**Л 14. Искусственные нейронные сети. Стохастические сети и сети с обратными связями**

Стохастические методы обучения. Обучение Больцмана. Алгоритм обучения Больцмана. Обучение Коши. Сети с обратными связями. Сеть Хопфилда. Правило обучения Хебба. Процедура ортогонализации образов. Сеть ДАП (двунаправленная ассоциативная память). Сеть APT (адаптивная резонансная теория).

**Стохастические методы обучения**

Искусственная нейронная сеть обучается посредством некоторого процесса, модифицирующего ее веса. Если обучение успешно, то предъявление сети множества входных сигналов приводит к появлению желаемого множества выходных сигналов.**Стохастические методы обучения** выполняют псевдослучайные изменения величин весов, сохраняя те изменения, которые ведут к улучшениям.

Локальные минимумы мешают всем алгоритмам обучения, основанным на поиске минимума функции ошибки, включая сети обратного распространения, и представляют серьезную и широко распространенную проблему.

Стохастические методы позволяют решить эту проблему. Стратегия коррекции весов, вынуждающая веса принимать значение глобального оптимума, возможна.

В качестве объясняющей аналогии предположим, что на рис. 1 изображен шарик на поверхности в коробке. Если коробку сильно потрясти в горизонтальном направлении, то шарик будет быстро перекатываться от одного края к другому.

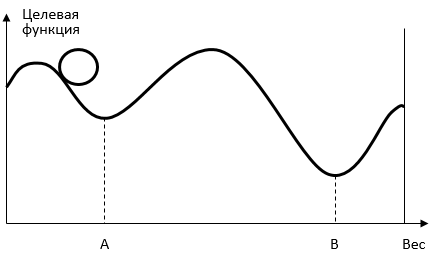


Рис. 1. Проблема локальных минимумов

Нигде не задерживаясь, в каждый момент шарик будет с равной вероятностью находиться в любой точке поверхности. Если постепенно уменьшать силу встряхивания, то будет достигнуто условие, при котором шарик будет на короткое время «застревать» в точке В. При еще более слабом встряхивании шарик будет на короткое время останавливаться как в точке А, так и в точке В. При непрерывном уменьшении силы встряхивания будет достигнута критическая точка, когда сила встряхивания достаточна для перемещения шарика из точки *А* в точку *В*, но недостаточна для того, чтобы шарик мог выбраться из *В* в *А.* Таким образом, окончательно шарик остановится в точке глобального минимума, когда амплитуда встряхивания уменьшится до нуля.

**Обучение Больцмана**

Искусственные нейронные сети могут обучаться, по существу, тем же самым образом посредством случайной коррекции весов. Вначале делаются большие случайные коррекции с сохранением только тех изменений весов, которые уменьшают целевую функцию. Затем средний размер шага постепенно уменьшается, и глобальный минимум в конце концов достигается.

Такая процедура напоминает отжиг металла, поэтому для ее описания часто используют термин **«имитация отжига».** В металле, нагретом до температуры, превышающей его точку плавления, атомы находятся в сильном беспорядочном движении. Как и во всех физических системах, атомы стремятся к состоянию минимума энергии, но при высоких температурах энергия атомных движений препятствует этому. В процессе постепенного охлаждения металла возникают все более низкоэнергетические состояния, пока в конце концов не будет достигнуто наиболее низкое из возможных состояний, глобальный минимум. В процессе отжига распределение энергетических уровней описывается следующим соотношением:

где - вероятность того, что система находится в состоянии с энергией *-* постоянная Больцмана;- температура по шкале Кельвина.

При высоких температурах вероятность приближается к единице для всех энергетических состояний. Таким образом, высокоэнергетическое состояние почти столь же вероятно, как и низкоэнергетическое. По мере уменьшения температуры вероятность высокоэнергетических состояний уменьшается по сравнению с низкоэнергетическими. При приближении температуры к нулю становится весьма маловероятным, чтобы система находилась в высокоэнергетическом состоянии.

Этот стохастический метод непосредственно применим к обучению искусственных нейронных сетей и относится к классу алгоритмов обучения с учителем.

**Алгоритм обучения Больцмана**

Шаг 1. Определить переменную , представляющую искусственную температуру. Придать большое начальное значение.

Шаг 2. Подать на вход сети один из входных образов обучающей выборки и вычислить реальный выход и значение функции ошибки сети (как в алгоритме обратного распространения).

Шаг 3. Придать случайное изменение выбранному весу и пересчитать выход сети и изменение функции ошибки в соответствии со сделанным изменением веса.

Шаг 4. Если функция ошибки уменьшилась, то сохранить изменение веса. Если изменение веса приводит к увеличению функции ошибки, то вероятность сохранения этого изменения вычисляется с помощью распределения Больцмана: . Выбирается случайное число из равномерного распределения от нуля до единицы. Если вероятность больше, чем , то изменение сохраняется, в противном случае величина веса возвращается к предыдущему значению.

Шаг 5. Повторять шаги 3 и 4 для каждого из весов сети, постепенно уменьшая температуру , пока не будет достигнуто допустимо низкое значение целевой функции.

Шаг 6. Повторять шаги 2—5 для всех векторов обучающей выборки (возможно неоднократно), пока функция ошибки не станет допустимой для каждого из них.

Замечание 1. На шаге 4 система может делать случайный шаг в направлении, портящем функцию ошибки, позволяя ей тем самым вырываться из локальных минимумов, где любой малый шаг увеличивает целевую функцию.

Замечание 2. В работах, посвященных больцмановскому обучению, показано, что для достижения сходимости к глобальному минимуму скорость уменьшения искусственной температуры должна подчиняться закону

где - номер итерации обучения. Этот результат предсказывает очень медленную сходимость процесса обучения, что является существенным недостатком метода.

**Обучение Коши**

В этом методе распределение Больцмана заменяется на распределение Коши. Распределение Коши имеет, как показано на рис. 2, более высокую вероятность больших шагов. Дисперсия распределения Коши бесконечна.

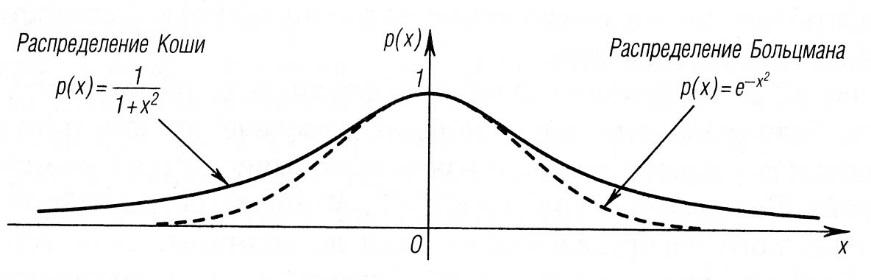


Рис. 2. Распределение Коши и распределение Больцмана

С помощью такого простого изменения максимальная скорость уменьшения температуры становится обратно пропорциональной линейной величине, а не логарифму, как для алгоритма обучения Больцмана. Эта связь может быть выражена следующим образом: .

Таким образом, время обучения резко уменьшается. Распределение Коши имеет вид

где - вероятность принять изменение веса .

Несмотря на улучшение скорости обучения, даваемое распределением Коши по сравнению с распределением Больцмана, время сходимости все еще может в 100 раз превышать время для алгоритма обратного распространения.

Комбинирование обратного распространения с обучением Коши.

Коррекция весов в комбинированном алгоритме, использующем обратное распространение и обучение Коши, состоит из двух компонент: компоненты, вычисляемой с использованием алгоритма обратного распространения, и случайной компоненты, определяемой распределением Коши.

Эти компоненты вычисляются для каждого веса, и их сумма является величиной, на которую изменяется вес. Как и в алгоритме Коши, после вычисления изменения веса вычисляется целевая функция. Если имеет место улучшение, изменение сохраняется. В противном случае оно сохраняется с вероятностью, определяемой распределением Коши.

Коррекция веса вычисляется с использованием представленных ранее уравнений для каждого из алгоритмов:

где - коэффициент, управляющий относительными величинами обучения Коши и обратного распространения в компонентах весового шага.

Если приравнивается нулю, метод становится обучением Коши. Если приравнивается единице, метод становится алгоритмом обратного распространения.

Комбинированная сеть, использующая обратное распространение и обучение Коши, обучается быстрее, чем каждый из алгоритмов в отдельности. Сходимость к глобальному минимуму гарантируется алгоритмом Коши, и во многих экспериментах по обучению сеть практически не попадала в локальные минимумы.

**Сети с обратными связями**

Рассмотренные ранее нейросетевые архитектуры относятся к классу сетей с направленным потоком распространения информации и не содержат обратных связей. После обучения на этапе функционирования сети каждый нейрон выполняет свою функцию - передачу выходного сигнала - ровно один раз. В общем случае может быть рассмотрена нейронная сеть, содержащая произвольные**обратные связи**, т. е. пути, передающие сигналы от выходов к входам. Отклик таких сетей является динамическим, т. е. после подачи нового входа вычисляется выход и, передаваясь по обратной связи, модифицирует вход. Затем выход повторно вычисляется, и процесс повторяется снова и снова. Для устойчивой сети последовательные итерации приводят к все меньшим изменениям выхода, и в результате выход становится постоянным. Для многих сетей процесс никогда не заканчивается, такие сети называются неустойчивыми. Неустойчивые сети обладают интересными свойствами и могут рассматриваться в качестве примера хаотических систем, но для большинства практических приложений используются сети, которые дают постоянный выход.

**Сеть Хопфилда**

Нейросетевые архитектуры получили всеобщее признание во многом благодаря исследованиям Джона Хопфилда, физика из Калифорнийского технологического института. Он изучал свойства сходимости сетей на основе принципа минимизации энергии, а также разработал на основе этого принципа семейство нейросетевых архитектур.

Рассмотрим однослойную сеть с обратными связями, состоящую из входов и нейронов (рис. 3). Каждый вход связан со всеми нейронами. Так как выходы сети заново подаются на входы, то *-* это значение -го выхода, который на следующем этапе функционирования сети становится -м входом.

