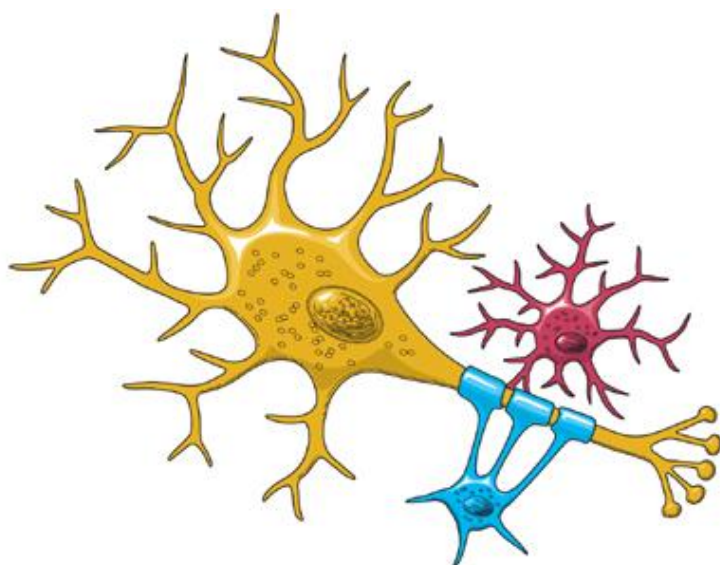


Кафедра ИУ5



Научно-исследовательская работа
по курсу «Интеллектуальные системы»
«Особенности обучения перцептронных сетей»



Выполнили: **Алмаев Владимир**
Петрунин Сергей

Группа: **ИУ5-92**

Дата: ____.

Подписи: _____

Аннотация

Искусственные нейронные сети индуцированы биологией, так как они состоят из элементов, функциональные возможности которых аналогичны большинству элементарных функций биологического нейрона. Эти элементы затем организуются по способу, который может соответствовать (или не соответствовать) анатомии мозга. Несмотря на такое поверхностное сходство, искусственные нейронные сети демонстрируют удивительное число свойств, присущих мозгу. Например, они обучаются на основе опыта, обобщают предыдущие прецеденты на новые случаи и извлекают существенные свойства из поступающей информации, содержащей излишние данные.

Основные свойства нейронных сетей:

- **Обучение** - искусственные нейронные сети могут менять свое поведение в зависимости от внешней среды.
- **Обобщение** - отклик сети после обучения может быть до некоторой степени нечувствителен к небольшим изменениям входных сигналов.
- **Абстрагирование** - некоторые из искусственных нейронных сетей обладают способностью извлекать сущность из входных сигналов.

Несмотря на такое функциональное сходство, даже самый оптимистичный их защитник не предположит, что в скором будущем искусственные нейронные сети будут дублировать функции человеческого мозга. Реальный «интеллект», демонстрируемый самыми сложными нейронными сетями, находится ниже уровня дождевого червя, и энтузиазм должен быть умерен в соответствии с современными реалиями. Однако равным образом было бы неверным игнорировать удивительное сходство в функционировании некоторых нейронных сетей с человеческим мозгом. Эти возможности, как бы они ни были ограничены сегодня, наводят на мысль, что глубокое проникновение в человеческий интеллект, а также множество революционных приложений, могут быть не за горами.

Потенциальными приложениями нейронных сетей являются те, где человеческий интеллект малоэффективен, а обычные вычисления трудоемки или неадекватны. Этот класс приложений во всяком случае не меньше класса, обслуживаемого обычными вычислениями, и можно предполагать, что искусственные нейронные сети займут свое место наряду с обычными вычислениями в качестве дополнения такого же объема и важности. [1]

1. Теоретические основы искусственных нейронных сетей

1.1. Биологический прототип

Развитие искусственных нейронных сетей вдохновляется **биологией**. Т.е., рассматривая сетевые конфигурации и алгоритмы, исследователи мыслят их в терминах организации мозговой деятельности. Но на этом аналогия может и закончиться. Наши знания о работе мозга столь ограничены, что мало бы нашлось руководящих ориентиров для тех, кто стал бы ему подражать. Поэтому разработчикам сетей приходится выходить за пределы современных биологических знаний в поисках структур, способных выполнять полезные функции. Во многих случаях это приводит к необходимости отказа от биологического правдоподобия, мозг становится просто метафорой, и создаются сети, невозможные в живой материи или требующие неправдоподобно больших допущений об анатомии и функционировании мозга.

