того, значения весов различны при разных проходящих сквозь них сигналах.

1.2. Однослойный перцептрон

Один нейрон способен выполнять элементарные вычисления, однако базовые функции нейронной сети обеспечиваются не отдельно взятыми нейронами, а целыми группами со связями между нейронами внутри каждой. Вообще, простейшая искусственная нейронная сеть представляет собой группу расположенных параллельно нейронов, на вход каждому из них подаётся каждый элемент вектора входных данных $X = (X_1, \dots, X_i)$ (рис. 3). Такая сеть называется однослойным перцептроном¹. В ней, независимо друг от друга, нейроны вычисляют выходное значение. В таком случае, размерность выхода будет равна количеству нейронов в сети, при этом у всех

¹ От англ. “*perceptron*”, также возможен вариант произношения “персептрон”

нейронов должно быть одинаковое количество синапсов, также совпадающее с размером сигнала, который подаётся на вход.

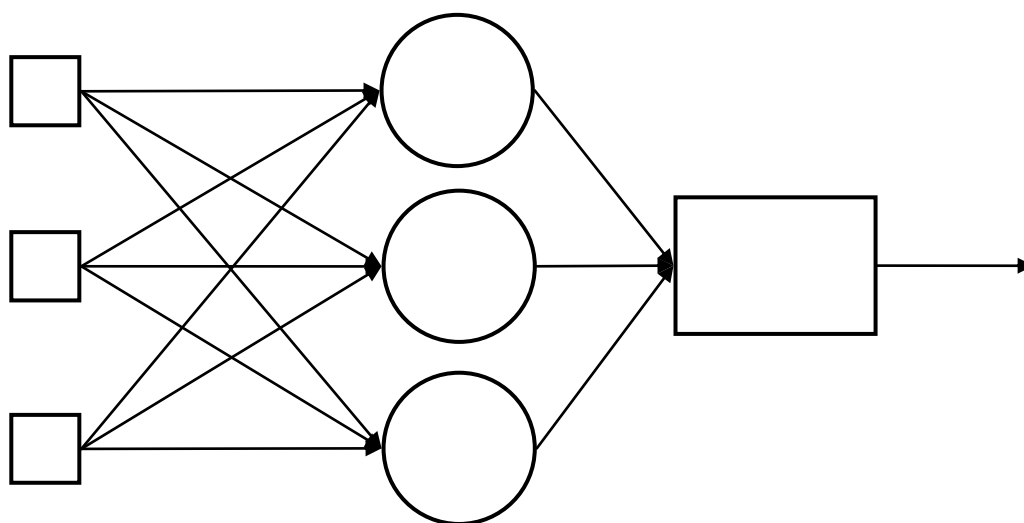


Рис. 3 Схема однослойного перцептрона

Однослойный перцептрон кажется очень простой системой, но при этом способен выполнять ряд полезных функций, таких как вычисление значений логических выражений или классификация образов [2].

1.3. Многослойный перцептрон

Многослойный перцептрон вычисляет выход Y для входа X . То есть, сеть находит значение векторной функции $Y = F(X)$. Следовательно, в качестве условия задачи, которая ставится перед перцептроном, формулируется в виде набора векторов $\{X^1, \dots, X^p\}$. Решение же задачи представимо в виде множества $\{Y^1, \dots, Y^p\}$, при этом $\forall p Y^p = F(X^p)$.

Сеть формирует отображение $F: X \rightarrow Y, \forall X^j \in X (j = \overline{1, p})$. Возможности извлечь полностью отображение из перцептрона нет, но при этом можно посчитать значение образов для произвольного числа входных параметров. Для нейронных сетей отсутствуют точные методики формализации, то есть выяснения смысла компонент векторов входа и

выхода, поэтому на данный момент эта задача решается человеком эмпирическим путём.

Для построения многослойного перцептрона параметры выбираются следующим образом:

- 1) Устанавливается смысл, вкладываемый во множество компонент вектора X . На вход должна подаваться вся необходимая для получения ответа информация, другими словами, необходимо, чтобы вектор входа содержал формализованное условие задачи.
- 2) Выход Y выбирается так, чтобы в его компонентах содержался полностью ответ к задаче.
- 3) Подбирается подходящий вид функции активации. Важно учитывать, что правильный выбор ускоряет процесс обучения, поэтому следует учитывать особенности поставленной задачи.
- 4) Определяется количество слоёв перцептрона и количество нейронов в каждом слое
- 5) основываясь на выборе функции преобразования, устанавливается область, в пределах которой возможны изменение значений для входных и выходных параметров, весовых коэффициентов и порога.
- 6) Весовым коэффициентам и уровням порога присваиваются начальные значения. К ним выставляются следующие требования: они не должны быть как слишком большими, так и слишком малыми, иначе обучение сети замедлится (в первом случае нейроны будут перенасыщены, во втором же большинство выходов будет равно нулю)
- 7) Производится обучение сети, другими словами, подбираются параметры для наилучшего решения задачи. Впоследствии сеть будет уметь решать задачи, подобные тем, на которых она обучалась

- 8) На вход перцептрону подаётся вектор X , являющий собой условие задачи. Формализованное решение данной задачи найдётся при расчёте выхода Y .