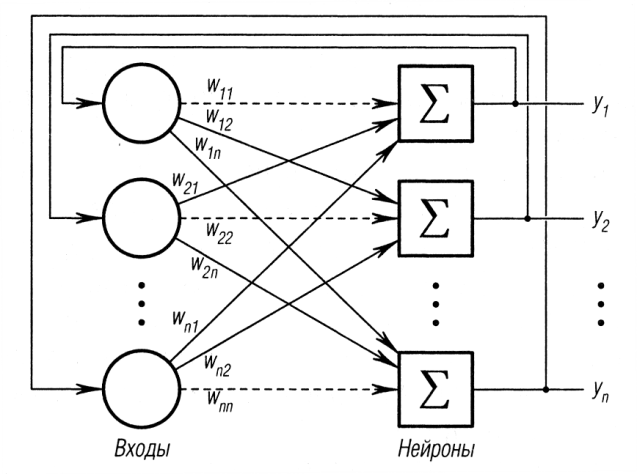


Рис. 3. Модель сети Хопфилда

Совокупность выходных значений всех нейроновна некотором этапеобразует**вектор состояния сети** Нейродинамика приводит к изменению вектора состояния на

Обозначим силу синаптической связи от -го входа к -му нейрону как . Каждый *-*й нейрон сети реализует пороговую активационную функцию следующего вида:

Здесь - значение выхода -го нейрона на предыдущем этапе функционирования сети; - пороговое значение -го нейрона.

В модели Хопфилда предполагается условие симметричности связей с нулевыми диагональными элементами . Устойчивость такой сети может быть доказана следующим образом. Введем в рассмотрение функцию, зависящую от состояния сети и называемую функцией энергии сети Хопфилда:

Вычислим изменение функции энергии , вызванное изменением состояния -го нейрона *.*

(здесь мы воспользовались симметричностью связей и тем, что ). Допустим, что величина больше порога . Тогда выражение в правой части положительно, а из вида активационной функции следует, что новый выход нейрона должен стать равным 1, т. е. измениться в положительную сторону (или остаться без изменения). Это значит, что , и тогда . Следовательно, энергия сети либо уменьшится, либо останется без изменения. Далее допустим, что величина меньше порога. Тогда получается новое значение и величина *.* может быть только отрицательной или нулем. Следовательно, опять энергия должна уменьшиться или остаться без изменения. Если величинаравна порогу изменение равно нулю и энергия остается без изменения.

Эти рассуждения показывают, что любое изменение состояния нейрона либо уменьшит функцию энергии, либо оставит ее без изменения. Так как функция энергии задана на конечном множестве то она ограничена снизу, и вследствие непрерывного стремления к уменьшению, в конце концов, должна достигнуть минимума и прекратить изменение. По определению такая сеть является устойчивой.

Поверхность функции энергии в пространстве состояний имеет весьма сложную форму с большим количеством локальных минимумов. Стационарные состояния, отвечающие минимумам, могут интерпретироваться как**образы памяти нейронной сети**. Сходимость к такому образу соответствует процессу извлечения из памяти. При произвольной матрице связей образы также произвольны. Для записи в память сети какой-либо конкретной информации требуется определенное значение весов , которое может получаться в процессе обучения.

Нейронная сеть Хемминга была предложена в 1987 году Р. Липпманом [Lippman, 1987]. Она представляет собой релаксационную многослойную сеть с отрицательными фиксированными связями между отдельными слоями, а весовые коэффициенты и пороговые значения нейронов определяются из условий задачи. Данная сеть характеризуется, по сравнению с сетью Хопфилда, меньшим количеством производимых вычислений и объемом используемой оперативной памяти компьютера.

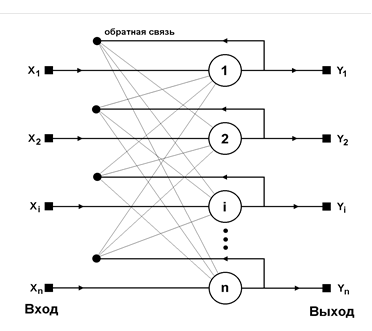


Рис. 4. ИНС Хопфилда

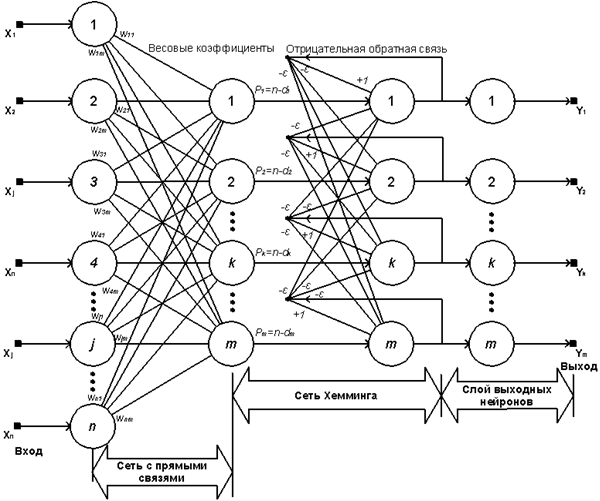
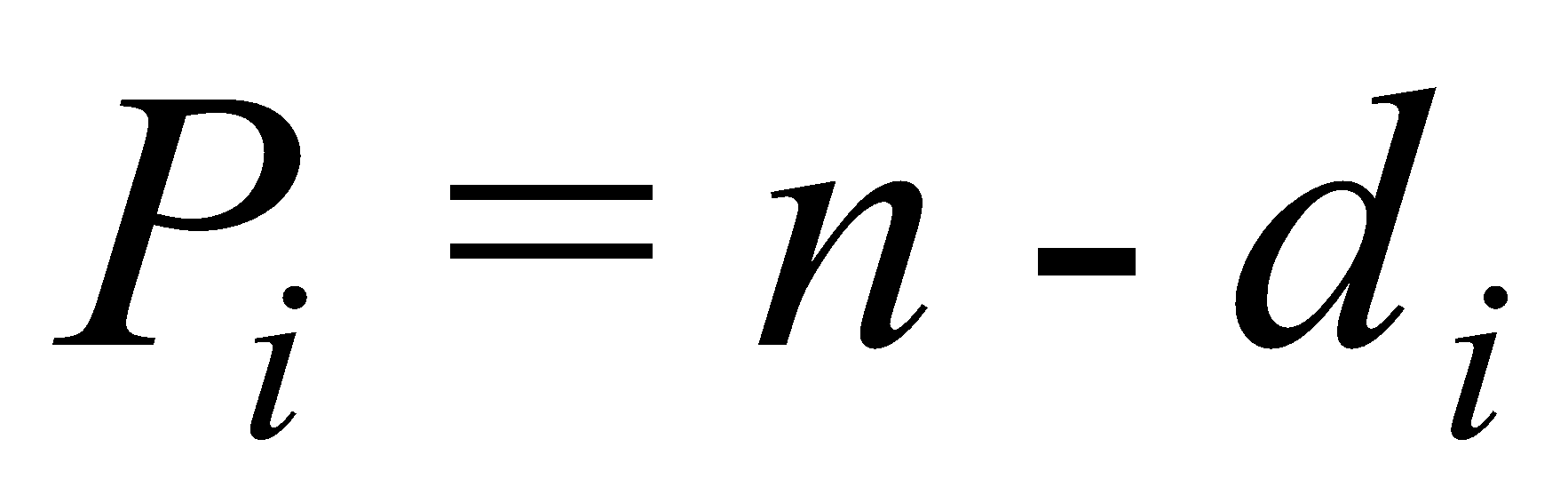
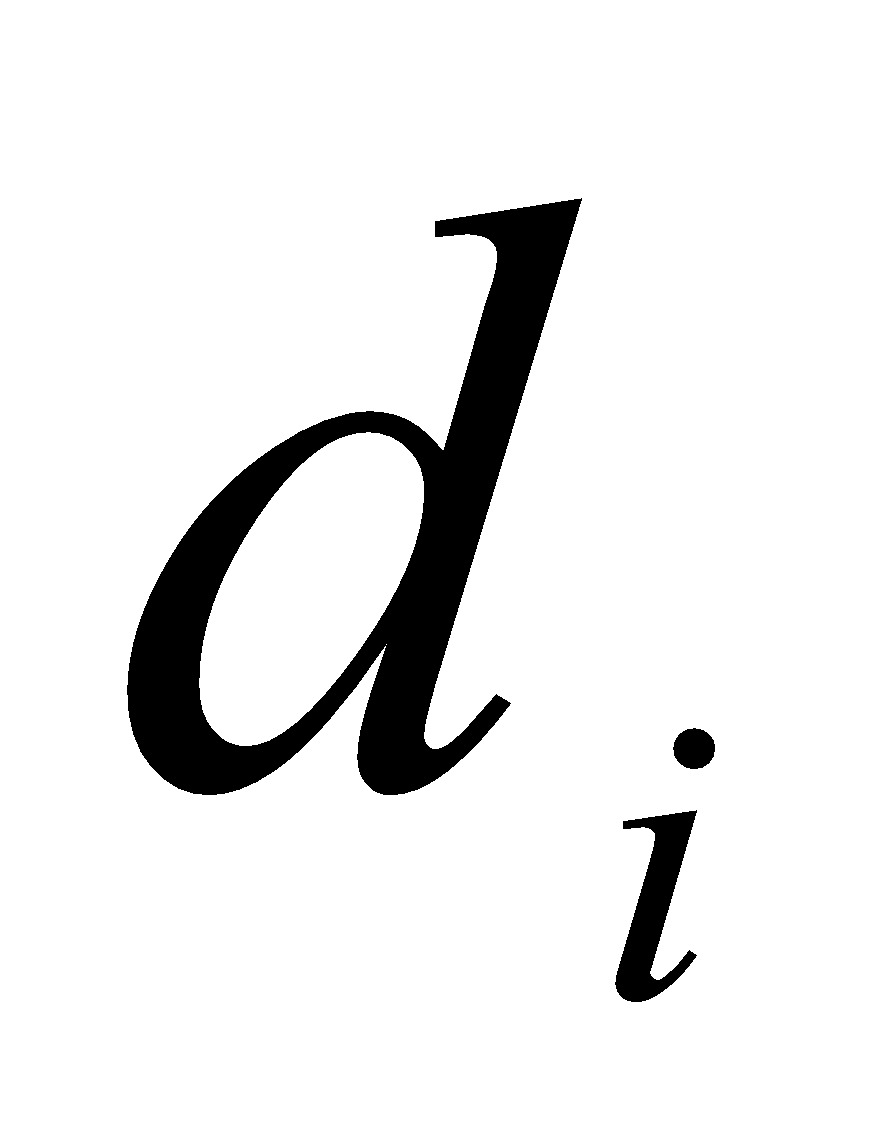