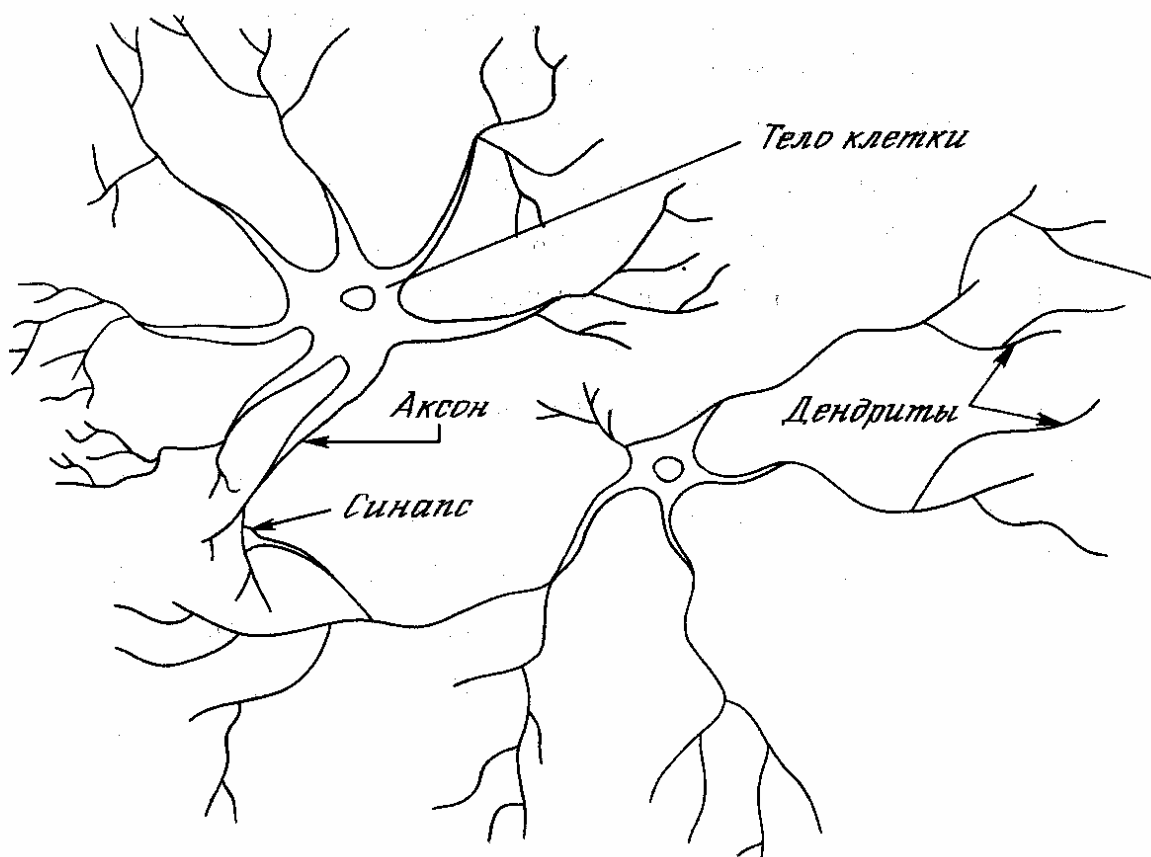


Рис.1.1. Биологический нейрон.

Несмотря на то, что связь с биологией слаба и зачастую несущественна, искусственные нейронные сети продолжают сравниваться с мозгом. Их функционирование часто напоминает человеческое познание, поэтому трудно избежать этой аналогии. [1]

На рис. 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.

1.2. Искусственный нейрон

Искусственный нейрон имитирует в первом приближении свойства биологического нейрона. На вход искусственного нейрона поступает некоторое множество сигналов, каждый из которых является выходом другого нейрона. Каждый **вход** умножается на соответствующий **вес**, аналогичный синаптической силе, и все произведения суммируются, определяя уровень активации нейрона.