1.4. Классификация нейронных сетей.

Вообще, нейронные сети делятся на классы в соответствии со следующими своими характеристиками:

1) Топология сети:

- В *полносвязных* сетях каждый нейрон сети связан со всеми остальными (на данный момент такие сети не используют в связи с высокой сложностью обучения).
- В *слоистых* сетях нейроны расположены отдельными слоями, при этом все нейроны одного слоя имеют связь с нейронами другого. Существуют однослойные и многослойные нейронные сети (простейшие модели описаны выше)

2) Характер связи между нейронами:

- В нейронных сетях *прямого распространения* выход предыдущего слоя сети связан с входом последующего слоя.
- В *рекуррентных* нейронных сетях возможны связи между выходами последующих слоёв и входами предыдущих.

3) Тип подхода к обучению сети:

- *Обучение с учителем* подразумевает обучение на выборках, для которых известны выходные значения. Эти сети часто применяют при решении задач, связанных с классификацией.
- При *обучении сети без учителя* она сама находит особенности между объектами и на основе этого делит их на классы. Данные сети применяются при решении задач кластеризации.
- В *смешанном* обучении оба подхода комбинируются.

4) Вид входящего сигнала:

- *Бинарные* нейронные сети получают на вход сигнал, состоящий лишь из нулей и единиц.
- *Аналоговым* нейронным сетям на вход подаются произвольные сигналы (действительные числа).

5) Структура сети:

- *Однородная* нейронная сеть – такая, в которой для всех нейронов сети используется одна и та же функция активации.
- В *неоднородной* нейронной сети нейроны могут иметь различные активирующие функции.

1.5. Области применения нейронных сетей

Следует заметить, что для задач классификации, например, задачи, связанной с распознаванием символов на изображениях, плохо поддаются алгоритмизации. Если в задаче распознавания ответ нам известен заранее, то в прикладных задачах более сложного типа нейронная сеть выступает в качестве опытной сущности, обладающей экспертным мнением относительно поставленного вопроса.

Рассмотрим, например, сферу медицинского обслуживания. При диагностике пациентов нейронная сеть способна учитывать большой набор характеристических данных (таких как давление, показатели роста и веса, возраста, кардиограмм и т.д.). На основе выводов, полученных от сети, можно составить какие-либо умозаключения относительно состоянии пациентов.

Зачастую нейронные сети используются при классификации компаний по перспективам их развития. Нейронная сеть делает выводы, основываясь на огромных массивах экономических параметров, имеющих большое число связей друг с другом [3].

Популярность использования нейронных сетей при необходимости экспертного мнения обуславливается тем, что сеть имеет вычислительную мощность компьютера и умение распознавать характерные черты и обобщать полученные примеры. При этом сеть способна к обработке большого числа различных показателей, вне зависимости от того, насколько наглядными они являются для человека при вынесении каких-либо выводов.

Применение нейронных сетей возможно везде, где необходимо установить значение неопределенных показателей или особенностей на основе уже полученных или заранее известных данных (таким образом решаются задачи, связанные с классификацией и группировкой данных, с анализом данных временного ряда и в задачах регрессии). При этом необходимо, чтобы данных, на которых бы при составлении выводов основывалась сеть, было достаточно, а также должна существовать какая-либо установившаяся связь между подаваемыми и неизвестными значениями. Также нейронные сети, в частности, сети Кохонена, используются в задачах кластеризации и разведочного анализа.

Продemonстрируем некоторые примеры, представленные из разных областей, в которых применяются нейронные сети [4]:

- Оптическое распознавание символов, в том числе распознавание подписи. Пример: система идентификации подписи, которая учитывает помимо рисунка скорость движение пера, таким образом, процесс подделки чужой подписи затрудняется.
- Обработка изображений. Пример: система, которая сканирует видеоизображения со станций метро для определения заполненности станции, при этом работа такой системы не имеет зависимости от освещенности и движений поездов.
- Прогноз финансовых временных рядов. Пример: «LBS Global Management» сделала заявление об успехах в области финансовых операции, благодаря прогнозированию цен на акции.