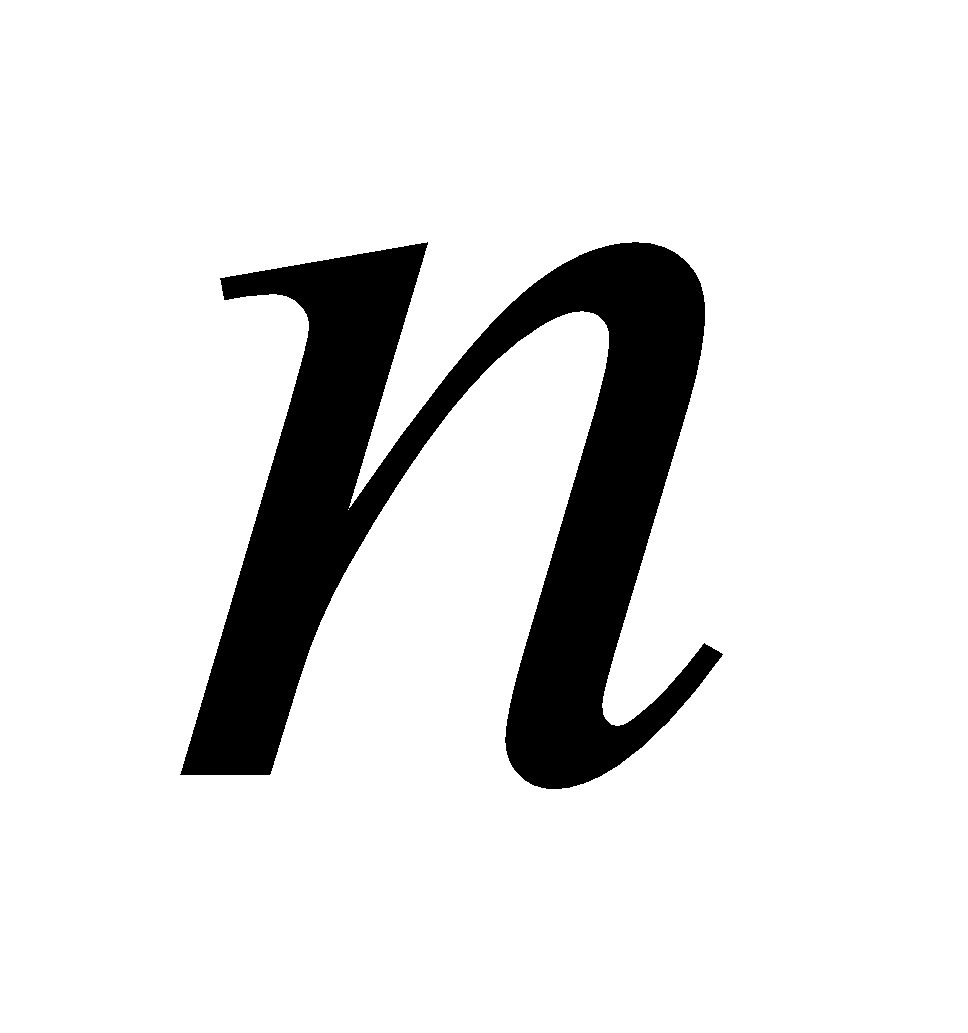
Рис. 5. ИНС Хемминга

ИНС Хемминга многослойна и состоит из: сети с прямыми связями, сети Хемминга и слоя выходных нейронов. Идея её работы состоит в нахождении эталона с минимальным расстоянием Хемминга (число отличающихся значений в двух бинарных векторах) до входного визуального образа и активизацией только одного выхода сети, соответствующего номеру этого эталона.

Сеть с прямыми связями состоит из первого - выполняющего роль распределительного и второго слоев, которые содержат *n* входных и *m* выходных нейронов, где *m* – число эталонных образов. Она вычисляет меру подобия между входным и эталонными образами, хранимыми в ней. Выходное значение *i*–го нейрона 2–го слоя представляет собой меру подобия *Pi* между входным и *i*–м эталонным образом:

,

где  - расстояние Хемминга между входным и *i*–м эталонным образом,

 - размерность входного вектора.

Сеть Хемминга (слои 2 и 3) используется для разрешения возникающих конфликтов, когда входной визуальный образ является подобным нескольким эталонам, хранимым в сети. При этом на выходе ИНС остается один нейрон – победитель.

Нейроны третьего слоя связаны между собой отрицательными обратными синаптическими связями. Единственный синапс с положительной обратной связью для каждого нейрона соединен с его же аксоном.

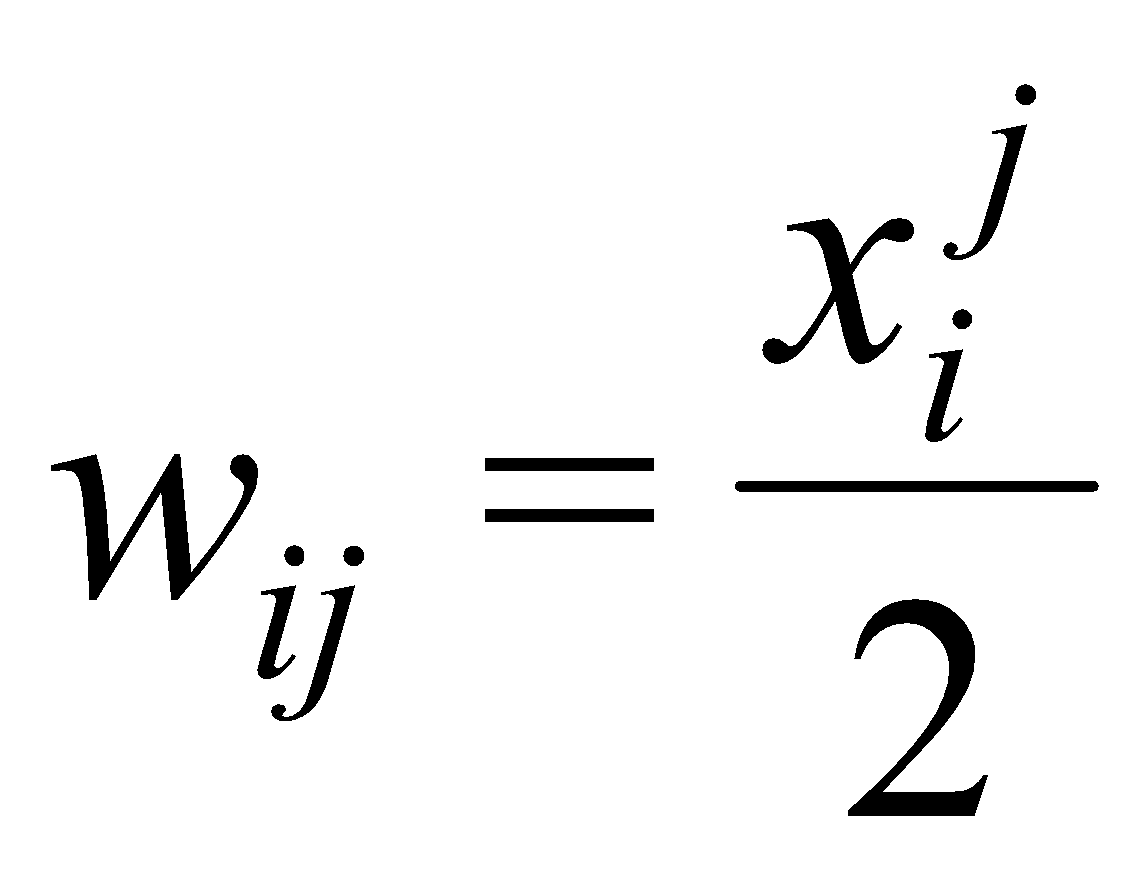
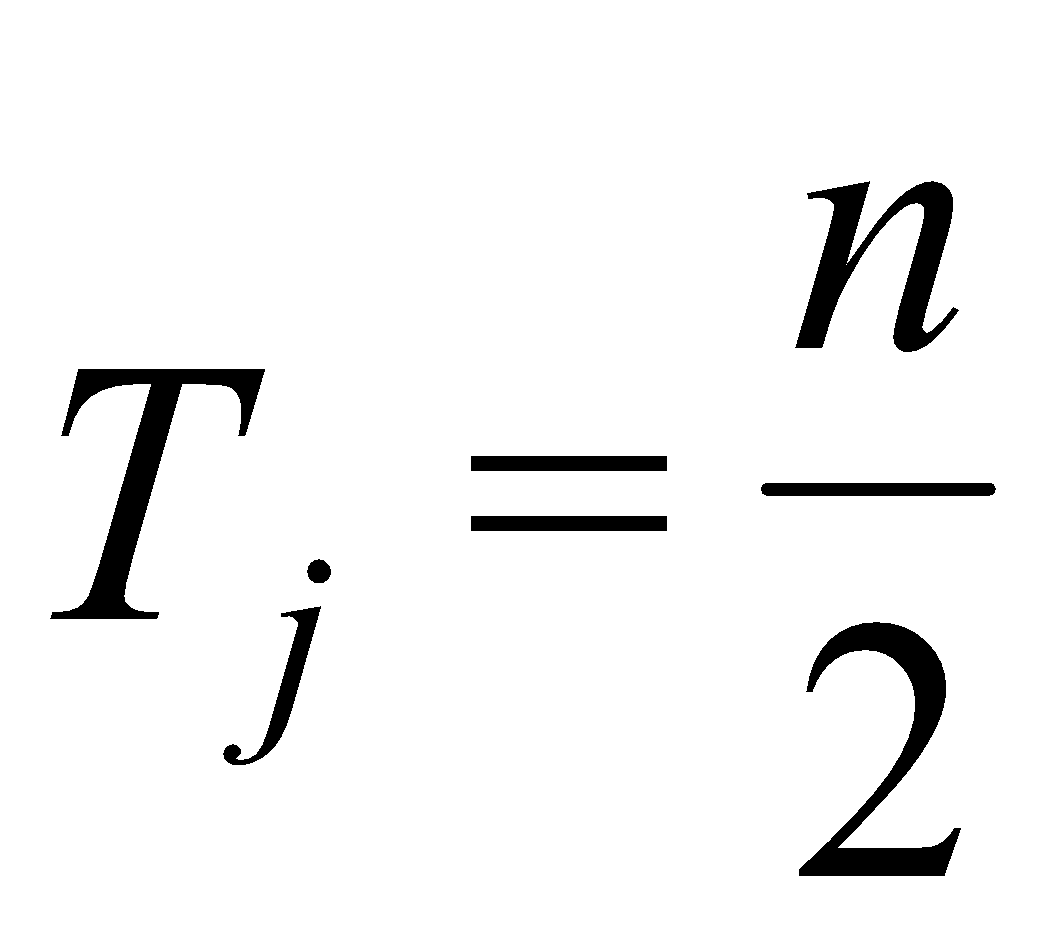
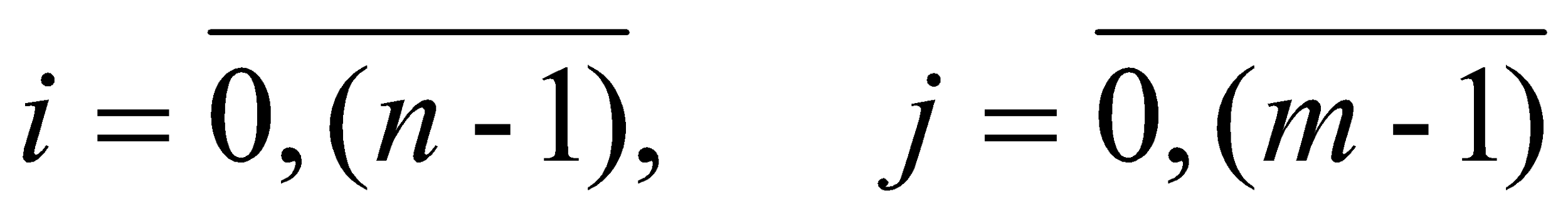
Выходной слой ИНС (4 слой) состоит из *m* нейронов, каждый из которых имеет пороговую функцию активации. Он предназначен для преобразования выхода активного нейрона – победителя сети Хемминга в единичное значение, а остальных нейронов выходного слоя в нулевое состояние. Таким образом, происходит идентификация входного визуального образа, который