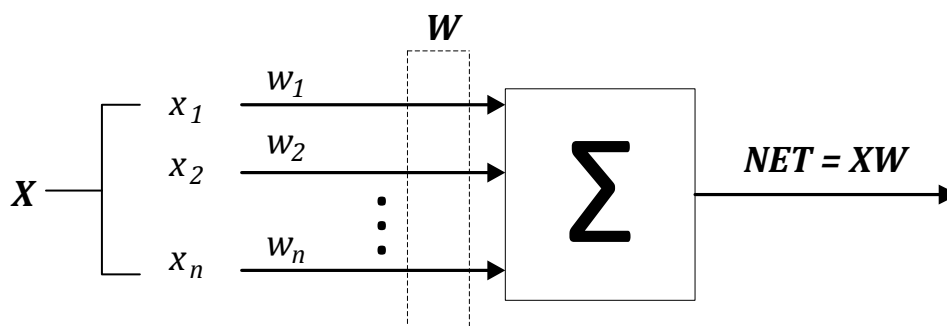


Рис.1.2. Искусственный нейрон.

На рис. 1.2 представлена модель искусственного нейрона. Множество входных сигналов, обозначенных x_1, x_2, \dots, x_n , поступает на искусственный нейрон. Эти входные сигналы, в совокупности обозначаемые вектором X , соответствуют сигналам, приходящим в синапсы биологического нейрона. Каждый сигнал умножается на соответствующий вес w_1, w_2, \dots, w_n , и поступает на суммирующий блок, обозначенный Σ . Каждый вес соответствует «силе» одной биологической синаптической связи. (Множество весов в совокупности обозначается вектором W .) Суммирующий блок, соответствующий телу биологического элемента, складывает взвешенные входы алгебраически, создавая выход, который мы будем называть NET. В векторных обозначениях это может быть компактно записано следующим образом:

$$NET = XW = \sum_i x_i w_i \quad (1)$$

1.3. Активационные функции

Сигнал NET далее, как правило, преобразуется **активационной функцией F** и дает выходной нейронный сигнал OUT. Активационная функция может быть:

- обычной линейной функцией вида

$$OUT = K(NET), \quad (2)$$

где K – постоянная, пороговой функции;

- пороговой функцией:

$$OUT = \begin{cases} 1, & \text{если } NET > T \\ 0 & \text{в остальных случаях} \end{cases} \quad (3)$$

где T – некоторая постоянная пороговая величина, или же функцией, более точно моделирующей нелинейную передаточную характеристику биологического нейрона и представляющей нейронной сети большие возможности.

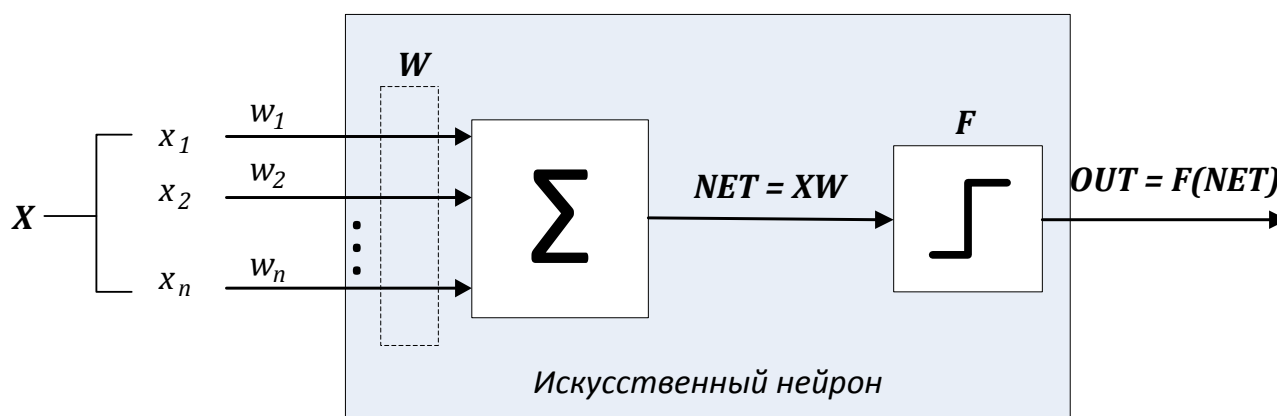


Рис. 1.3. Искусственный нейрон с активационной функцией.

