

# **ВИДЫ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ И СПОСОБЫ ОРГАНИЗАЦИИ ИХ ОБУЧЕНИЯ И ФУНКЦИОНИРОВАНИЯ**

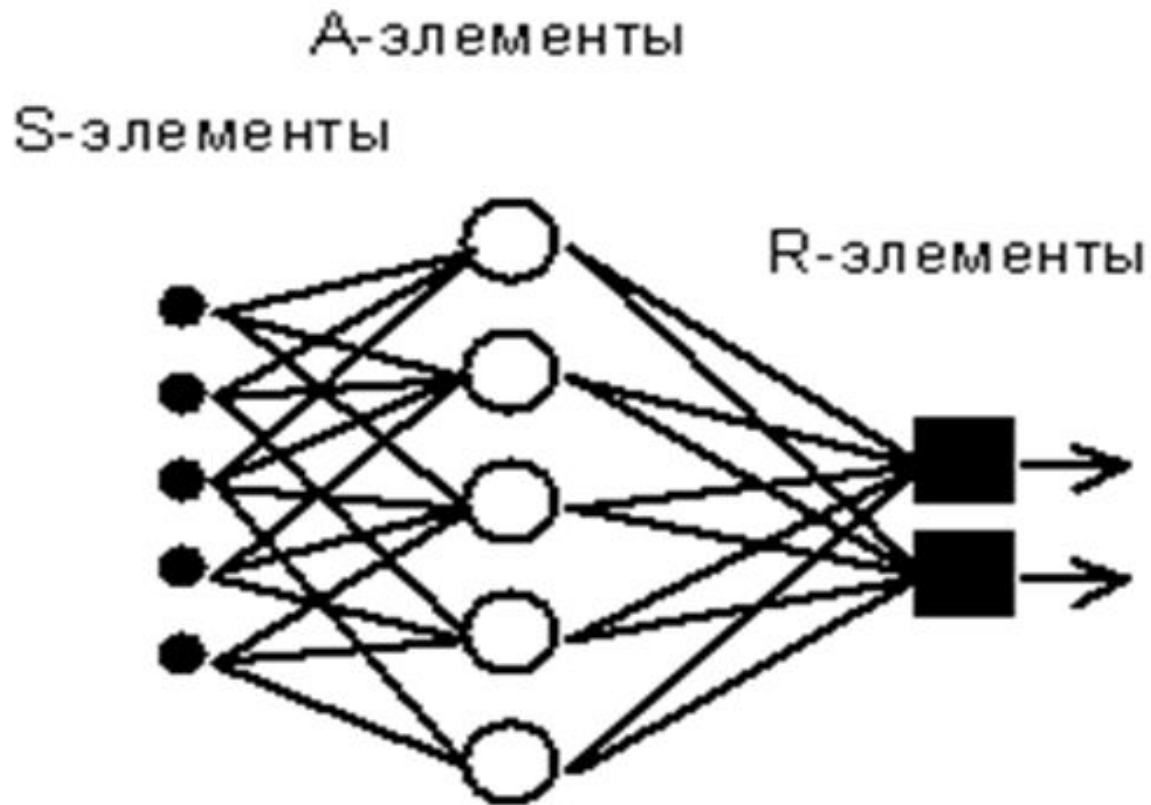
1

# ПЕРСЕПТРОН РОЗЕНБЛАТТА

- Одной из первых искусственных сетей, способных к **перцепции** (восприятию) и формированию реакции на воспринятый стимул, явился **PERCEPTRON** Розенблатта (F.Rosenblatt, 1957).

Персептрон рассматривался его автором не как конкретное техническое вычислительное устройство, а как модель работы мозга.

# ЭЛЕМЕНТАРНЫЙ ПЕРСЕПТРОН РОЗЕНБЛАТТА



## ОБУЧЕНИЕ СЕТИ

- Обучение сети состоит в подстройке весовых коэффициентов каждого нейрона.

Пусть имеется набор пар векторов  $(x^i, y^i)$ ,  $i = 1..p$ , называемый *обучающей выборкой*.

Нейронная сеть называется обученной на данной обучающей выборке, если при подаче на входы сети каждого вектора  $x^i$  на выходах всякий раз получается соответствующий вектор  $y^i$

- Предложенный Ф.Розенблаттом метод обучения состоит в итерационной подстройке матрицы весов, последовательно уменьшающей ошибку в выходных векторах. Алгоритм включает несколько шагов:

**Шаг 0.**

**Шаг 1.**

**Шаг 2.**

Шаг 3.

Шаг 4.

Шаги 1-3 повторяются для всех обучающих векторов.

Используемая на шаге 3 формула

$$W(t + \Delta t) = W(t) + \eta x^\alpha \cdot (\delta^\alpha)^T$$

учитывает следующие обстоятельства:

- а) модифицируются только компоненты матрицы весов, отвечающие ненулевым значениям входов;
- б) знак приращения веса соответствует знаку ошибки, т.е. положительная ошибка ( $\delta > 0$ , значение выхода меньше требуемого) приводит к усилению связи;
- в) обучение каждого нейрона происходит независимо от обучения остальных нейронов, что соответствует важному с биологической точки зрения, принципу *локальности* обучения.

- Данный метод обучения был назван Ф. Розенблаттом “методом коррекции с обратной передачей сигнала ошибки”. Позднее более широко стало известно название “ $\delta$  -правило”.