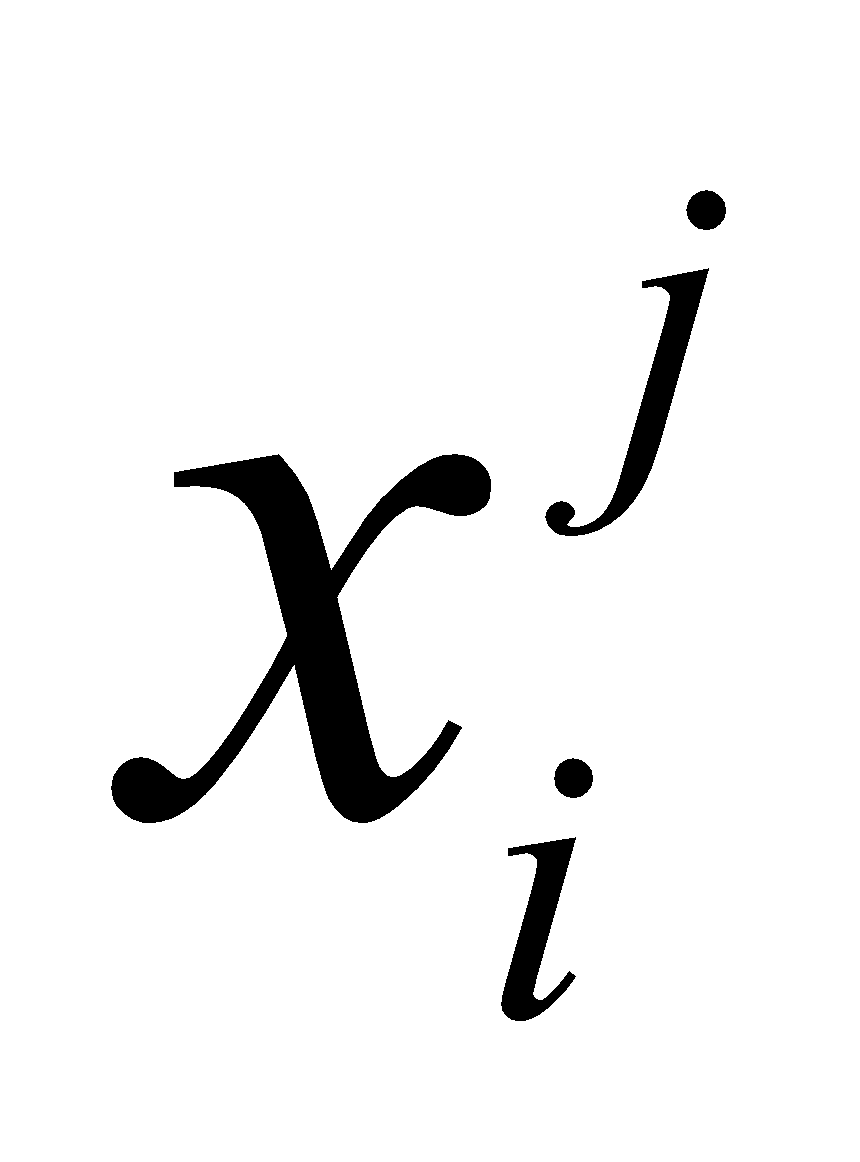
кодируется номером нейрона выходного слоя, имеющего единичное значение.

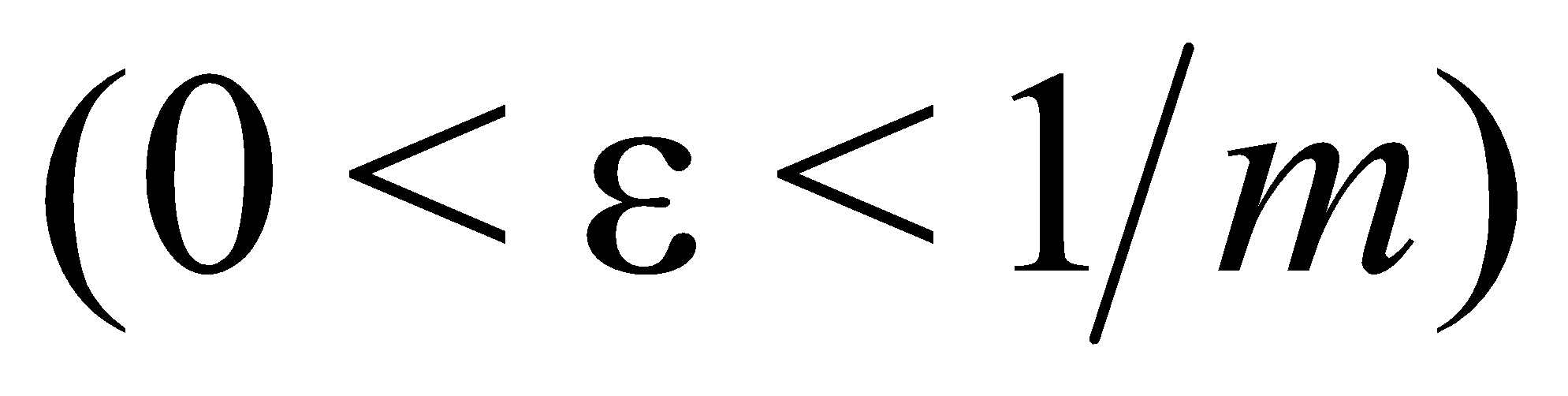
Если входной визуальный образ не совпадет ни с одним из эталонных, то на выходе ИНС будет формироваться номер такого эталонного образа, который имеет минимальное расстояние Хемминга по отношению к входному визуальному образу. Если в выходном слое существует несколько нейронов-победителей, то выбор одного из них осуществляется случайным образом.