На рис. 1.3 блок, обозначенный F , является активационной функцией. Он принимает сигнал NET и выдает сигнал OUT.

1.4. Обучение перцептрона

Рассмотрим метод обучения перцептрона, основанный на коррекции ошибок (другое название метода – обучение с учителем).

Пусть на вход перцептрона k поступает **вектор сигнала** $x_k(n)$. Под n подразумевается дискретное время, или, более конкретно – номер шага итеративного процесса настройки синаптических весов перцептрона k . **Выходной сигнал** нейрона k обозначается $y_k(n)$. Он будет сравниваться с **желаемым (целевым) выходом**, обозначенным $d_k(n)$. В результате получим **сигнал ошибки** $e_k(n)$:

$$e_k(n) = d_k(n) - y_k(n) \quad (4)$$

Сигнал ошибки инициализирует **механизм управления**, цель которого заключается в применении последовательности корректировок к синаптическим весам перцептрона k . Эти изменения нацелены на пошаговое приближение выходного сигнала $y_k(n)$ к желаемому $d_k(n)$. Эта цель достигается за счет минимизации **функции стоимости** E , определяемой в терминах сигнала ошибки следующим образом:

$$E(n) = \frac{1}{2} e_k^2(n), \quad (5)$$

где $E(n)$ – **текущее значение энергии ошибки**. Пошаговая корректировка синаптических весов перцептрона k продолжается до тех пор, пока система не достигнет устойчивого состояния (т.е. такого, при котором синаптические веса практически стабилизируются). В этой точке процесс обучения заканчивается.

Минимизация функции стоимости выполняется по так называемому **дельта-правилу**, или **правилу Видроу-Хоффа**. Обозначим $w_{kj}(n)$ текущее значение синаптического веса w_{kj} перцептрона k , соответствующего элементу $x_j(n)$, на шаге дискретизации n . В соответствии с дельта-правилом изменение $\Delta w_{kj}(n)$, применяемое к синаптическому весу w_{kj} на этом шаге дискретизации, задается выражением

$$\Delta w_{kj}(n) = \eta e_k(n) x_j(n), \quad (6)$$

где η – некоторая положительная константа, определяющая **скорость обучения**.

Вербально дельта правило можно определить следующим образом.

Дельта правило (правило Видроу-Хоффа): корректировка, применяемая к синаптическому весу нейрона, пропорциональна произведению сигнала ошибки на входной сигнал, его вызвавший.

Вычислив величину изменения синаптического веса $\Delta w_{kj}(n)$, можно определить его новое значение для следующего шага дискретизации:

$$w_{kj}(n + 1) = w_{kj}(n) + \Delta w_{kj}(n). \quad (7)$$

Решающим фактором, влияющим на обучение перцептрона в рамках этой модели, является подбор коэффициента скорости обучения η . Для обеспечения устойчивости и сходимости итеративного процесса обучения требуется тщательный подбор этого параметра. Выбор параметра скорости обучения влияет также на точность обучения. Другими словами, **параметр скорости обучения η играет ключевую роль в обеспечении производительности процесса обучения на практике.** [2]

1.5. Резюме

Рассмотренная простая модель искусственного нейрона игнорирует многие свойства своего биологического двойника. Например, она не принимает во внимание задержки во времени, которые воздействуют на динамику системы. Входные сигналы сразу же порождают выходной сигнал. И, что более важно, она не учитывает воздействий функции частотной модуляции или синхронизирующей функции биологического нейрона, которые ряд исследователей считают решающими.

Несмотря на эти ограничения, сети, построенные из этих нейронов, обнаруживают свойства, сильно напоминающие биологическую систему. Только время и исследования смогут ответить на вопрос, являются ли подобные совпадения случайными или следствием того, что в модели верно схвачены важнейшие черты биологического нейрона.

2. Исследовательская часть

Ниже представлены результаты исследования процесса обучения персептрона при различных методиках обучения, начальных значениях и критериях. Были рассмотрены следующие методики:

- M1. Динамическое изменение скорости обучения.
- M2. Постоянная скорость обучения.

