СОДЕРЖАНИЕ

ВВЕДЕНИЕ 3

1 МАТЕМАТИЧЕСКИЕ ОСНОВЫ 5

2 ОБЛАСТИ ПРИМЕНЕНИЯ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ 9

2.1 Распознавание и классификация образов 9

2.2 Другие области использования искусственных нейронных сетей 11

3 ПО ДЛЯ РАБОТЫ С ИСКУССТВЕННЫМИ СЕТЯМИ 12

3.1 Multiple Back-Propagation 12

3.2 Пакет STATISTICA Neural Networks 12

3.3 Matlab Neural Network Toolbox 13

ЗАКЛЮЧЕНИЕ 15

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ 16

ВВЕДЕНИЕ

Теория нейронных сетей включают широкий круг вопросов из разных областей науки: биофизики, математики, информатики, схемотехники и технологии. Поэтому понятие "нейронные сети" детально определить сложно.

Искусственные нейронные сети (НС) – совокупность моделей биологических нейронных сетей. Представляют собой сеть элементов – искусственных нейронов – связанных между собой синаптическими соединениями. Сеть обрабатывает входную информацию и в процессе изменения своего состояния во времени формирует совокупность выходных сигналов.

Большинство моделей НС требуют обучения. В общем случае, обучение – такой выбор параметров сети, при котором сеть лучше всего справляется с поставленной проблемой. Обучение – это задача многомерной оптимизации, и для ее решения существует множество алгоритмов.

Выделим характерные черты искусственных нейросетей как универсального инструмента для решения задач:

1. НС дают возможность лучше понять организацию нервной системы человека и животных на средних уровнях: память, обработка сенсорной информации, моторика;

2. НС – средство обработки информации:

* гибкая модель для нелинейной аппроксимации многомерных функций;
* средство прогнозирования во времени для процессов, зависящих от многих переменных;
* классификатор по многим признакам, дающий разбиение входного пространства на области;
* средство распознавания образов;
* инструмент для поиска по ассоциациям;
* модель для поиска закономерностей в массивах данных.

3. НС свободны от ограничений обычных компьютеров благодаря параллельной обработке и сильной связанности нейронов.

4. В перспективе НС должны помочь понять принципы, на которых построены высшие функции нервной системы: сознание, эмоции, мышление.

Существенную часть в теории нейронных сетей занимают биофизические проблемы. Для построения адекватной математической модели необходимо детально изучить работу биологических нервных клеток и сетей с точки зрения химии, физики, теории информации и синергетики.

1. МАТЕМАТИЧЕСКИЕ ОСНОВЫ

Искусственная нейронная сеть (ИНС) – это существенно параллельно распределенный процессор, который обладает способностью к сохранению и репрезентации опытного знания. Она сходна с мозгом в двух аспектах:

* знание приобретается сетью в процессе обучения;
* для сохранения знания используются силы межнейронных соединений, называемые также синаптическими соединениями.

Работа нейронной сети (НС) состоит в преобразовании входных сигналов во времени, в результате чего меняется внутреннее состояние сети и формируются выходные воздействия. Обычно НС оперирует цифровыми, а не символьными величинами. Большинство моделей НС требуют обучения. В общем случае обучение — это такой выбор параметров сети, при котором сеть лучше всего справляется с поставленной проблемой. Обучение — это задача многомерной оптимизации, и для ее решения существует множество алгоритмов.

Современные искусственные НС демонстрируют следующие ценные свойства:

* обучаемость. Выбрав одну из моделей НС, создав сеть и выполнив алгоритм обучения, мы можем обучить сеть решению задачи, которая ей по силам.
* способность к обобщению. После обучения сеть становится нечувствительной к малым изменениям входных сигналов (шуму или вариациям входных образов) и дает правильный результат на выходе.
* способность к абстрагированию. Если предъявить сети несколько искаженных вариантов входного образа, то сеть сама может создать на выходе идеальный образ, с которым она никогда не встречалась.

К задачам, успешно решаемым НС на данном этапе их развития, относятся:

* распознавание зрительных, слуховых образов;
* ассоциативный поиск информации и создание ассоциативных моделей, синтез речи, формирование естественного языка;
* формирование моделей и различных нелинейных и трудноописываемых математически систем, прогнозирование развития этих систем во времени;
* системы управления и регулирования с предсказанием;
* разнообразные конечные автоматы: системы массового обслуживания и коммутации, телекоммуникационные системы;
* принятие решений и диагностика, исключающие логический вывод, особенно в областях, где отсутствуют четкие математические модели: в медицине, криминалистике, финансовой сфере.

На рисунке 1.1 показана структура пары типичных биологических нейронов. Дендриты идут от тела нервной клетки к другим нейронам, где они принимают сигналы в точках соединения, называемых синапсами. Принятые синапсом входные сигналы подводятся к телу нейрона. Здесь они суммируются, причем одни входы стремятся возбудить нейрон, другие – воспрепятствовать его возбуждению. Когда суммарное возбуждение в теле нейрона превышает некоторый порог, нейрон возбуждается, посылая по аксону сигнал другим нейронам. У этой основной функциональной схемы много усложнений и исключений, тем не менее большинство искусственных нейронных сетей моделируют лишь эти простые свойства.