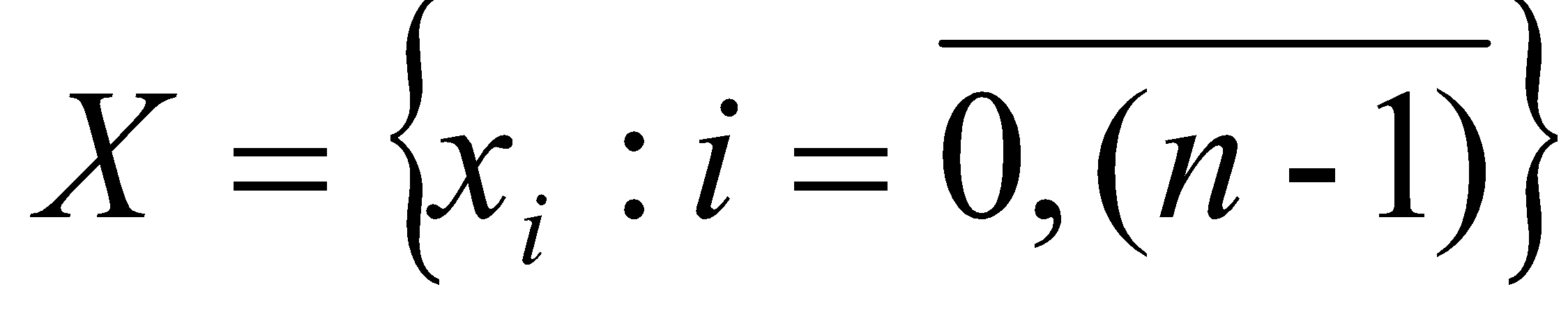
Алгоритм функционирования ИНС Хемминга

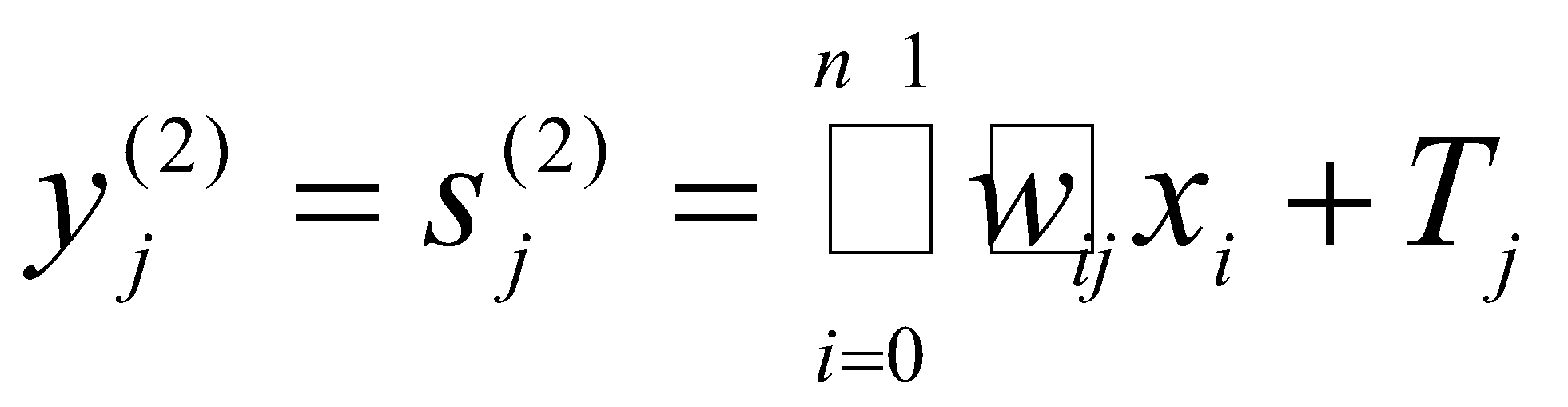
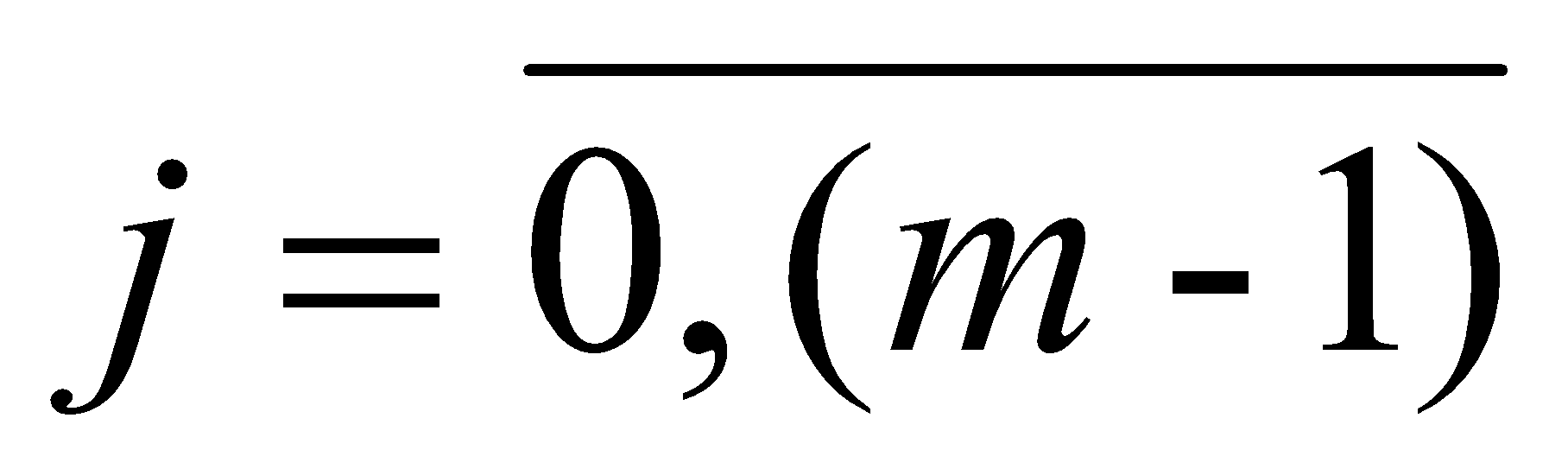
1. На стадии инициализации весовым коэффициентам второго слоя и порогу активационной функции присваиваются следующие значения

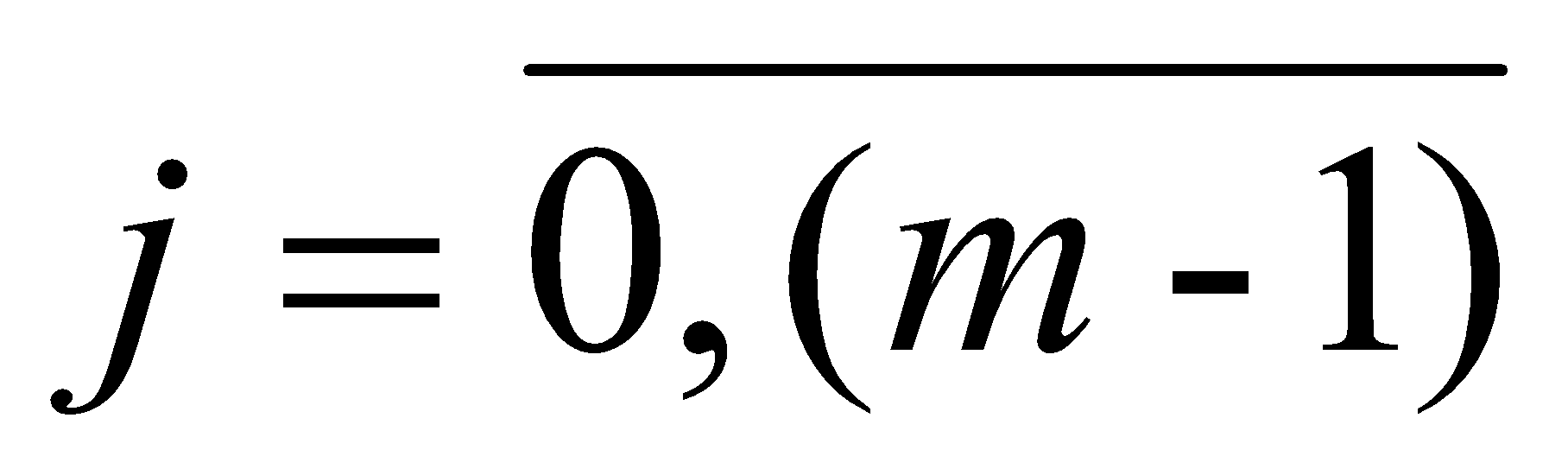
*,, *;

где *n* – число входов, *m* – число входов ИНС,  – *i*-ый элемент *j*-ого образца.

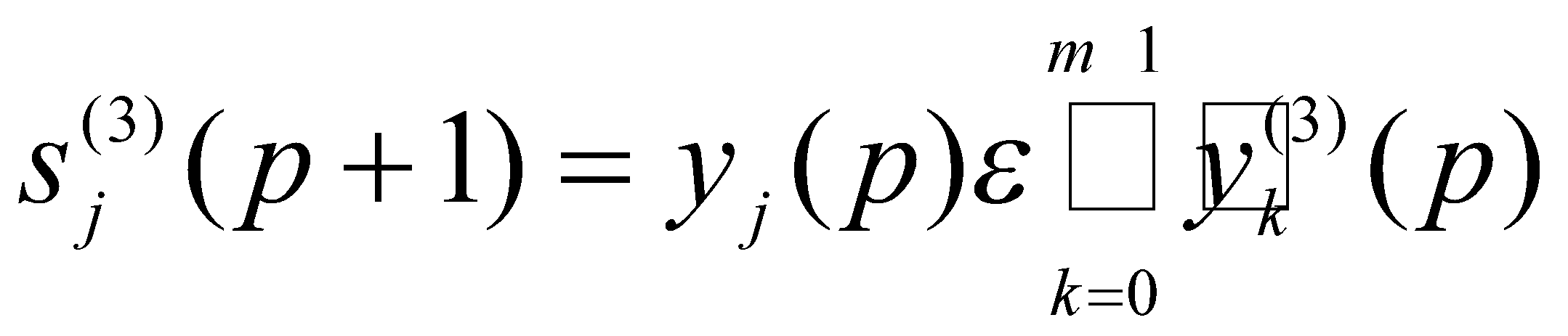
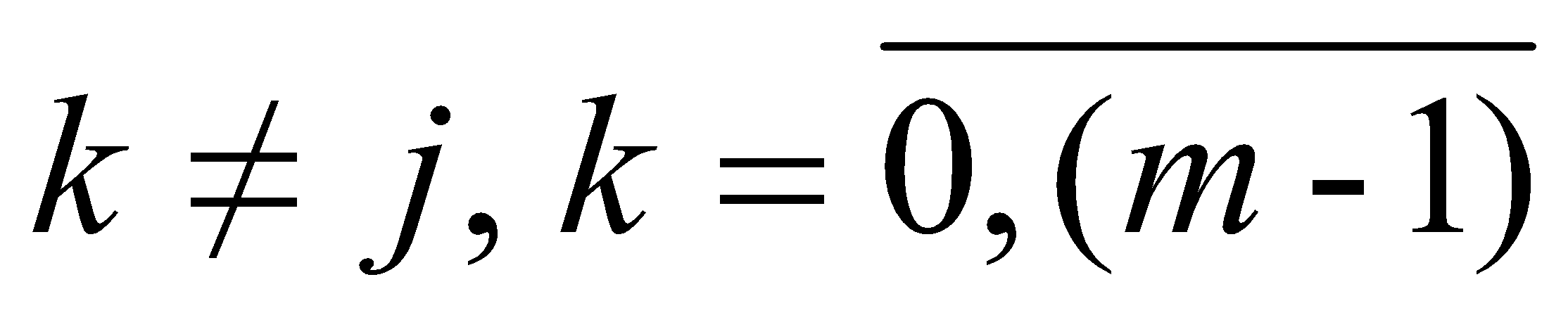
Весовые коэффициенты тормозящих синапсов во втором слое равны величине *ε* . Синапс нейрона, связанный с его же аксоном имеет вес +1.