Рассмотрено влияние начальных значений на скорость обучения:

- A1. Варьирование начальных значений синаптических весов персептрона.
- A2. Изменение размера обучающей выборки

Также даны рекомендации по выбору критерия окончания обучения.

2.1. Проведение эксперимента

Необходимо установить каким образом будут влиять такие параметры, как скорость обучения, начальные веса входов персептрона, критерий окончания обучения влиять на обучаемость персептрона, а также выдать рекомендации по заданию этих параметров для достижения наилучшего результата. В рамках эксперимента рассмотрим персептрон, распознающий изображение размером 3 на 3, т.е. персептрон имеет 9 информационных входов (см. рис. 2.1)

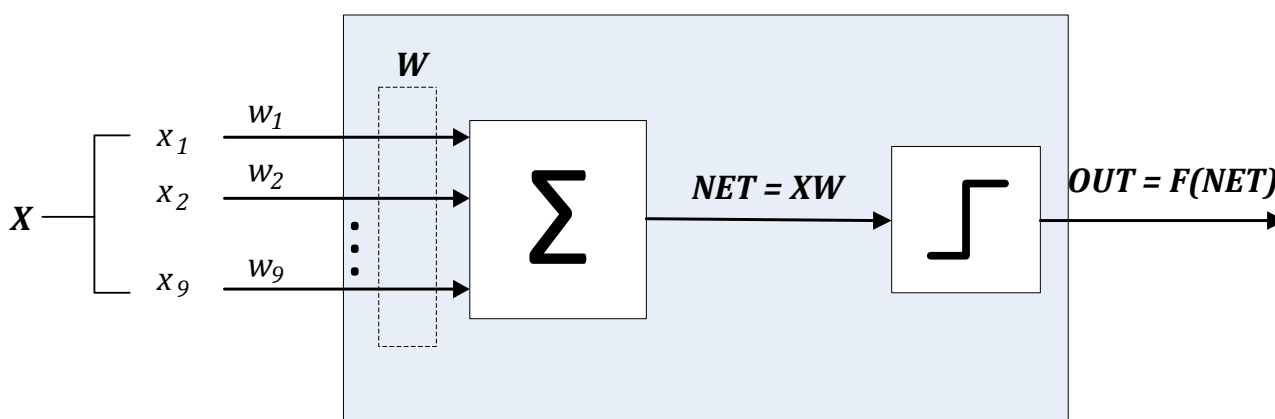


Рис. 2.1. Модель персептрона с 9 информационными входами.

2.1.1. Исследование скорости обучения

Рассмотрим метод с постоянной скоростью обучения. Зададим скорость обучения, равную 0.1, и начальные веса, равные 0. Зададим правильное изображение, сгенерируем обучающую выборку, проведём обучение на ней. По умолчанию у нас один цикл обучения на выборке из 10 элементов.

Метод с постоянной скоростью обучения. Исходные данные

Скорость обучения	0.1
Начальные веса входов	0
Количество циклов обучения	1
Размер обучающей выборки	10

Были получены следующие экспериментальные данные.

Значение на выходе перцептрона для правильных выборок

Изображения из одного пикселя	NET = 0.076
Изображения из двух пикселей	NET = 0.111
Изображения из трех пикселей	NET = 0.119
Изображения из четырех пикселей	NET = 0.123
Изображения из пяти пикселей	NET = 0.134
Изображения из девяти пикселей	NET = 0.857

Значение на выходе перцептрона для неправильных выборок (наихудший вариант)

Изображения из одного пикселя	NET = 0.059
Изображения из двух пикселей	NET = 0.068
Изображения из трех пикселей	NET = 0.080
Изображения из четырех пикселей	NET = 0.080
Изображения из пяти пикселей	NET = 0.108
Изображения из девяти пикселей	NET = 0.799

Разница между значениями на выходе перцептрона для правильного изображения и для неправильного изображения.

Изображения из одного пикселя	$\Delta = 0.017$
Изображения из двух пикселей	$\Delta = 0.043$
Изображения из трех пикселей	$\Delta = 0.039$
Изображения из четырех пикселей	$\Delta = 0.043$
Изображения из пяти пикселей	$\Delta = 0.026$
Изображения из девяти пикселей	$\Delta = 0.058$

$$\Delta_{\text{CP}} = 0.038$$

Результат: перцептрон в состоянии распознать изображение, однако грань между правильным и неправильным изображением минимальна, поэтому данный метод обучения не рекомендуется к использованию.