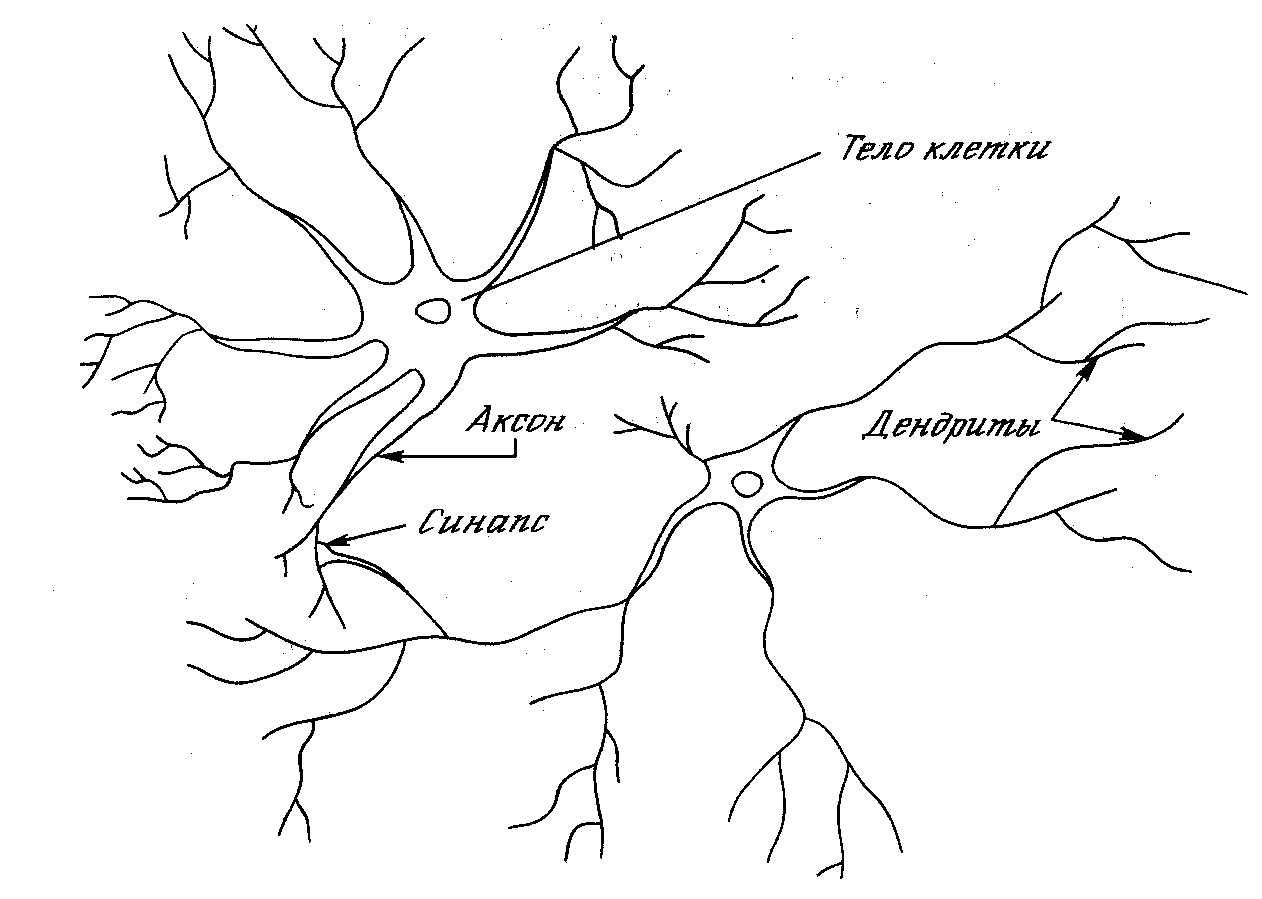


Рисунок 1.1 – Биологический нейрон

Основным элементом нейронной сети является нейрон, который осуществляет операцию нелинейного преобразования суммы произведений входных сигналов на весовые коэффициенты:

 (1)

где X = (x1,x2,…,xn) – вектор входного сигнала; W=(w1,w2,…,wn) – весовой вектор; T – порог; f – так называемая функция активации.

Схема нейронного элемента изображена на рисунке 1.2. Каждому i-му входу нейрона соответствует весовой коэффициент wi (синапс), который характеризует силу синаптической связи по аналогии с биологическим нейроном.

Весовой вектор W, порог T и пороговая функция f определяют поведение любого нейрона – его реакцию на входные данные. Величина веса wi определяет степень влияния входа i на выход нейрона, а знак – характер влияния. Каждый из входов связан с некоторым источником информации (рецептор, формирующий признаки, распределительная ячейка, выход другого нейрона). Положительные веса характерны для возбуждающих связей, способствующих повышению активности нейрона. Отрицательные, как правило, соответствуют тормозящим связям.



Рисунок 1.2 – Искусственный нейрон

Порог, если он используется, является характеристикой, задающей начальный уровень активности (при нулевом входе) и помогающей настроить нейрон на пороговую функцию. Изменение порога эквивалентно сдвигу пороговой функции по оси абсцисс. Ряд авторов [2] вводят дополнительный вход нейрона x0, всегда равный 1, и обозначают порог как его вес w0. Это позволяет упростить выражение (1) и математическую запись некоторых алгоритмов обучения. Однако на практике при программной реализации это не приводит к экономии времени и способствует ошибкам, кроме этого, порог может настраиваться иначе, чем весовой вектор.

Функция активации используется для ограничения выхода нейрона в заданном диапазоне и для нелинейного преобразования взвешенной суммы. Последнее позволяет нейронному классификатору аппроксимировать любую нелинейную границу между классами в пространстве образов. Функция активации выбирается для конкретной задачи и является неизменной характеристикой отдельного нейрона.

Могут использоваться следующие функции активации и их гибриды:

1) линейная функция y = Ax;

2) пороговая функция ;

3) биполярная пороговая функция ;

4) сигмоидная функция

; (2)



Рисунок 1.3 – Сигмоидные функции активации

5) биполярная сигмоидная функция (рисунок 1.3)

; (3)

6) гиперболический тангенс (рисунок 1.4)

. (4)



Рисунок 1.4 – Гиперболический тангенс

Для классификаторов чаще всего применяют функции (2) – (4), поскольку они нелинейные и хорошо дифференцируются.

Веса и порог конкретного нейрона являются настраиваемыми параметрами. Они содержат знания нейрона, определяющие его поведение. Процесс настройки этих знаний с целью получения нужного поведения называется обучением.

Нейронная сеть – совокупность нейронных элементов и связей между ними. Слоем нейронной сети называется множество нейронных элементов, на которые в каждый такт времени параллельно поступает информация от других нейронных элементов сети. Помимо слоев нейронов часто используется понятие входного распределительного слоя. Распределительный слой передает входные сигналы на первый обрабатывающий слой нейронных элементов (рисунок 1.5).



Рисунок 1.5 – Распределительный слой

У любой нейронной сети можно выделить 2 режима работы: обучение и воспроизведение. На этапе обучения настраиваются веса и пороги всех слоев. Воспроизведение является этапом обработки информации, следующим за обучением, при этом веса и пороги, как правило, не изменяются. Большое влияние на функционирование сети оказывает ее топология – архитектура слоев и связей между нейронами.

На способе обработки информации решающим образом сказывается наличие или отсутствие в сети обратных связей. Если обратных связей нет (каждый нейрон получает информацию только от нейронов предыдущих слоев), то обработка информации происходит в одном направлении за число тактов, равное числу слоев. При наличии обратных связей информация может проходить через сеть много раз до достижения какого-либо условия. В случае если условие достигнуто, говорят, что сеть стабилизировалась. В общем случае сходимость не гарантируется. Тем не менее наличие обратных связей позволяет решать задачи с привлечением меньшего числа нейронов, что ускоряет процесс обучения.

Для лучшего понимания данного вопроса приведем одну из возможных классификаций НС в зависимости от различных характеристик [2].