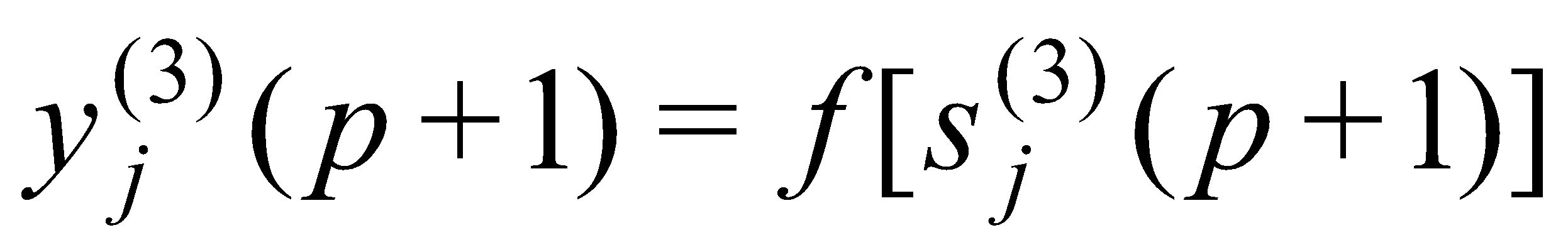
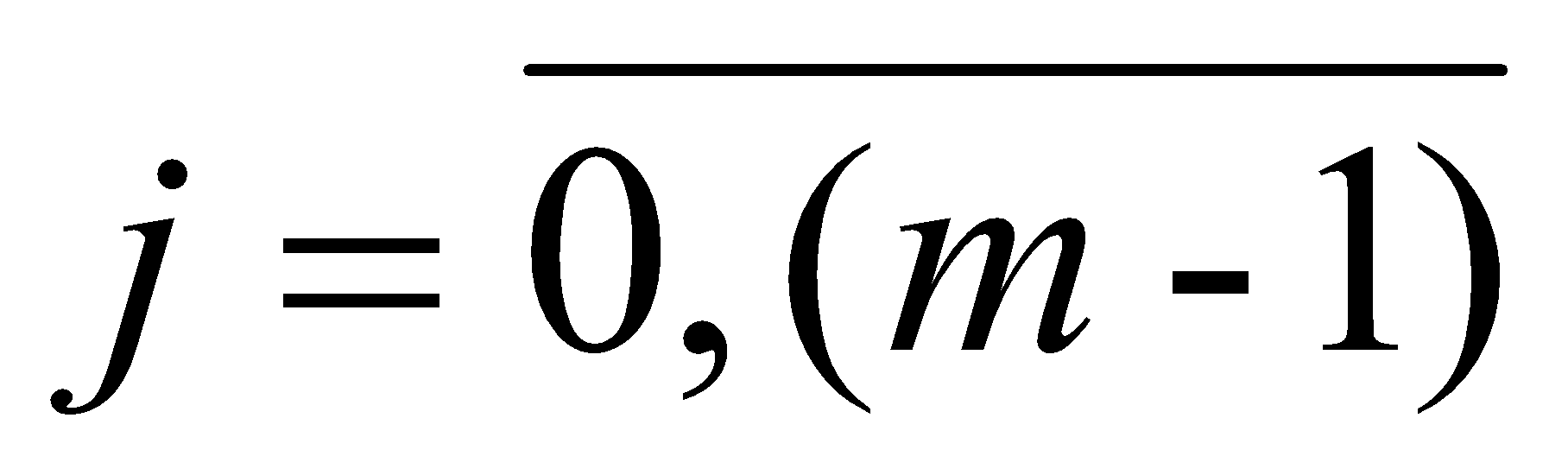
2. На входы сети подается неизвестный вектор *,* исходя из которого рассчитываются состояния нейронов второго слоя (верхний индекс в скобках указывает номер слоя)

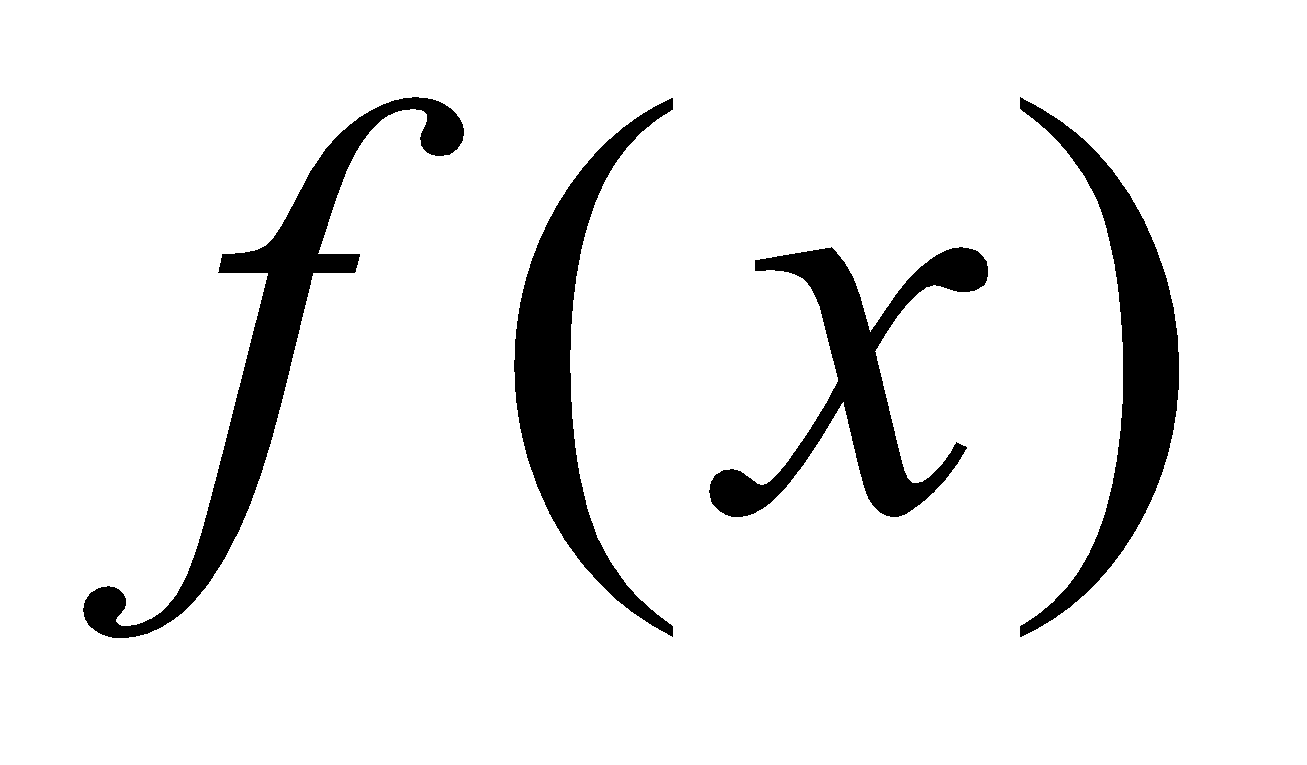
, **.

Полученными значениями инициализировать состояния нейронов третьего слоя *yj(3) = yj(2)*, **.

3. Вычисляются новые состояния нейронов третьего слоя

, **

и значения их аксонов , **.

Активационная функция имеет вид порога (рис. 6), причем её величинадолжна быть достаточно большой, чтобы любые возможные значения аргумента не приводили к насыщению.