Введём понятие «**динамическое изменение скорости обучения**», подразумевая под ним смену скоростей обучения в зависимости от того, является ли выборка правильной или нет.

В качестве иллюстрации рассматриваемой методики приведем результаты следующего эксперимента: скорость обучения для правильной выборки равна 0.4, для неправильной – 0.01, значения начальных весов равны -0.2.

Метод с динамическим изменением скорости обучения. Исходные данные

Скорость обучения для правильной выборки	0.4
Скорость обучения для неправильной выборки	0.01
Начальные веса входов	-0.2
Количество циклов обучения	1
Размер обучающей выборки	10

Были получены следующие экспериментальные данные.

Значение на выходе перцептрона для правильных выборок

Изображения из одного пикселя	NET = 0.4
Изображения из двух пикселей	NET = 0.858
Изображения из трех пикселей	NET = 0.954

Значение на выходе перцептрона для неправильных выборок

Изображения из одного пикселя	NET = 0.125
Изображения из двух пикселей	NET = 0.524
Изображения из трех пикселей	NET = 0.611

Разница между значениями на выходе перцептрона для правильного изображения и для неправильного изображения

Изображения из одного пикселя	$\Delta = 0.275$
Изображения из двух пикселей	$\Delta = 0.334$
Изображения из трех пикселей	$\Delta = 0.343$

$$\Delta_{\text{CP}} = 0.318$$

Рассмотренный метод позволяет увеличить зазор между правильными и неправильными изображениями при их распознавании. Это даёт большой простор для выбора оптимального порогового значения.

Результат: персептрон в состоянии распознать изображение, причем грань между правильным и неправильным изображением возросла на порядок по сравнению с методом статической скорости обучения. Данный метод обучения рекомендуется к использованию.

2.1.2. Исследование начальных весов

Рассмотрим следующие методы обучения:

- метод с нулевыми начальными весами;
- метод с положительными начальными весами;
- метод с отрицательными начальными весами.

Результаты **метода с нулевыми начальными весами** идентичны результатам метода с постоянной скоростью обучения (см. пункт 2.1.1).

Рассмотрим **метод с положительными начальными весами**. Зададим скорость обучения, равную 0.1, и начальные веса, равные 0.2. Зададим правильное изображение, сгенерируем обучающую выборку, проведём обучение на ней. По умолчанию у нас один цикл обучения на выборке из 10 элементов.

Метод с положительными начальными весами. Исходные данные

Скорость обучения	0.1
Начальные веса входов	0.2
Количество циклов обучения	1
Размер обучающей выборки	10

Были получены следующие экспериментальные данные.

Значение на выходе персептрона для правильных выборок

Изображения из одного пикселя	NET = 0.256
Изображения из двух пикселей	NET = 0.311
Изображения из трех пикселей	NET = 0.319
Изображения из четырех пикселей	NET = 0.323
Изображения из пяти пикселей	NET = 0.334
Изображения из девяти пикселей	NET = 1.0

Значение на выходе персептрона для неправильных выборок (наихудший вариант)

Изображения из одного пикселя	NET = 0.239
Изображения из двух пикселей	NET = 0.262
Изображения из трех пикселей	NET = 0.290
Изображения из четырех пикселей	NET = 0.292
Изображения из пяти пикселей	NET = 0.208
Изображения из девяти пикселей	NET = 0.999

Разница между значениями на выходе персептрона для правильного изображения и для неправильного изображения

Изображения из одного пикселя	$\Delta = 0.017$
Изображения из двух пикселей	$\Delta = 0.049$
Изображения из трех пикселей	$\Delta = 0.029$
Изображения из четырех пикселей	$\Delta = 0.033$
Изображения из пяти пикселей	$\Delta = 0.027$
Изображения из девяти пикселей	$\Delta = 0.001$

$$\Delta_{\text{CP}} = 0.026$$

Таким образом, граница весьма невелика, и при более сложных задачах распознавания результат будет не всегда положительным.