1. По типу входной информации:
   * аналоговые НС (используют информацию в форме действительных чисел);
   * двоичные НС (оперируют с информацией, представленной в двоичном виде).
2. По характеру обучения:

* с учителем (известно входное пространство решений НС);
* без учителя (НС формирует выходное пространство решений только на основе входных воздействий – самоорганизующиеся сети).

1. По характеру настройки синапсов:

* сети с фиксированными связями (весовые коэффициенты НС выбираются сразу, исходя из условия задачи);
* сети с динамическими связями (в процессе обучения происходит настройка синаптических связей).

1. По методу обучения:

* НС с алгоритмом обратного распространения ошибки;
* НС с конкурентным обучением;
* НС, использующие правило Хебба;
* НС с гибридным обучением, в которых используются различные алгоритмы обучения.

1. По характеру связей:

* НС с прямыми связями;
* НС с обратным распространением информации.

1. По архитектуре и обучению:

* персептронные сети с прямыми связями;
* самоорганизующиеся НС (НС Кохонена, НС адаптивного резонанса, рециркуляционные НС);
* НС с обратными связями (НС Хопфилда, НС Хэмминга, двунаправленная ассоциативная память, рекуррентные НС);
* гибридные НС (НС встречного распространения, НС с радиально-базисной функцией активации).

1. ОБЛАСТИ ПРИМЕНЕНИЯ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

Однонаправленные нейронные сети с сигмоидальной функцией активации широко применяются на практике, составляя важное звено процесса выработки решений. Далее приводится несколько приложений, позволяющим подчеркнуть универсальность и разнородность функций, которые искусственные нейронные сети могут выполнять.

2.1 Распознавание и классификация образов

Распознаванием и классификацией образа будем называть его идентификацию и отнесение к соответствующему классу данных. При решении этой задачи нейронная сеть может выполнять функцию как экстрактора свойств, так и классификатора, приписывающего образ конкретному классу. Однако чаще всего экстракция свойств производится на отдельном этапе предварительного преобразования данных. В качестве примера рассмотрим определение достоверности цифровой подписи.

Определение достоверности цифровой подписи – это последовательность действий, в результате которых дается ответ на вопрос: принадлежит данная подпись конкретному человеку или нет. В примере использован пакет STATISTICA Neural Networks.

Упростим задачу – рассмотрим одиночный символ, рукописная буква “А”. Как будет видно ниже, сложность данной задачи – в структуре данных и их представления для обучения нейронной сети.

В качестве исходных данных берётся набор картинок. Все картинки разделены на две группы:

1. подпись, принадлежащая конкретному человеку;
2. подпись, не принадлежащая конкретному человеку (искаженная подпись).

Внутри каждой группы картинки различны. Действительно, ни один человек не может расписаться два раза абсолютно одинаковым образом: изменяются стартовая точка, наклон, форма отдельных элементов. Также присутствуют изображения с “шумами”.

Приведём несколько примеров изображений для каждой группы.