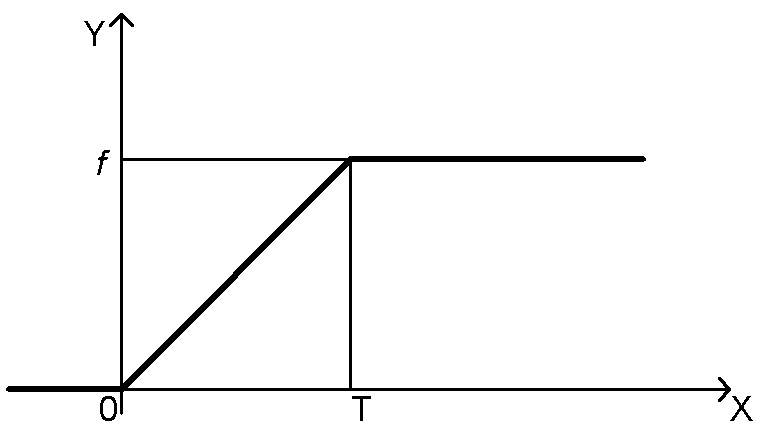
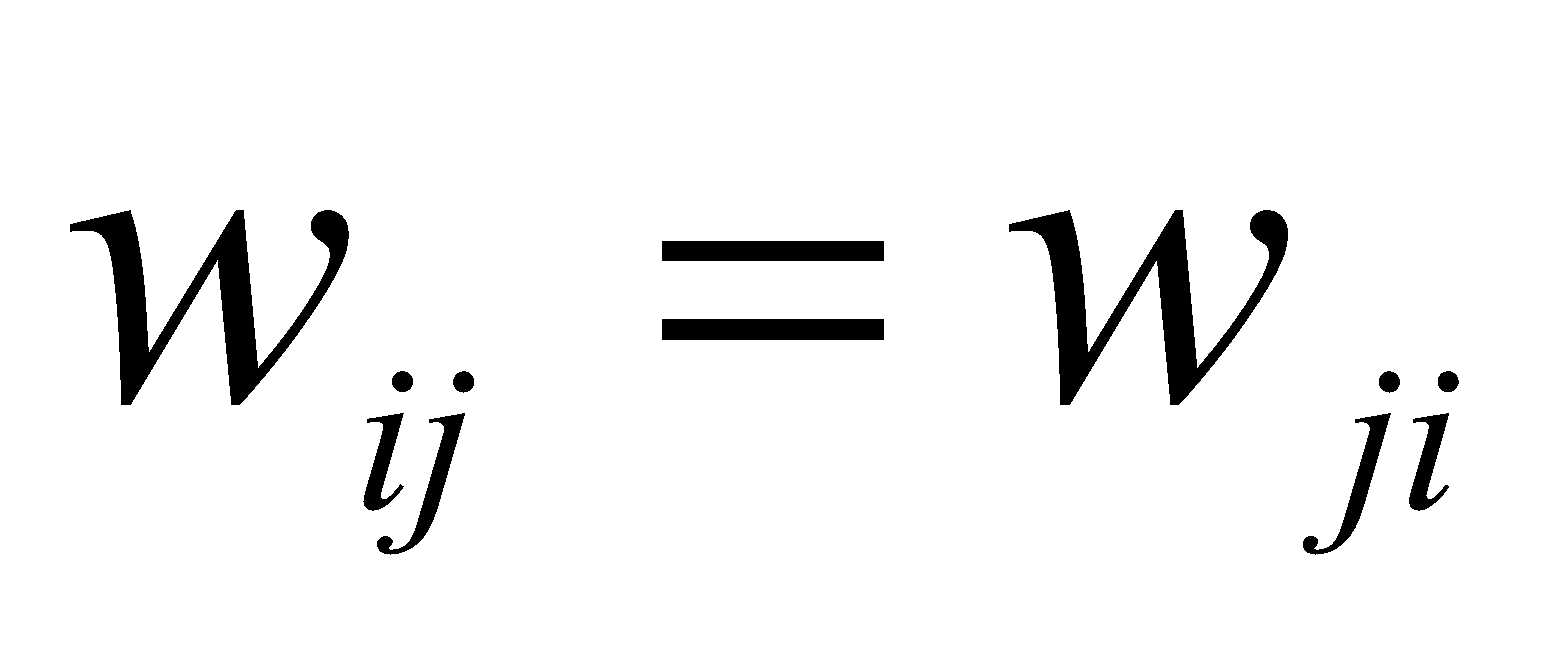
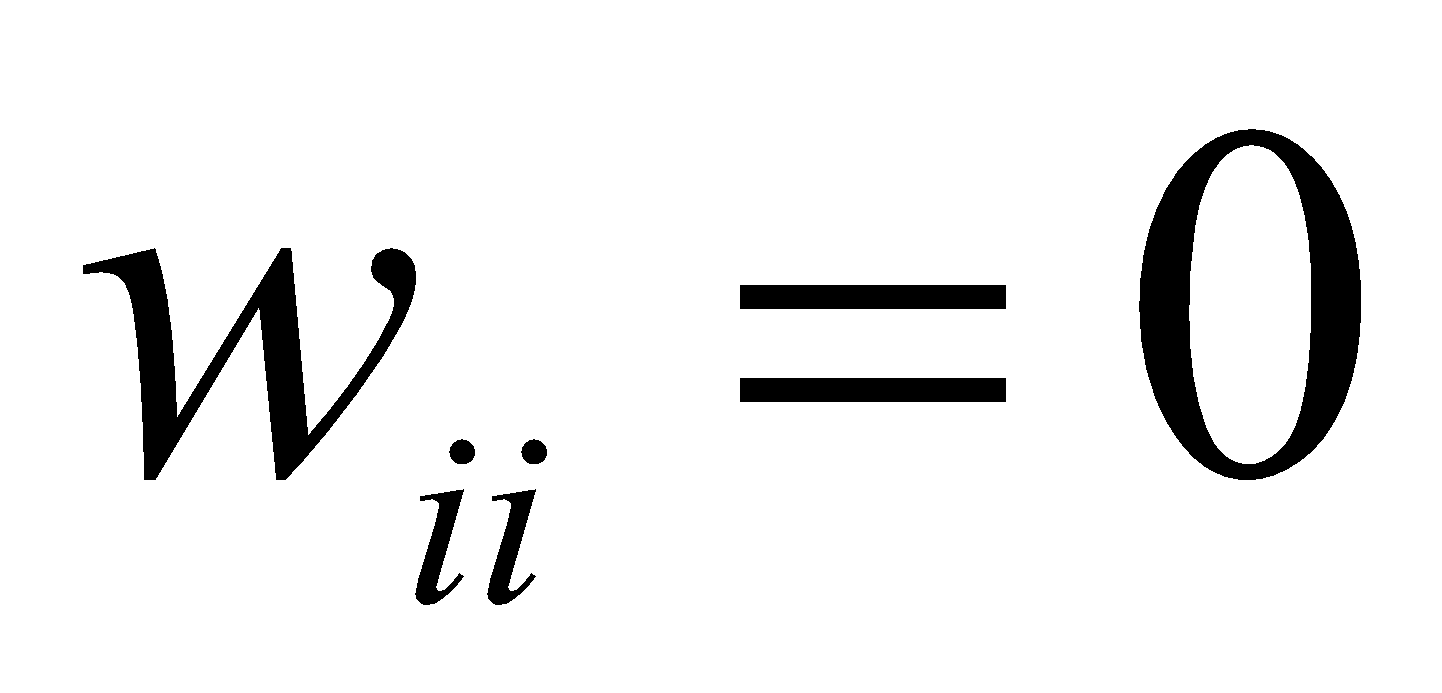
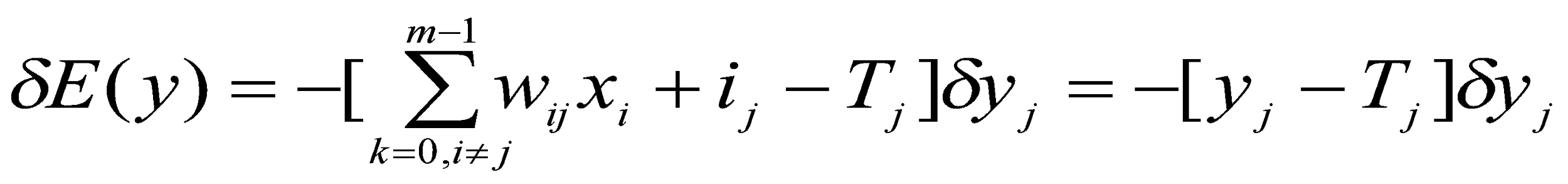
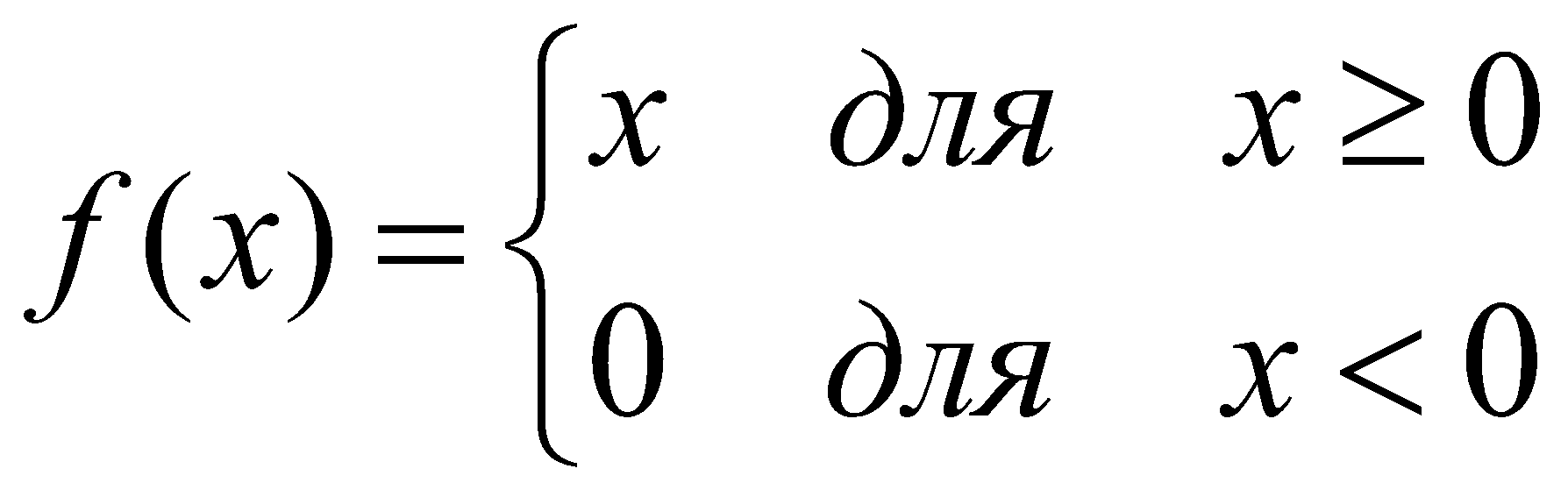


Рис. 6. Активационная функция нейрона ИНС Хемминга

Это следует из того, что теорема об устойчивости работы ИНС Хемминга [Hopfield, 1982; Тэнк, Хопфилд, 1988; Lippman, 1987] требует от матрицы весов симметричности и наличия нулевой диагонали ,. При этом изменение энергии, вызванное изменением состояния нейрона с номером *j*, определяется по формуле

,

где *ij* - взвешенная сумма возбуждений *j*-го нейрона, *Тj* - пороговое значение заданное внешним источником для *j*-го нейрона. Любое изменение состояния одного нейрона не увеличивает величины энергии всей системы. Поэтому активационная функция нейронов имеет следующий вид

.