- Оценка кредитной способности. Пример: имеются анкетные данные, необходимо определить уровень надежности заемщика.
- Торговля с помощью прямой рассылки. Пример: используя базы данных, определить клиентов, которые могут откликнуться на почтовую рекламу.
- Медицинская диагностика. Пример: прогноз эпилептических припадков.
- Техническая диагностика. Пример: измеряя шум и вибрацию механизма, на ранней стадии выявить неисправность.
- Синтез речи. Пример: система «NetTalk», которая произносит фонемы из текста.
- Прогноз хаотических временных рядов. Множество исследований продемонстрировали отличную способность нейронных сетей к прогнозированию таких рядов.
- Управление процессами. Пример: мониторинг производственных процессов с регулированием параметров, которые являются управляемыми.
- Системы управления двигателями. Пример: используя сенсорные данные, производится оценивание уровня потребления топлива и управление им.
- Анализ лингвистического характера. Пример: имеем сеть с неконтролируемым обучением, которая используется как идентификатор фраз и слов в различных языках.
- Системы детекторов с использованием переключения в реальном времени в физике. Пример: нейронные сети имеют устойчивость к шуму, поэтому допускается использования различных шаблонов в данных физического характера с большим шумом.

Уже сейчас использование нейронных сетей широко распространено в различных сферах жизни, однако пока еще не решён ряд важных вопросов, связанных с риском человеческих жизней либо вложения большого объёма средств. Соответственно, основываясь на информации о задаче, требуется определить допустимый уровень ошибки сети. Компьютеры не являются пока что авторитетным источником информации в связи с тем, что могут ошибаться в связи с разного рода ошибками и неполадками, как со стороны

самой сети, так и со стороны пользователя. Следовательно, нейронные сети имеют скорее характер дополнительного источника знаний, когда отсутствует возможность решить задачу стандартным образом.

Еще одна проблема в применении нейронных сетей заключается в том, что классические сети не умеют объяснять, как именно они получили подобное решение задачи. Интерпретация выходных данных, полученная от сетей, является трудоёмким, а иногда и вовсе невозможным для осуществления процессом. Исключение, пожалуй, составляют лишь некоторые элементарные случаи, не являющиеся привлекательными и интересными с точки зрения подробного изучения.

С недавних пор человек все чаще пытается объединить экспертную систему и нейронную сеть. В таком симбиозе предполагается, что нейронная сеть решает простейшие задачи, а более сложные передаются на оценку экспертной системе. Таким образом, для принятия решений относительно наиболее сложных случаев производится сбор дополнительной информации, либо привлекаются опытные специалисты в данной области [5].

Глава 2. Обучение нейронных сетей.

Во второй главе предоставляется информация о концептуальных подходах к обучению нейронных сетей, а также алгоритмах обучения. Кроме того, дается обзор некоторых методов, применяемых для обучения нейронных сетей,

2.1. Общее описание подходов к обучению нейронных сетей

Умение обучаться является главной отличительной чертой мозга. В рамках искусственных нейронных сетей обучение может представлять собой подбор правильной архитектуры сети и весовых коэффициентов связей, дабы выполнение задачи, поставленной перед сетью, было эффективным. Искусственная нейронная сеть обычно подбирает веса, основываясь на данных, предоставленных ей для обучения, что даёт ей преимущество перед системами, где данные заложены изначально. Стандартно обучение нейронной сети происходит на выборке. В зависимости от степени процесса обучения, происходящему по определенному алгоритму, сеть все лучше и лучше реагирует на входные данные.

Стабилизация весов сети происходит после многократного предъявления примеров, при этом нейронная сеть дает верные ответы на подавляющее большинство примеров из базы данных. Таким образом, можно сказать, что нейронная сеть выучила все примеры, то есть сеть обучена. Благодаря программным реализациям заметно, что величина ошибки в процессе обучения уменьшается. В том случае, когда ошибка достигает нуля, тренировку останавливают, а нейронная сеть считается обученной и готовой для применения на новых данных.

Также важно выделить то, что вся информация, имеющаяся у нейронной сети о задаче, содержится в наборе примеров. В связи с этим качество обучения сети сильно зависит от количества примеров в выборке, предназначенной для обучения, а также от степени полноты описания задачи

этими примерами. Например, не имеет смысла пользоваться нейронной сетью для предсказания финансового кризиса, если в выборке обучения кризисов не предоставлено. Будем считать, что для хорошей тренировки нейронной сети необходимо несколько десятков примеров, а лучше сотен.

Вообще обучение нейронной сети считается достаточно сложным, наукоемким процессом. Самые алгоритмы обучения имеют самые разные параметры и настройки, и для управления ими необходимо понимание их влияния.

Существующие на данный момент методы обучения нейронных сетей делятся на следующие классы:

- *Детерминированный метод* – такой, при котором параметры нейронной сети изменяются итеративно, в зависимости от значений параметров, входных данных, а также основываясь на значениях выхода: тех, которые необходимо получить и тех, которые получились в процессе вычислений. Наиболее ярким примером служит метод обратного распространения ошибки.
- В *стохастических методах* параметры сети меняются без каких-либо закономерностей, случайно. Изменения принимаются лишь в том случае, когда они приводят к улучшению результатов. К проблемам данного метода обучения относится так называемая “ловушка локального минимума” (рис. 4).