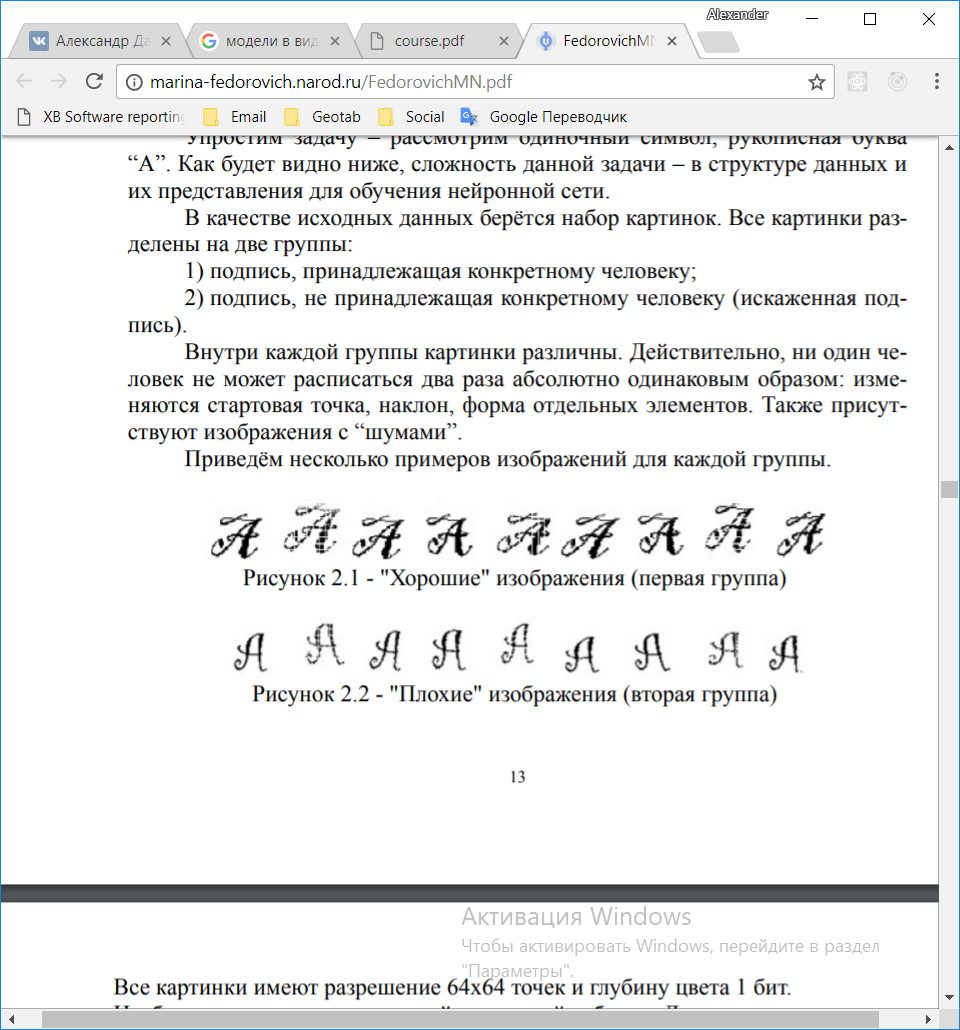


Рисунок 2.1 – «Хорошие» изображения

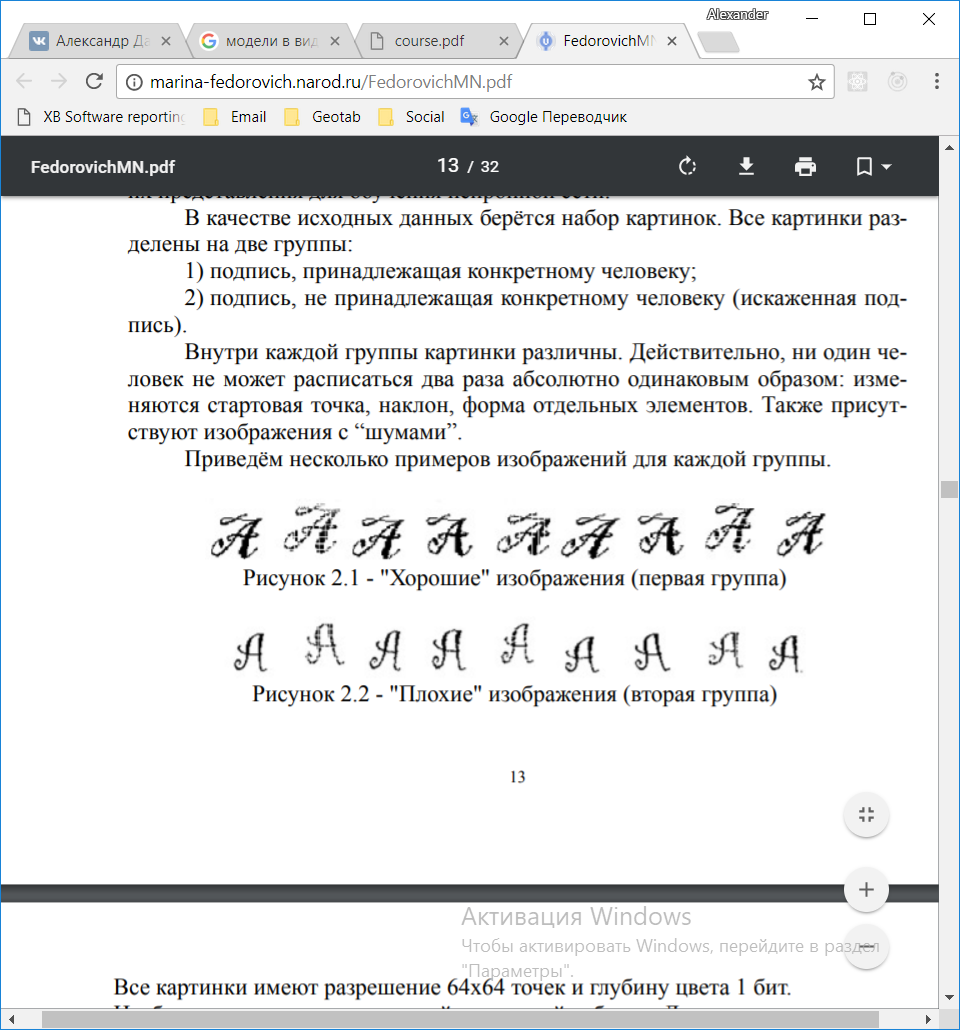


Рисунок 2.2 – «Плохие» изображения

Все картинки имеют разрешение 64х64 точек и глубину цвета 1 бит.

Изображения представим в одной двумерной таблице. Для этого применим простую развёртку: двумерный массив значений преобразуется в одномерный. Каждому изображению соответствует одно наблюдение – строка в таблице данных; элементы строки – значения соответствующих пикселей в исходном изображении.

В итоге имеем таблицу с 4096 столбцами (4096 = 64\*64) и 2275 наблюдениями. Была добавлена переменная Type, принимающая значение 1, если подпись правильная, 0 – если подпись неправильная.

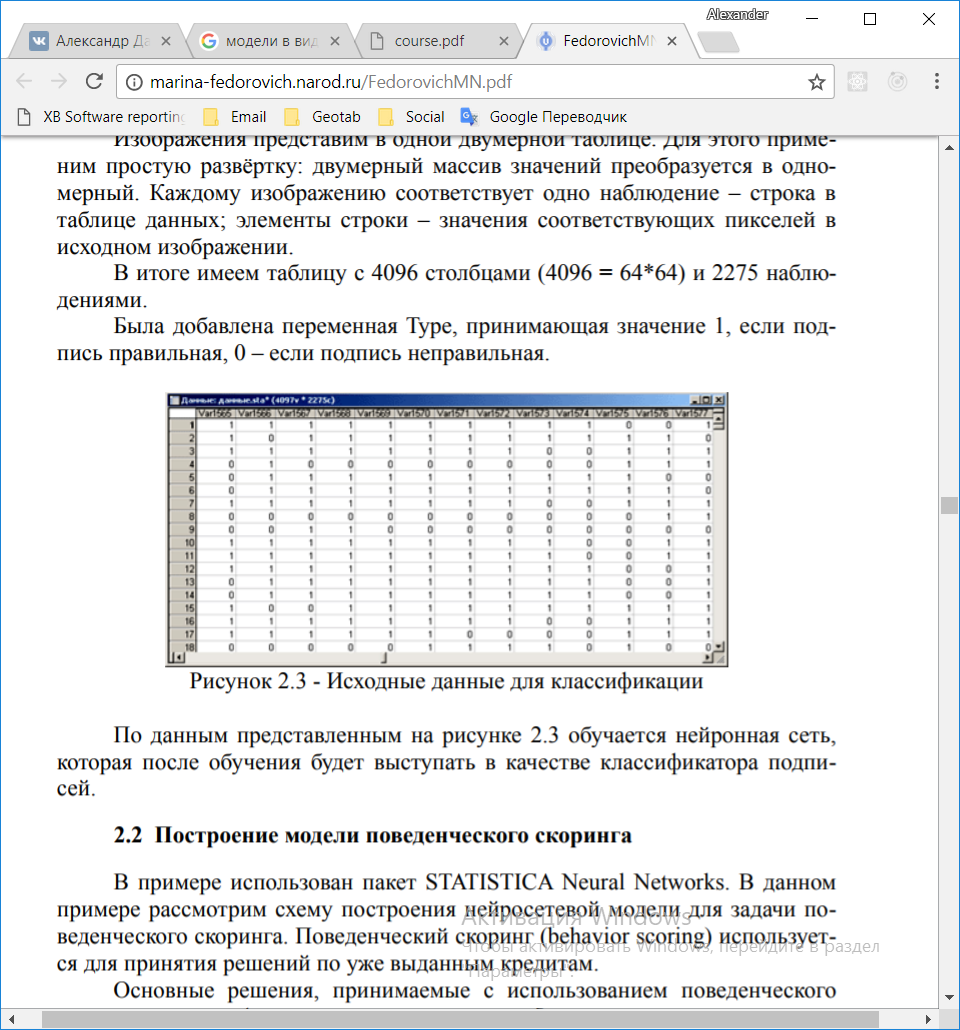


Рисунок 2.3 – Исходные данные для классификации

По данным представленным на рисунке 2.3 обучается нейронная сеть, которая после обучения будет выступать в качестве классификатора подписей.