4. Проверить, изменились ли выходы нейронов второго слоя по сравнению с предыдущим шагом алгоритма. Если да – перейди к шагу 3, в противном случае – конец.

Оценивая алгоритм функционирования нейронной сети Хемминга можно сделать вывод о том, что роль второго слоя весьма условна: воспользовавшись один раз на шаге 1 значениями его весовых коэффициентов, сеть больше не обращается к нему, поэтому второй слой может быть вообще исключен из сети - заменен на матрицу весовых коэффициентов.

**Правило обучения Хебба**

Метод обучения для сети Хопфилда опирается на исследования Дональда Хебба, реализовавшего простой механизм обучения, названный**правилом Хебба***.*

Пусть задана обучающая выборка образов Требуется построить матрицу связей такую, что соответствующая нейронная сеть будет иметь в качестве стационарных состояний образы обучающей выборки (значения порогов нейронов положим равными нулю). В случае одного обучающего образа , правило Хебба приводит к матрице Покажем, что состояние является стационарным для сети Хопфилда с данной матрицей

Действительно, значение функции энергии в состоянии является для нее глобальным минимумом:

т. е. сеть прекратит изменения, достигнув состояния .

Для запоминания образов применяется итерационный процесс:

(считаем, что ). Этот процесс приводит к полной матрице связей:

Сеть Хопфилда нашла широкое применение в системах**ассоциативной памяти**, позволяющих восстанавливать идеальный образ по имеющейся неполной или зашумленной его версии.

**Процедура ортогонализации образов**

Два различных запоминаемых векторных образа сети называются **ортогональными**, если их скалярное произведение равно нулю: *.* Если все запоминаемые образы сетипопарно ортогональны, емкость памяти сети Хопфилда увеличивается до , т. е. сеть может запомнить количество образов, не превосходящее число нейронов в ней. На этом свойстве основано улучшение правила Хебба: перед запоминанием в нейронной сети исходные образы следует ортогонализовать. Процедура расчета весовых коэффициентов в этом случае имеет следующий вид:

Шаг 1. Вычисляются элементы матрицы *:*

Шаг 2. Определяется матрица , обратная к матрице *.*

Шаг 3. Задаются весовые коэффициенты сети Хопфилда:

Существенным недостатком метода является его нелокальность: прежде чем начать обучение, необходимо наперед знать все обучающие образы. Добавление нового образа требует полного переобучения сети.

**Сеть ДАП** **(двунаправленная ассоциативная память)**

Сеть Хопфилда реализует так называемую автоассоциативную память. Это означает, что образ может быть завершен или исправлен, но не может быть ассоциирован с другим образом. Двунаправленная ассоциативная память (ДАП), разработанная в 1988 году Бертом Коско (В. Kosko), является гетероассоциативной: она сохраняет пары образов и выдает второй образец пары, когда ассоциированный с ним первый образец подается на вход сети. Как и сеть Хопфилда, сеть ДАП способна к обобщению, вырабатывая правильные реакции, несмотря на искаженные входы. Сеть ДАП (рис. 7) содержит два слоя нейронов. Элементы весовой матрицы отражают связь между -м нейроном первого слоя и *-м* нейроном второго слоя, .

В процессе функционирования сети входной вектор умножается на транспонированную матрицу весов сетии подается на вход первого слоя, в результате чего вырабатывается вектор выходных сигналов нейронов первого слоя *.* Вектор затем умножается на матрицуи подается на вход второго слоя, который вырабатывает выходные сигналы, представляющие собой новый входной вектор . Этот процесс повторяется до тех пор, пока сеть не достигнет стабильного состояния, в котором ни вектор *,* ни вектор не изменяются. Нейроны в обоих слоях сети ДАП функционируют аналогично нейронам сети Хопфилда.

Этот процесс может быть выражен следующим образом:

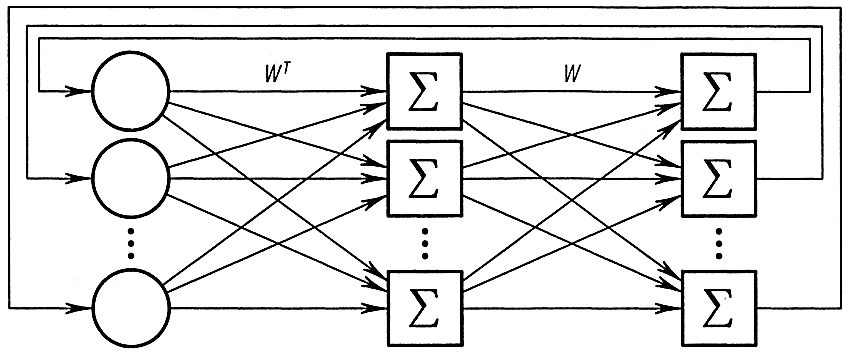


Рис. 7. Структура сети ДАП

где

значение функции активации данного нейрона на предыдущем шаге.

Пусть задана обучающая выборка ассоциированных образов . Весовая матрица сети ДАП вычисляется как сумма произведений всех векторных пар обучающего набора:

В отличие от сети Хопфилда, весовая матрица в сети ДАП не квадратная, что во многих случаях позволяет оптимизировать вычислительные затраты, необходимые для функционирования сети. Если рассмотреть пример с запоминанием букв А, В, С, то сеть Хопфилда в этом случае имела бы 10x7=70 входов и требовала для своей работы хранения весовой матрицы размером 70х70, содержащей 4900 элементов. Ассоциируем с каждым из входных образов сети двухбитовый вектор: символ А будет связан с вектором (1,-1,-1), символ В с вектором (-1, 1,-1), символ С с вектором (-1, -1, 1). Таким образом, например, при подаче на вход искаженной версии буквы А, сеть после стабилизации должна выдавать образ (1, -1, -1). Так как ассоциированные пары заранее известны, это приведет к правильному распознаванию зашумленного входа. Но для работы такой сети требуется хранение всего 70x3 = 210 элементов весовой матрицы.

Основным недостатком сети ДАП, как и сети Хопфилда, является небольшая емкость памяти. Так, число запоминаемых ассоциаций не может превышать числа нейронов в меньшем слое. Если все пороговые значения нулевые, то оценка еще ухудшается: размер запоминаемой выборки не должен превосходить , где — число нейронов в меньшем слое. Если этот лимит превышен, сеть начинает вырабатывать неверные выходные сигналы, воспроизводя ассоциации, которым не обучена.

**Сеть APT (адаптивная резонансная теория)**

Адаптивная резонансная теория включает две парадигмы, каждая из которых определяется формой входных данных и способом их обработки. АРТ-1 разработана для обработки двоичных входных векторов, в то время как АРТ-2 может классифицировать как двоичные, так и непрерывные векторы. Рассмотрим подробно сеть АРТ-1, так как несмотря на более простую архитектуру, именно она используется в большинстве практических приложений.

Сеть АРТ-1 обучается без учителя и реализует простой алгоритм кластеризации. В соответствии с этим алгоритмом первый входной сигнал считается образцом первого кластера. Следующий входной сигнал сравнивается с образцом первого кластера. Говорят, что входной сигнал принадлежит первому кластеру, если расстояние до образца первого кластера меньше порога. В противном случае второй входной сигнал - образец второго кластера. Этот процесс повторяется для всех следующих входных сигналов. Таким образом, число кластеров растет с течением времени и зависит как от значения порога, так и от метрики расстояния, использующейся для сравнения входных сигналов и образцов классов.

Сеть АРТ-1 содержит два слоя нейронов. Число нейронов первого слоя совпадает с размерностью входных образов. Число нейронов второго слоя изменяется в процессе настройки сети и совпадает с числом сформированных кластеров.

**Алгоритм функционирования сети АРТ-1**

Шаг 1. Инициализация сети:

где - синаптический вес связи от -го нейрона первого слоя к –му нейрону второго слоя на итерации с номером , синаптический вес связи от -го нейрона второго слоя к -му нейрону первого слоя на итерации с номером . Веса и определяют образец, соответствующий нейрону *.*

Задать - значение порога.

Шаг 2. Предъявление сети нового бинарного входного сигнала *.*

Шаг 3. Вычисление значений соответствия:

Шаг 4. Выбор образца с наибольшим соответствием:

Если , создать новый кластер, соответствующий входному образцу с весами положить и перейти на шаг 8.

Шаг 5. Сравнение с порогом:

Если , перейти к шагу 7.

Шаг 6. Исключение примера с наибольшим значением соответствия. Значение соответствия образца временно устанавливается равным нулю. Переход к шагу 4 (поиск нового значения *).*

Ш а г 7. Адаптация примера с наибольшим значением соответствия:

Шаг 8. Включение всех исключенных на шаге 6 образцов. Положить. Возврат к шагу 2.

Замечание. Порог показывает, насколько должен входной сигнал совпадать с одним их запомненных образцов, чтобы они считались похожими. Близкое к единице значение порога требует почти полного совпадения. При малых значениях порога даже сильно различающиеся входной сигнал и образец считаются принадлежащими одному кластеру.

На шаге 5 вычисляется отношение скалярного произведения входного сигнала и образца с наибольшим значением соответствия к числу единичных бит входного сигнала. Значение отношения сравнивается с порогом, введенном на первом шаге.

Если значение отношения больше порога, то входной сигнал считается похожим на образец с наибольшим значением соответствия. В этом случае образец кластера модифицируется путем выполнения операции AND (логическое «И») с входным вектором.

Если значение отношения меньше порога, то считается, что входной сигнал отличается от данного образца и осуществляется поиск другого похожего вектора. Если входной вектор отличается от всех образцов, то он рассматривается как новый образец. В сеть вводится нейрон, соответствующий новому образцу и рассчитываются значения синаптических весов.