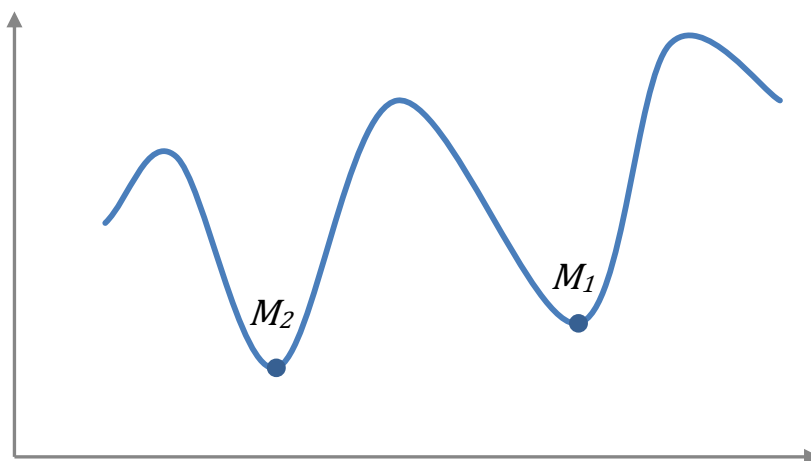


Рис. 4 Проблема локальных минимумов

Предположим, что значение ошибки изначально близко к точке M_1 , либо равно её значению. При малых произвольных шагах коррекции параметров любые отклонения от нее увеличат значение ошибки, и, следовательно, не будут приняты. Получается, что невозможно найти наименьшее значение ошибки в точке M_2 . Если же случайные шаги корректировки достаточно велики, ошибка не сможет установиться в одном из двух минимумов, так как будет меняться слишком резко.

В качестве варианта решения проблемы предполагается постепенное уменьшение среднего размера случайных корректирующих шагов. С равной вероятностью же значение ошибки принимает все возможные значения при больших произвольных шагах. В случае, когда размер случайных шагов корректировки изменяется постепенно, значение ошибки будет останавливаться в точке M_2 в течение некоторого времени. При последующем уменьшении шага корректировки значение ошибки может «застревать» не только в точке M_2 , но и в M_1 . Дальнейшее непрерывное уменьшение шага может привести к исходу, когда преодолевается локальный минимум M_1 , при этом попадая в локальный минимум M_2 .

2.2. Алгоритмы обучения сетей

Алгоритм обучения подразумевает процедуру, использующую правила обучения для настройки весов. Перечислим наиболее распространенные алгоритмы:

- Обратное распространение
- Обучение Кохонена
- Левенберга- Марквардта
- Сопряженных градиентов
- Псевдообратный
- Быстрое распространение
- Анализ главных компонентов
- Пометка ближайших классов
- Метод «К-средних»

- «Дельта – дельта с чертой» ($\Delta - \bar{\Delta}$)
- Квази-Ньютоновский
- Обучающий векторный квантователь
- Линейный дискриминантный анализ
- Обобщенно-регрессионная нейронная сеть
- Пошаговый прямой или обратный отбор входных данных
- Радиальная (под-) выборка
- Метод «К-ближайших соседей (KNN)»
- Вероятностная нейронная сеть
- Проекция Саммона
- Генетический алгоритм отбора входных данных
- Установка явных отклонений
- Установка изотропных отклонений

Как только определяется количество слоев и элементов в каждом из них, требуется найти значения для весов нейронной сети, эти значения должны минимизировать ошибку прогноза, которую выдвигает сеть. Для этой цели и служат алгоритмы обучения. Данный процесс характеризуется как подгонка модели, реализуемая сетью относительно обучающих данных, которые у нас имеются. Определение ошибки для какой-либо конфигурации сети происходит путем прогона через сеть имеющихся наблюдений и сравнении желаемых значений с реально выдаваемыми выходными значениями.

Отдадим предпочтение алгоритмам, позволяющим обучить сеть за достаточно малое число шагов и требующих небольшое количество дополнительных параметров. Этот выбор связан с тем, что обучение происходит на машинах (компьютерах) сравнительно малой производительности с небольшим объемом оперативной памяти.

Все основные алгоритмы обучения строятся на функции, оценивающей качество работы нейронной сети. При этом имеем некий алгоритм, подстраивающий параметры системы, которые меняются в зависимости от значений оценки, которая была получена. Такие алгоритмы обычно являются

достаточно простыми, но не позволяют построить хорошую систему управления за приемлемое время.

Теория об обучении нейронных сетей включает в себя рассмотрение трёх основных характеристик процесса обучения по примерам: сложность вычислений, сложность обучающей выборки и ёмкость [4]. Последняя представляет собой информацию о том, как много образцов из выборки сеть способна запомнить, а также о функциях и границах в принятии решений.

Сложность же обучающей выборки показывает, какое число образцов необходимо сети для получения способности к обобщению. Если набор примеров будет слишком малым, то сеть может оказаться “переобученной”: на обучающей выборке она будет хорошо функционировать, при этом плохо функционируя на тестовых образцах, подверженных сходному статистическому распределению [6].

2.3. Концептуальные подходы к обучению сетей.