2.2 Другие области использования искусственных нейронных сетей

Прогнозирование финансовых временных рядов (компания LBS Capital Management объявила о значительных успехах в финансовых операциях, достигнутых за счет прогнозирования цен акций с помощью многослойных персептронов, которые они будут рассмотрены ниже).

Повышение эффективности процесса добычи полезных ископаемых (выделение значимых факторов, влияющих на показатели эффективности добычи).

Оптическое распознавание символов, включая распознавание подписи (например, система идентификации подписи, учитывающая не только окончательный ее рисунок, но и скорость движения пера на различных участках, что значительно затрудняет подделку чужой подписи).

Обработка изображений (например, система сканирует видеоизображения станций метро и определяет, насколько станция заполнена людьми, причем работа системы не зависит от условий освещенности и движения поездов).

Медицинская диагностика (например, прогнозирование эпилептических припадков, определение размеров опухоли простаты).

Синтез речи (знаменитая экспериментальная система Nettalk, способная произносить фонемы из написанного текста).

Лингвистический анализ (пример: сеть с неконтролируемым обучением используется для идентификации ключевых фраз и слов в языках туземцев Южной Америки).

1. РАБОТА С НЕЙРОННЫМИ СЕТЯМИ
   1. Нейронная сеть Хопфилда

Сеть Хопфилда – однослойная, симметричная, нелинейная, автоассоциативная нейронная сеть, которая запоминает бинарные / биполярные образы. Сеть характеризуется наличием обратных связей. Топология сети Хопфилда показана на рисунке 3.1. Информация с выхода каждого нейрона поступает на вход всех остальных нейронов. Образы для данной модификации сети Хопфилда кодируются биполярным вектором, состоящим из *1* и *–1*.



Рисунок 3.1 – Топология сети Хопфилда

Обучение сети осуществляется в соответствии с соотношением:

, для , (1.1)

где *wij* – вес связи от *i*-го нейрона к *j*-му;

*n* – количество нейронов в сети;

*m* – количество образов, используемых для обучения сети;

*aik* – *i*-й элемент *k*-го образа из обучающей выборки.

Матрица весовых коэффициентов:

. (1.2)

В качестве матрицы весовых коэффициентов Хопфилд использовал симметричную матрицу (*wij=wji*) с нулевой главной диагональю (*wii=0*). Последнее условие соответствует отсутствию обратной связи нейронного элемента на себя. В качестве функции активации нейронных элементов может использоваться как пороговая, так и непрерывная функции, например сигмоидная или гиперболический тангенс.

Будем рассматривать нейронную сеть Хопфилда с дискретным временем. Тогда при использовании пороговой функции активации она называется нейронной сетью с дискретным состоянием и временем. Нейронная сеть с непрерывной функцией активации называется нейронной сетью с непрерывным состоянием и дискретным временем. При использовании непрерывного времени модель Хопфилда называется непрерывной.

Для описания функционирования таких сетей Хопфилд использовал аппарат статистической физики. При этом каждый нейрон имеет два состояния активности *(1, –1)*, которые аналогичны значениям спина некоторой частицы. Весовой коэффициент *wji* можно интерпретировать как вклад поля *j*-частицы в величину потенциала *i*-частицы. Хопфилд показал, что поведение такой сети аналогично поведению лизингового спинового стекла. При этом он ввел понятие вычислительной энергии, которую можно интерпретировать в виде ландшафта с долинами и впадинами. Структура соединений сети определяет очертания ландшафта. Нейронная сеть выполняет вычисления, следуя по пути, уменьшающему вычислительную энергию сети. Это происходит до тех пор, пока путь не приведет на дно впадины. Данный процесс аналогичен скатыванию капли жидкости по склону, когда она минимизирует свою потенциальную энергию в поле тяготения. Впадины и долины в сети Хопфилда соответствуют наборам информации, которую хранит сеть. Если процесс начинается с приближенной или неполной информации, то он следует по пути, который ведет к ближайшей впадине. Это соответствует операции ассоциативного распознавания.

Матрица весов является диагонально симметричной, причем все диагональные элементы равны *0*.

Нейронная сеть Хопфилда может функционировать синхронно и асинхронно. Для воспроизведения используется соотношение:

, (1.3)

где *aj(t)* – выход *j*-го нейрона в момент времени *t*, а *f* – бинарная / биполярная функция активации;

 (1.4)

При работе в синхронном режиме на один такт работы сети все нейроны одновременно меняют состояние по формуле (1.4). В случае асинхронной работы состояние меняет только один случайный нейрон. Итерации продолжаются до тех пор, пока сеть не придет в стабильное состояние.

Во время воспроизведения исходным вектором *a(0)* является некоторый тестовый образ, не совпадающий с образами из обучающей выборки. В процессе функционирования по формуле (1.4) сеть должна прийти в состояние, соответствующее образу из обучающей выборки, наиболее похожему на тестовый.

Максимальное количество образов, которое можно запомнить в матрице *W*, не превышает:

, (1.5)

где *n* – количество нейронов, что следует отнести к недостаткам этой сети.

На рисунке 3.2 представлены результаты распознавания букв при различном искажении исходного изображения. Нужно обратить внимание на то, что в при зашумлении между 40-60% зашумления сеть перестает идентифицировать исходный образ, а начиная с 60% начинает идентифицировать зеркальное отражение исходного образа.