Результат: персептрон не всегда в состоянии распознать изображение, грань между правильным и неправильным изображением минимальна, поэтому данный метод обучения не рекомендуется к использованию.

Рассмотрим **метод с отрицательными начальными весами**. Зададим скорость обучения, равную -0.2, и начальные веса, равные 0. Зададим правильное изображение, сгенерируем обучающую выборку, проведём обучение на ней. По умолчанию у нас один цикл обучения на выборке из 10 элементов.

Метод с положительными начальными весами. Исходные данные

Скорость обучения	-0.2
Начальные веса входов	0
Количество циклов обучения	1
Размер обучающей выборки	10

Были получены следующие экспериментальные данные.

Значение на выходе персептрона для правильных выборок

Изображения из одного пикселя	NET = - 0.144
Изображения из двух пикселей	NET = - 0.089
Изображения из трех пикселей	NET = - 0.045
Изображения из четырех пикселей	NET = - 0.007
Изображения из пяти пикселей	NET = 0.137
Изображения из девяти пикселей	NET = 0.436

Значение на выходе персептрона для неправильных выборок (наихудший вариант)

Изображения из одного пикселя	NET = - 0.639
Изображения из двух пикселей	NET = - 0.762
Изображения из трех пикселей	NET = - 0.840
Изображения из четырех пикселей	NET = - 0.883
Изображения из пяти пикселей	NET = - 0.956
Изображения из девяти пикселей	NET = - 0.999

Разница между значениями на выходе персептрона для правильного изображения и для неправильного изображения

Изображения из одного пикселя	$\Delta = 0.495$
Изображения из двух пикселей	$\Delta = 0.673$
Изображения из трех пикселей	$\Delta = 0.795$
Изображения из четырех пикселей	$\Delta = 0.876$
Изображения из пяти пикселей	$\Delta = 0.819$
Изображения из девяти пикселей	$\Delta = 0.563$

$$\Delta_{\text{CP}} = 0.704$$

Рассмотренный метод позволяет увеличить зазор между правильными и неправильными изображениями при их распознавании на порядок по сравнению с методами нулевых и положительных начальных весов. Это даёт большой простор для выбора оптимального порогового значения.

Результат: персептрон в состоянии распознать изображение, причем грань между правильным и неправильным изображением возросла на порядок по сравнению с методами нулевых и положительных начальных весов.

В рамках эксперимента пороговое значение выбиралось равным значению правильной выборки, таким образом правильным признавалось изображение, коэффициент NET которого был больше, либо равен NET правильного изображения. В общем случае пороговое значение необходимо выбирать несколько меньше NET правильной выборки, для того, чтобы изображение с незначительными отклонениями от правильного имело возможность считаться правильным также.

В качестве критерия окончания обучения можно брать несколько аспектов: расстояние между правильной и наиболее близким к правильному достигло определённого значения, значение NET правильного изображения достигло определённого значения. Критерии можно использовать одновременно, поскольку каждый из них может быть в частном случае недостижим, и тогда наша нейронная сеть окажется необученной.

С каждым новым циклом обучения показатели NET для всех случаев изменяются.

Анализ полученных результатов

В итоге получаем комплекс различных характеристик, которые, так или иначе, влияют на скорость обучения персептрона. По результатам эксперимента можно дать следующие рекомендации для выбора оптимальных параметров обучения:

1. **Скорость обучения** - лучше использовать динамический подход.
2. **Начальные коэффициенты** лучше выбирать отрицательными. Это позволит получить наибольшую разницу между верными и неверными значениями NET.
3. Обучать персептрон нужно на **характерных выборках**. Чем меньше размер обучающей выборки, тем быстрее происходит процесс обучения.
4. **Количество циклов обучения** важно выбирать в зависимости от необходимой точности (чем больше циклов, тем больше точность).
5. **Пороговое значение** лучше выбирать как среднее значение между значением NET и максимальным из неправильных NET.

Список литературы

1. Уоссермен Ф. Нейрокомпьютерная техника: теория и практика. М.: Мир, 1992 г. – 126 с.
2. Хайкин С. Нейронные сети: полный курс, 2-е издание. : Пер. с англ. – М.: Издательский дом «Вильямс», 2006 г. – 1104 с.