Обучение с учителем.

Процесс обучения с учителем полагает, что для каждого из множества входных векторов будет существовать целевой вектор, который и представляет из себя желаемый выход. Другими словами, задаётся массив пар векторов $\{(X^M, s^M)\}$, где $X^M \in X$ – описывающий условие задачи вектор, а $s^M \in Y$ – уже известное для заданного вектора X^M решение задачи. Такая пара называется *обучающей*. Для обучения сети обычно требуется некоторое количество таких пар. Оно происходит следующим образом: вычисляется выход сети, основываясь на предоставленной информации о выходном векторе, после чего выход и соответствующий ему целевой вектор сравниваются. Разность этих значений (ошибка) считывается сетью с

помощью обратной связи, после чего весовые коэффициенты меняются по принципу алгоритма, выбранного для минимизации ошибки. Иначе говоря, сеть меняет свои параметры так, чтобы получалось требуемое отображение $X \rightarrow Y$. При этом важно учитывать, что размер набора $\{(X^M, s^M)\}$ должен быть достаточно большим, дабы алгоритм сформировал желаемое отображение [7]. Векторы из обучающего набора представляются сети поочередно, после чего происходит вычисление ошибок и подстраивание весового коэффициента каждого вектора до того момента, пока общая ошибка на всём обучающем наборе не будет на достаточно низком уровне. Стоит отметить, что коррекция весовых коэффициентов происходит только при условии ошибочного ответа [8].

Данная модель впервые была предложена Фрэнком Розенблаттом в 1957 году [9].

Хотя обучение с учителем и достигло определенных высот в применении, оно подвергалось критике из-за биологической неправдоподобности модели. Невозможно представить в мозге механизм обучения, способный сравнивать требуемые и действительные значения выходов, после чего корректируя данные посредством обратной связи. Даже допустив существование подобного механизма в мозге, остаётся вопрос: каким образом определяются желаемые выходы?

Обучение без учителя.

С точки зрения биологии, обучение без учителя представляет собой более правдоподобную модель. В процессе исследования этой модели обучения Тойво Кохоненом [10], а также многими другими, установлено, что она не требует целевого вектора выходов, значит, нет необходимости в сравнении полученного выхода с изначально определенными эталонными ответами. Массив обучающих данных представляет собой лишь входные векторы. Наилучшее значение для выхода определяется самим алгоритмом. Обычно он подстраивает весовые коэффициенты таким образом, чтобы на выходе

векторы получались согласованными, другими словами, если входные векторы будут достаточно близкими, то их предъявление алгоритму должно давать схожие выходные векторы. Во время обучения сетью выделяются характерные черты обучающей выборки, на основе чего схожие векторы формируются в некоторые группы. При поступлении входного вектора из данной группы даёт какой либо конкретный вектор выхода, но до начала процесса обучения неясно, какой именно выход получится при направлении в сеть данной группы входных векторов выборки. То есть важен вопрос интерпретации выхода сети, основанный на процессе обучения. Это представляется вполне возможным, так как определить связь входа и выхода, установленную сетью, обычно несложно.

2.4. Другие методики обучения

Метод Хебба

Большая часть существующих на данный момент алгоритмов обучения основана на концепциях Дональда Хебба, впервые описанных им еще в 1949 году [11]. Опираясь на своих исследования в области психологии и физиологии, он предположил, что вес соединения будет возрастать при активации нейрона-источника и нейрона-приёмника сигнала. Следовательно, наиболее используемые пути прохождения сигнала в сети получают более высокий приоритет, что объясняет такие явления, как обучение через повторение одинаковых действий и существование привычек.

Искусственная нейронная сеть, основанная на принципе обучения по Хеббу, определяет весовые коэффициенты как произведение уровней возбуждения нейрона, от которого идёт сигнал и нейрона, который этот сигнал получает. Математически это описывается следующим образом:

$$\omega_{ij}(m + 1) = \omega_{ij}(m) + \gamma W_i W_j ,$$

где:

- $\omega_{ij}(m)$ – значение весового коэффициента от нейрона i к нейрону j до подстраивания,
- γ – коэффициент, характеризующий скорость обучения,
 $\omega_{ij}(m + 1)$ – значение весового коэффициента от нейрона i к нейрону j после процесса подстройки,
- W_i – выход i и вход j ,
- W_j – выход j .

Обучение по Хеббу активно использовалось при разработке нейронных сетей в течение длительного времени. Тем не менее, существуют и более эффективные методики обучения сетей, разработанные и усовершенствованные за последние 20 лет. Например, развитие получили алгоритмы, основанные на принципе обучения с учителем, имеющие более обширный набор характеристик для обучающей выборки, и обладающие большими скоростями, нежели модели обучения, основанные на методиках Хебба

Соревновательный метод обучения

В обучении Хебба большое число выходных нейронов имеют возможность одновременно возбуждаться в отличие от нейронов при обучении методом соревнования, где им приходится конкурировать между собой за возможность активации. Другими словами, из всех выходных нейронов выбирается один, имеющий самый большой выход. Подобный алгоритм схож с методиками обучения живых нейронных сетей. С помощью него появляется возможность классификации входных сигналов: из схожих образцов сетью составляется группа, которая уже сама служит в качестве единого сигнала-образца. Важно отметить, что каждый выходной нейрон “несёт ответственность” лишь за один класс. Соответственно, количество классов, с которыми нейронная сеть может взаимодействовать, эквивалентно

числу нейронов выхода.