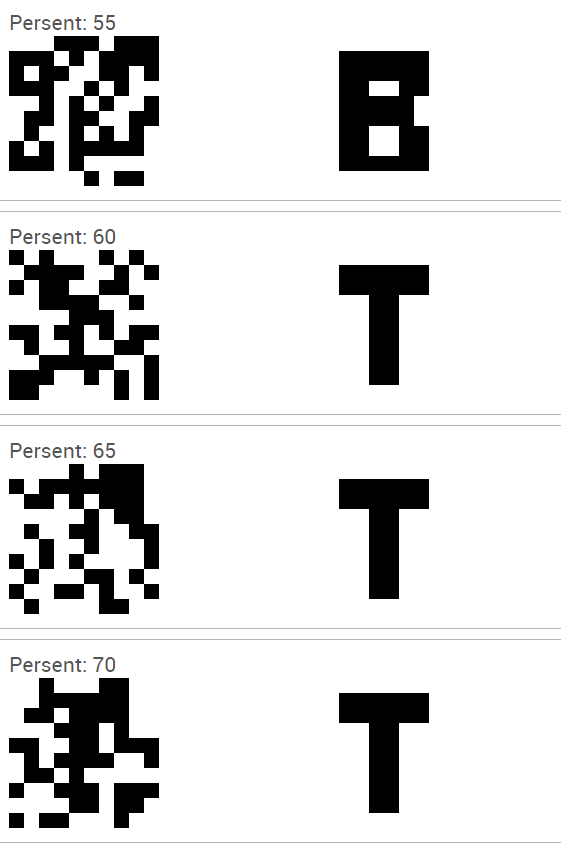
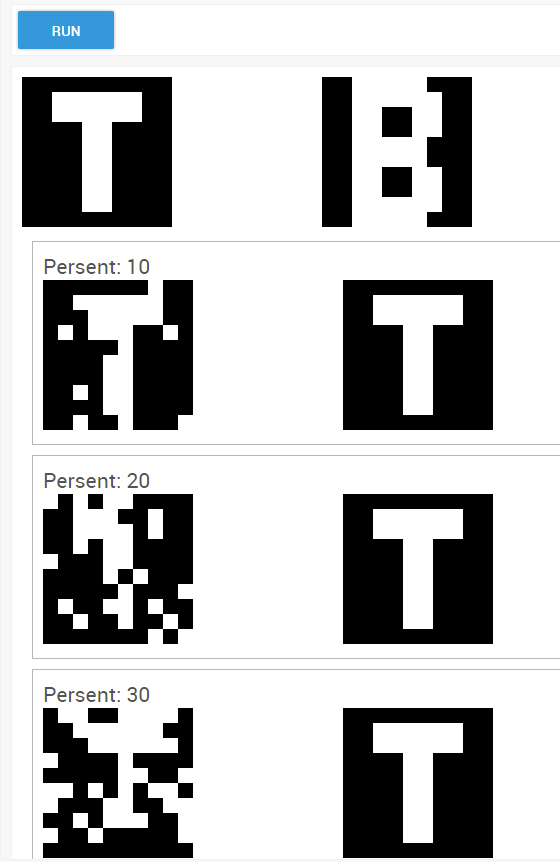


Рисунок 3.2 – Распознавание образов сетью Хопфилда при различном зашумлении исходного образа

* 1. Многослойный персептрон

Многослойный персептрон является сетью с прямым распространением сигнала (без обратных связей), обучаемой с учителем. Такая сеть способна аппроксимировать любую непрерывную функцию или границу между классами со сколь угодно высокой точностью. Для этого достаточно одного скрытого слоя нейронов с сигмоидной функцией активации (2) – (4), то есть многослойный персептрон обычно состоит из 3 слоев: первого распределительного, второго скрытого и третьего выходного (рисунок 3.3).



Рисунок 3.3 – Многослойный персептрон

Такая сеть имеет n входов и n нейронов распределительного слоя, h нейронов скрытого слоя и m выходных нейронов. Используются две матрицы весов: скрытого слоя v размером n×h и выходного слоя w размером h×m. Кроме этого, с каждым слоем нейронов связан массив порогов: Q – для скрытого слоя, T – для выходного. Эти данные представляют собой знания сети, настраиваемые в процессе обучения и определяющие ее поведение. Персептрон функционирует по следующим формулам:

, (2.1)

. (2.2)

В качестве функции активации используется одна из функций (2) – (4). Вид функции определяет диапазон чисел, в котором работает сеть. В дальнейшем будет использоваться сигмоидная функция (2), имеющая область значений от 0 до 1.

Обучение с учителем ставит перед сетью задачу обобщить p примеров, заданных парами векторов (xr, yr), . Вектор  в случае задачи классификации задает входной образ (вектор признаков), а вектор , задающий эталонный выход, должен кодировать номер класса. При этом есть множество вариантов кодирования. Оптимальным представляется кодирование, когда номер класса определяется позицией единичной компоненты в векторе yr, а все остальные компоненты равны 0. Каждый выходной нейрон соответствует одному классу. Такой способ позволяет при классификации определять вероятность каждого класса по величине на выходе соответствующего нейрона (чем ближе к единице, тем вероятность больше).

Обучение персептрона проводится с помощью алгоритма обратного распространения ошибки, который минимизирует среднеквадратичную ошибку нейронной сети. Для этого с целью настройки синаптических связей используется метод градиентного спуска в пространстве весовых коэффициентов и порогов нейронной сети. Рассмотрим алгоритм обратного распространения ошибки.

1. Происходит начальная инициализация знаний сети. Простейший вариант такой инициализации – присвоить всем весам и порогам случайные значения из диапазона [-1,1].

2. Для каждой пары векторов (xr, yr) выполняется следующее:

2.1. Для входного вектора рассчитываются выходы нейронов скрытого слоя и выходы сети по формулам (2.1), (2.2).

2.2. Происходит коррекция знаний сети, при этом главное значение имеет отклонение реально полученного выхода сети y от идеального вектора yr. Согласно методу градиентного спуска, изменение весовых коэффициентов и порогов нейронной сети происходит по следующим формулам:

, (2.3)

, (2.4)

где E - среднеквадратичная ошибка нейронной сети для одного образа, а α – параметр, определяющий скорость обучения.

Формулы записаны в терминах выходного слоя, аналогично выглядят формулы для скрытого слоя.

Среднеквадратичная ошибка сети вычисляется как

. (2.5)

Ошибка k-го нейрона выходного слоя определяется как:

. (2.6)

Выразим производные из формул (2.3), (2.4) через легко вычисляемые величины. Определим взвешенную сумму, аргумент функции активации как:

. (2.7)

Из соотношения  можно представить как:

, (2.8)

где  – ошибка k-го нейрона;  – производная функции активации;  – значение j-го нейрона предыдущего слоя. Получаем

. (2.9)

Аналогично (2.8), с учетом того, что , получаем

. (2.10)

Веса и пороги скрытого слоя также корректируются по формулам, аналогичным (2.3), (2.4), с учетом (2.9) и (2.10). При этом главной трудностью является определение ошибки нейрона скрытого слоя. Эту ошибку явно определить по формуле, аналогичной (2.6), невозможно, однако существует возможность рассчитать ее через ошибки нейронов выходного слоя (отсюда произошло название алгоритма обратного распространения ошибки):

. (2.11)

Производные от функций активации тоже легко рассчитываются, например, для сигмоидной функции получаем

. (2.12)

Аналогично

;

.

Таким образом, можно записать окончательные выражения (2.3), (2.4) для двух слоев, использующих сигмоидную функцию:

, (2.13)

, (2.14)

, (2.15)

. (2.16)

В эти формулы вводится дополнительный параметр β – скорость обучения скрытого слоя, который может отличаться от аналогичного параметра для выходного слоя. Рекомендуется изменять скорости обучения обратно пропорционально количеству шагов алгоритма обучения, однако это не всегда оправдывает себя на практике.