В процессе данного обучения веса изменяются только у нейрона, “одержавшего победу”, в связи с чем образец и входной сигнал становятся чуть более схожи.

Генетические алгоритмы

Генетическими алгоритмами называют класс алгоритмов, базирующийся на моделях процессов развития биологической популяции.

Пусть популяция – множество векторов $P = \{\rho^1, \dots, \rho^L\} = \{\rho^l\}$, при этом L определяет размер популяции. Каждый вектор ρ^l обозначает особь и представляет собой набор параметров, описывающих её.

Необходимо найти минимум функции $E = E(\rho^l)$, вычисляющей ошибку.

Также определены правила, согласно которым каждая составляющая множества P имеет способность к эволюционным изменениям:

- В случае, когда значение функции $E = E(\rho^0)$ достаточно малое, особь, соответствующая вектору, определяется как удачная и размножение данной особи имеет повышенный приоритет.
- В обратном случае, когда $E = E(\rho^0)$ достаточно велико, это означает, что ρ^0 – неудачная особь, следовательно, она имеет повышенную вероятность гибели и малую вероятность способности размножиться.
- Также определены *мутационные процессы* в рамках популяции: каждый элемент множества P способен мутировать, значит, менять своё значение на некоторую малую величину: $\rho^{0*} = \rho^0 + \Delta\rho$. $\Delta\rho$ представляет собой вектор, который характеризует степень мутации, при этом имеет небольшое по модулю значение.
- Процесс *размножения* заключается в делении каждой точки (особи), основываясь на определенной вероятности размножения. Законы, по которым происходит размножение, определяются выбранной моделью.
- *Гибель* особи возможна в соответствии с определенной вероятностью гибели. Другими словами, точка удаляется из множества векторов P .

Пока не существует точной модели для теории эволюции, соответственно, отсутствуют правила, по которым можно вычислить точные вероятностные

значения для размножения и гибели особей. Поэтому определить, какие алгоритмы являются наилучшими можно лишь эмпирически [12].

2.5. Методы увеличения скорости процесса обучения сети.

Стоит отметить, что нейронные сети – это направление искусственного интеллекта, часто используемое для решения задач распознавания образов и оптимизации. В наше время уже разработано огромное количество моделей нейронных сетей, предназначенных для решения прикладных задач. Каждая модель имеет свои предложенные методы обучения. Но, тем не менее, работы по созданию новых и усовершенствованию имеющихся моделей не прекращаются.

Выделим два основных этапа в создании нейронной сети:

- 1) Этап выбора модели сети, ее структура и алгоритм обучения. Назовем его структурным синтезом.
- 2) Этап обучения нейронной сети и проверки результатов. Он будет иметь название “параметрический синтез”.

Решения о возвращении на предыдущие стадии структурного или параметрического синтеза принимаются в зависимости от результатов проверки.

Имеется некоторое количество методов, позволяющих уменьшить время обучения многослойной нейронной сети, и основанных на принципе достаточности (ошибка сети не больше некоторого значения). В качестве таких методов рассмотрим изменение весовых коэффициентов сети, корректировка шага изменения весовых коэффициентов и реорганизацию классов.

Будем обучать сеть до тех пор, пока ошибка не будет близка к нулю. Обычно это ведет к большим тратам временных ресурсов в связи с тем, что, во время обучения ошибка не превышала значения, значительно удаленного от нуля.

Исходя из условия задачи, и определяется степень достаточности обучения.

Идея принципа достаточности подразумевает отказ от стремления к идеалу при поиске решения. То есть имеется в виду, что 100% точно подобрать решение требуется не всегда. Иногда бывает достаточно, чтобы ошибка не превышала некоторого заданного δ , максимального значения, которое зависит от характера обучающей выборки [13].

Полноту и противоречивость обучающей выборки рассмотрим в качестве ее характеризующих параметров.

Обеспеченность классов обучающими наборами характеризуется полнотой. Количество обучающих наборов для каждого класса должно в 4-5 раз превышать количество признаков, используемых в данных наборах. Расчет полноты можно произвести по следующей формуле:

$$F_{O.B.} = \frac{n_{F_{O.B.}}}{N}$$

где:

- $n_{F_{O.B.}}$ – количество классов, удовлетворяющих правилу, описанному выше,
- N – число всех классов.

Наборы, содержащие объекты, определенные к различным классам, но при этом содержащие одинаковые классификационные признаки, считаются противоречивыми. Противоречивость все обучающей выборки высчитывается по формуле:

$$I_{O.B.} = \frac{M_I}{M}$$

где:

- M_I – число наборов, содержащих противоречия,
- M – число всех наборов выборки.

Следовательно, от того, насколько большими будут величины $F_{O.V.}$ и $I_{O.V.}$, зависит, насколько большие значения может принимать величина δ , и, соответственно, насколько быстро будет проходить обучение сети [14].

2.6. Алгоритм обратного распространения ошибки

При работе с многослойной нейронной сетью отсутствует информация об оптимальных величинах выходов нейронов каждого из слоёв, исключая последний. Соответственно, если нейронная сеть имеет более одного слоя, обучение, основанное лишь на значениях ошибок выходов из сети будет невозможным. Данную проблему можно попробовать решить, определив наборы сигналов на выходе из каждого слоя сети, что представляет собой энергозатратный и не всегда возможный процесс. Кроме того, возможно подстраивать веса динамически, при этом обычно выбирают самые слабые связи и подвергаются незначительным изменениям, которые сохраняются лишь в том случае, когда сеть стала лучше работать, то есть уменьшилась ошибка на выходе. Данный эмпирический метод является трудоёмким вычислительным процессом, несмотря на простоту описываемых действий. Наиболее привлекательным вариантом является распространение ошибки от выходов искусственной нейронной сети к её входам, что является обратным направлением относительно обычного рабочего режима сети. Подобный алгоритм носит название «метод обратного распространения ошибки» [15].

При обучении нейронной сети методом обратного распространения ошибки по всем слоям сети производится два прохода: прямой и обратный. При первом производится приём входных данных, а также распространение их в направлении выходов. Во время второго этапа вычисляется ошибка, которая

впоследствии распространяется обратно, а также корректируются весовые коэффициенты. Когда процесс обучения сети будет закончен, данные будут распространяться лишь в прямом направлении. Важно отметить, что хотя тренировка сети может длиться достаточно долго, но при этом уже обученная сеть способна быстро вычислять результаты по полученным данным. Для метода обратного распространения ошибки имеются различные модификации, направленные на уменьшение времени обучения сети.

Нейронная сеть, имеющая лишь один слой, имеет множество ограничений на шаблоны данных входа, многослойная нейронная сеть подобного изъяна лишена.

Архитектура сети, на которой применяется метод

Рассмотрим сеть с множеством слоёв, у которой при этом один слой нейронов является скрытым (слой Z , состоящий из нейронов Z_1, Z_j, Z_p). (рис. 5).

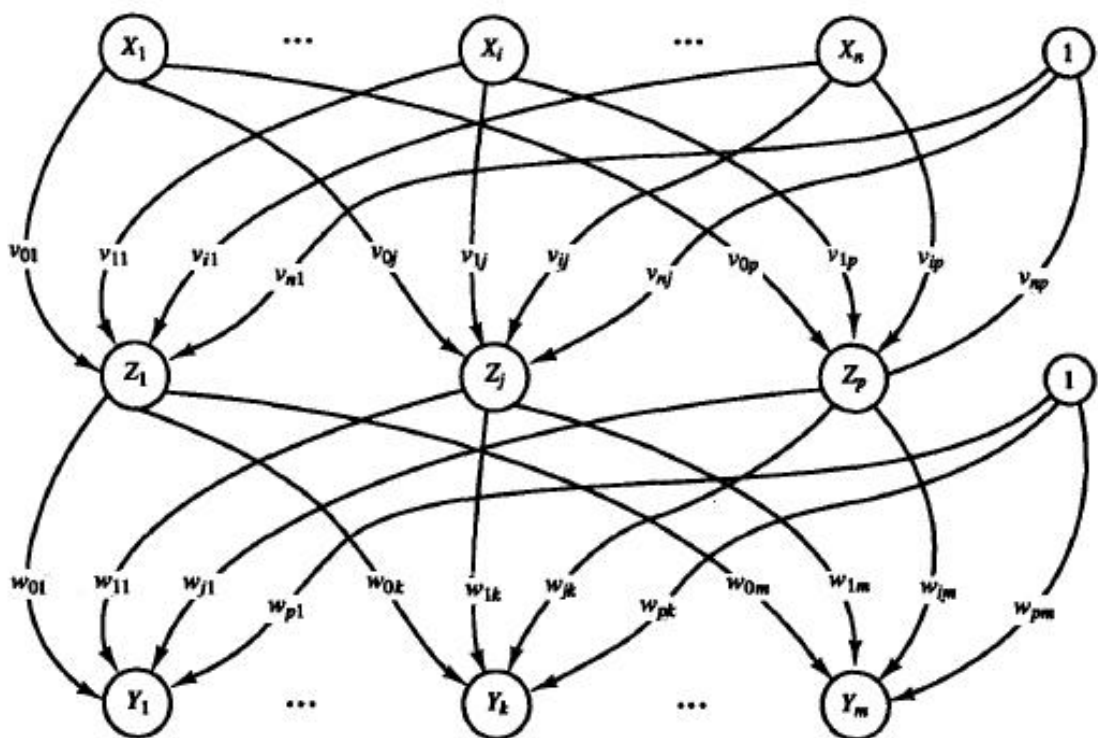


Рис. 5 Нейронная сеть, применяемая в методе обратного распространения ошибки с одним скрытым слоем