3. После того как коррекция знаний произведена для каждой пары векторов, можно оценить степень успешности обучения сети для определения момента завершения алгоритма. Для этого можно использовать в качестве критерия максимальную по модулю ошибку на выходе dk, полученную на шаге 2. Условием прекращения обучения в этом случае будет:

, (2.17)

где D – достаточно маленькая константа – величина максимальной ошибки, которую требуется достичь в процессе обучения. Если условие (2.17) не выполняется, то шаг 2 повторяется.

Способность персептрона-классификатора разделять образы в пространстве признаков прежде всего зависит от его скрытого слоя. Именно на этот слой возлагается задача сделать множество классов линейно разделимым для успешной работы выходного слоя. Очевидно, что чем больше нейронов в скрытом слое, тем большее количество примеров этот слой может разделять. Кроме этого, увеличение числа признаков входных образов также способствует успешному их разделению в пространстве признаков. Однако увеличение этих параметров приводит к росту ошибок сети и времени обучения. Увеличение размерности входов n приводит к росту ошибки аппроксимации сети, возникающей из-за обобщения данных. Увеличение числа нейронов h скрытого слоя приводит к росту ошибки, связанной со сложностью модели. Персептрону легче провести функцию через эталонные точки, однако при этом обобщающая способность сети ухудшается. Он хуже предсказывает поведение аппроксимируемой функции на образах, не входящих в обучающую выборку. Такое состояние сети называется переобучением.

Оптимальное соотношение между этими параметрами оценивают как

. (2.18)

Эксперименты показывают, что обучение максимально успешно проходит на множестве классов, хорошо (желательно линейно) разделенных в пространстве признаков. Это достигается удачным подбором информативных признаков. Если классы в пространстве признаков хорошо кластеризуются, т.е. образы каждого класса составляют компактную группу, достаточно удаленную от других групп, то есть возможность уменьшить размер обучающей выборки p (используются только центры кластеров) и затем уменьшить число нейронов h. Это приводит к ускорению обучения и улучшает работу классификатора.

Другой проблемой является то, что алгоритм градиентного спуска не гарантирует нахождение глобального минимума среднеквадратичной ошибки сети (2.5), а гарантируется определение только локального минимума. Проблемы, возникающие в процессе градиентного спуска, можно проанализировать на примере функции ошибки, схематически изображенной на рис. 2.2.



Рисунок 3.4 – Схематическая функция ошибки сети

На рисунке 3.4 показаны четыре критические точки, производная функции ошибки в которых близка к нулю. Точка A соответствует локальному минимуму. Признаком достижения локального минимума в процессе обучения является полное прекращение уменьшения ошибки (2.17). В этом случае может помочь повторное обучение с другим начальным распределением знаний (тут может помочь случайная инициализация). Точка B – локальный максимум. В случае попадания в окрестность такой точки скорость резко падает, затем снова быстро растет. Не существует способа предсказать, в какую сторону лучше двигаться из точки с производной, близкой к нулю. Точка C – точка перегиба, характеризуется длительным уменьшением скорости. Точка D – глобальный минимум – цель алгоритма.

Существуют многочисленные способы оптимизации метода градиентного спуска, призванные улучшить поведение алгоритма в подобных критических точках. Эффективной модификацией является введение момента, накапливающего влияние градиента на веса со временем. Тогда величина Δw (2.3) в момент времени t будет вычисляться как

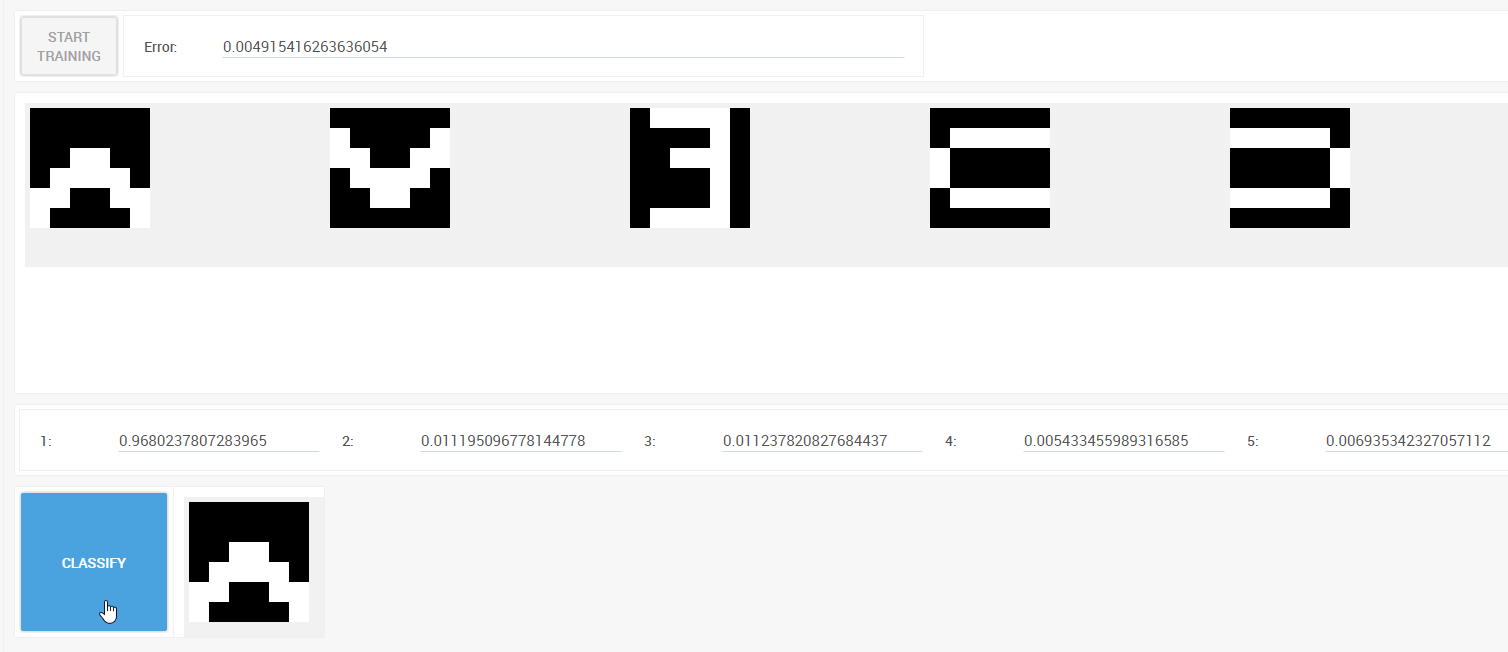
, (2.19)

где μ – параметр, определяющий величину влияния момента.