Выходы из сети – нейроны множества Y , а также нейроны скрытого слоя, имеют смещение. Элементу скрытого слоя Z_j соответствует смещение v_{0j} , элементу Y_k выходного слоя – смещение ω_{0k} . Данные смещения символизируют веса связей нейронов, всегда имеющих на выходе значение 1. Обычно они не имеют внешнего изображения и лишь подразумеваются.

Помимо этого, стрелками обозначено направление при перемещении информации на этапе распространения входных данных к выходам. Во время тренировки сети данные распространяются также и в обратном направлении.

2.7. Обобщение информации о нейронных сетях.

Методик и алгоритмов обучения нейронных сетей существует множество, каждая из которых имеет свои преимущества и недостатки, связанные, в первую очередь с наличием способности обрабатывать входные данные различных типов, а также с временем тренировки сети на основе обучающей выборки. Для решения различного типа задач выбирается подходящий тип нейронной сети и алгоритм её обучения, при этом оптимальный вариант выбирается эмпирическим путём, либо на основе чьего-либо опыта.

Нейронные сети качественно строят сглаживание или подбор значений функции по точкам, и могут подобрать большинство функций, как одной, так и нескольких переменных.

Исследования, проведенные нейронными сетями достаточно наглядные, и имеют преимущество перед другими вычислительными методами. Это следует из того, что теоретические требования к моделям, полученным нейронными сетями, очень гибкие, также им нужны относительно небольшие объемы предварительных знаний для формирования задачи.

Нейронные сети представляют собой мощный механизм для решения задач в различных областях. Однако они никогда не смогут заменить людей в этом процессе. Сети необходимо использовать для обобщения, а не определения данных. Нейронные сети – адаптивны, они подражают решению человека, но они не сообщают, какой именно критерий решения задачи стоит брать во внимание перед сбором данных. Также обучающие машины используются для формализации знаний из реального мира, но сами машины не способны при этом генерировать принципы формализации.

Заключение.

Все поставленные в работе цели и задачи были выполнены. Были описаны различные математические модели искусственных нейронных сетей, рассмотрены принципы классификации сетей, изучены концептуальные подходы к их обучению, а также была дана краткая информация о некоторых методах обучения. Кроме того, в работе предоставлена информация о возможности увеличения скорости обучения сети, а также подробно рассказывается о сферах применения нейронных сетей.

Список используемой литературы

1. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр "Академия", 2005. — 176 с.
2. Короткий С. Нейронные сети: основные положения.
<http://www.gotai.net/documents/doc-nn-002.aspx>.
3. Галушкин А.И. Применения нейрокомпьютеров в финансовой деятельности. – Новосибирск: Наука, 2002. – 215с.
4. Белов В.Г. О перспективах искусственного интеллекта. М.: Дело, 2006. – 82 с.
5. Ефимов Д.В. Нейросетевые системы управления. М.: Высшая школа, 2002. — 184 с.
6. Круглов В.В., Борисов В.В. Искусственные нейронные сети. Теория и практика. М.: Горячая линия-Телеком, 2002. – 380 с.
7. Kröse B., Smagt P. An introduction to Neural Networks. University of Amsterdam, 1996.
8. Smith L. An Introduction to Neural Networks. Unpublished draft, University of Stirling, 2001.
(<http://www.cs.stir.ac.uk/~lss/NNIntro/InvSlides.html>).
9. Rosenblatt F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. Psychological Review, 1958.
10. Kohonen, T. (1989/1997/2001), Self-Organizing Maps, Berlin — New York: Springer-Verlag. First edition 1989, second edition 1997, third extended edition 2001, ISBN 0-387-51387-6, ISBN 3-540-67921-9
11. John Wiley & Sons. *The Organization of Behaviour*. 1949. ISBN 978-0-471-36727-7
12. Chellapilla K., Fogel D.B. *Evolution, neural networks, games and intelligence* / Proc. IEEE, 1999. Vol.87, No.9. - pp.1471-1496.
13. Обучение нейронной сети: <http://www.aiportal.ru/articles/neural-networks/learning-neunet.html>

14. Парадигмы обучения нейронных сетей:

<http://apsheronk.bozo.ru/Neural/Lec3.htm>

15. Алгоритмы обучения нейронных сетей будущего:

<http://ai.h16.ru/note/algo/algo.php>

16. Васенков Д. В. Методы обучения искусственных нейронных сетей, Журнал «Компьютерные инструменты в образовании». № 1, 2007 г.