С использованием (2.19) скорость изменения весов возрастает на участках с постоянным знаком производной. В окрестностях минимума скорость резко падает за счет колебания знака.

Достоинства алгоритма – большая скорость в точках перегиба, возможность по инерции преодолевать небольшие локальные минимумы. Недостатки – еще один параметр, величину которого следует подбирать и настраивать.

Этот и другие алгоритмы оптимизации обучения персептрона позволяют улучшить работу сети в условиях плохой сходимости. Однако они усложняют процесс обучения, не гарантируя в то же время полного успеха во всех случаях. Успех обучения классификатора зависит от самого алгоритма обучения и качества обучающей выборки.



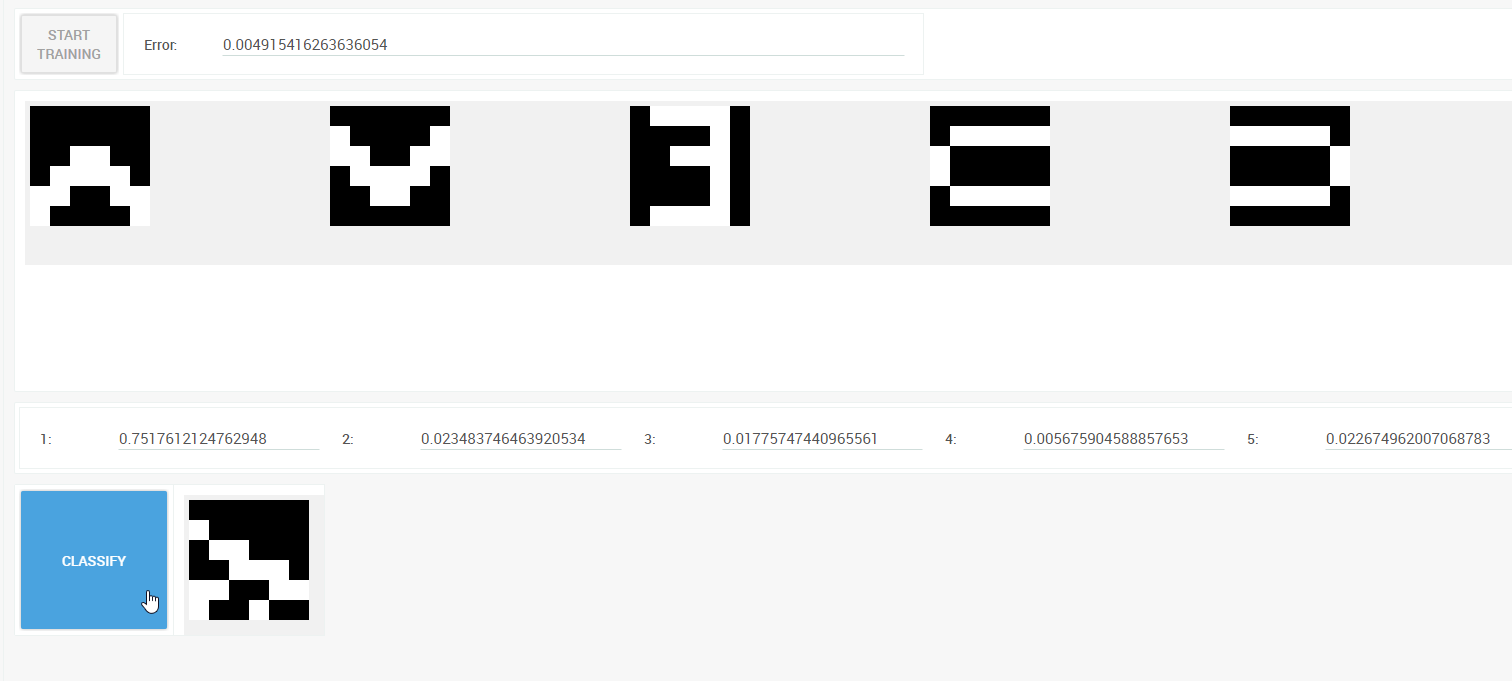


Рисунок 3.4 – Распознавание образов многослойный персептроном с различной степенью шума

ЗАКЛЮЧЕНИЕ

Развитие нейронных сетей вызвало немало энтузиазма и критики. Некоторые сравнительные исследования оказались оптимистичными, другие – пессимистичными. Для многих задач, таких как распознавание образов, пока не создано доминирующих подходов. Нужно пытаться понять возможности, предпосылки и область применения различных подходов и максимально использовать их дополнительные преимущества для дальнейшего развития интеллектуальных систем.

Множество надежд в отношении нейронных сетей сегодня связывают именно с аппаратными реализациями, но пока время их массового выхода на рынок, видимо, еще не пришло. Они или выпускаются в составе специализированных устройств, или достаточно дороги, а зачастую и то и другое. На их разработку тратится значительное время, за которое программные реализации на самых последних компьютерах оказываются лишь на порядок менее производительными, что делает использование нейропроцессоров нерентабельным. Но все это только вопрос времени – нейронным сетям предстоит пройти тот же путь, по которому еще совсем недавно развивались компьютеры, увеличивая свои возможности и производительность, захватывая новые сферы применения по мере возникновения новых задач и развития технической основы для их разработки.

Сегодня нейронные сети используются для работы в относительно узких областях, и неизвестно, доверят ли им когда-нибудь решение вопросов, которые требуют понимания социального контекста. Между тем нейронные сети уверенно продолжают проникать в нашу жизнь, и примеров тому немало.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

[1] Осовский, С. Нейронные сети для обработки информации / С. Осовский. – М.: Финансы и статистика, 2002. – 344 с.

[2] Р.Х. Садыхов, М.М. Лукашевич, Лабораторный практикум по дисциплине «Цифровая обработка сигналов и изображений»: В 2 ч. Ч. 2 / Р.Х. Садыхов, М.М. Лукашевич – Мн. : БГУИР, 2006. – 32 с.

[2] Хайкин, С. Нейронные сети: полный курс, 2-е издание / С. Хайкин. – М.: Издательский дом «Вильямс», 2006. – 1104 с.

[3] Каллан, Р. Основные концепции нейронных сетей / Р. Каллан. – М.: Издательский дом «Вильямс», 2001. – 287 с.

[4] Marquardt, D. An algorithm for least squares estimation of nonlinear parameters / D. Marquardt // SIAM. – 1963. – C. 431- 442.

[5] Bishop, C. M. Neural network for pattern recognition / C.M. Bishop. – Oxford: Oxford University Press, 2005. – 482 с.