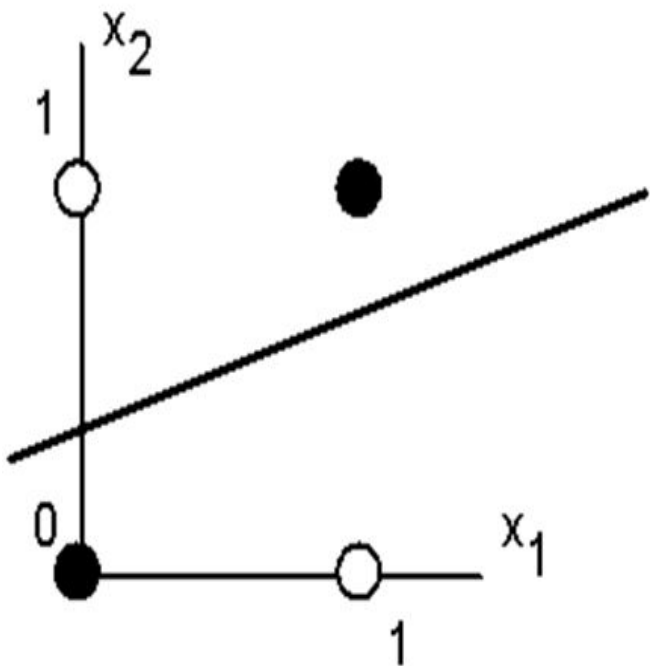
Представленный алгоритм относится к широкому классу алгоритмов обучения *с учителем*, поскольку известны как входные вектора, так и требуемые значения выходных векторов.

- Доказанная Розенблаттом теорема о сходимости обучения по  $\delta$  -правилу говорит о том, что **персептрон способен обучится любому обучающему набору, который он способен представить.**



## ЭЛЕМЕНТАРНЫЙ ПЕРСЕПТРОН РОЗЕНБЛАТТА

- Однако, как было показано позднее (М.Мinsky, S.Papert, 1969), ЭТОТ ВЫВОД ОКАЗАЛСЯ НЕТОЧНЫМ:
- Были выявлены принципиальные неустранимые ограничения однослойных перцептронов, и впоследствии стал в основном рассматриваться **многослойный вариант перцептрона**, в котором имеются несколько слоев процессорных элементов.



Белые точки не могут быть отделены одной прямой от черных

Требуемая активность нейрона для этого рисунка определяется таблицей задания логической функции “исключающее или”.

$X_1$	$X_2$	$Y$
0	0	0
1	0	1
0	1	1
1	1	0

Число переменных N	Полное число возможных логических функций (= $2^{2^N}$ )	Из них линейно разделимых функций
1	4	4
2	16	14
3	256	104
4	65536	1882
5	> 10000000000	94572

# МНОГОСЛОЙНЫЙ ПЕРСЕПТРОН

*Многослойными персептронами называют нейронные сети прямого распространения.*

Входной сигнал в таких сетях распространяется в прямом направлении, от слоя к слою.

Многослойный персептрон в общем представлении состоит из следующих элементов:

- множества входных узлов, которые образуют входной слой;
- одного или нескольких скрытых слоев вычислительных нейронов;
- одного выходного слоя нейронов.

- Основное прикладное значение этого класса сетей состоит в том, что они могут решать задачу аппроксимации многомерных функций, т.е. построения многомерного отображения  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , обобщающего заданный набор примеров  $\{(x^\alpha, y^\alpha)\}$ .

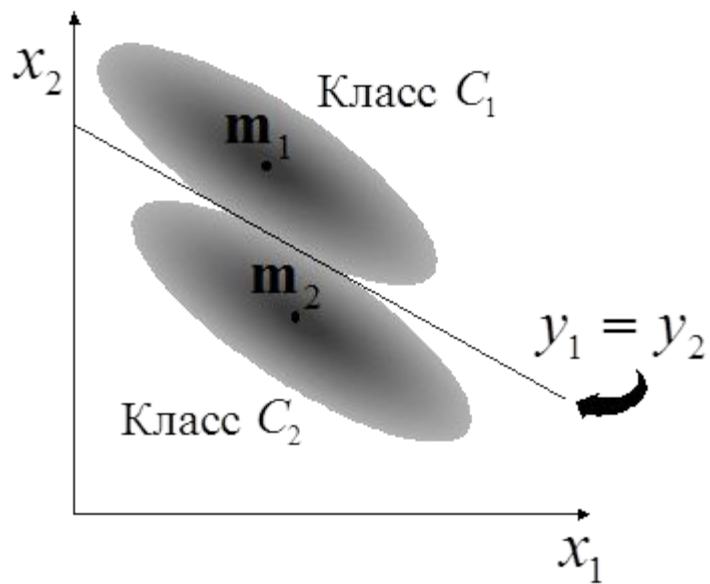
В зависимости от типа выходных переменных, аппроксимация функций может принимать вид классификации или регрессии
- К этой постановке сводятся многие практические задачи распознавания образов, фильтрации шумов, предсказания временных рядов и др.

## ДИСКРИМИНАТОР

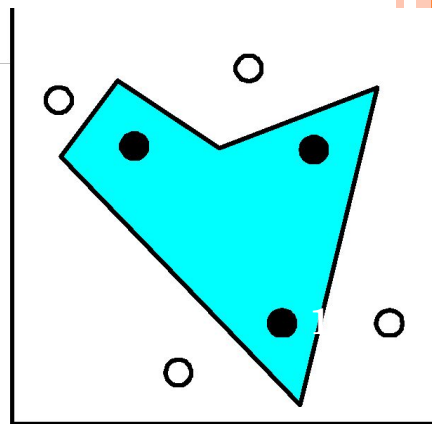
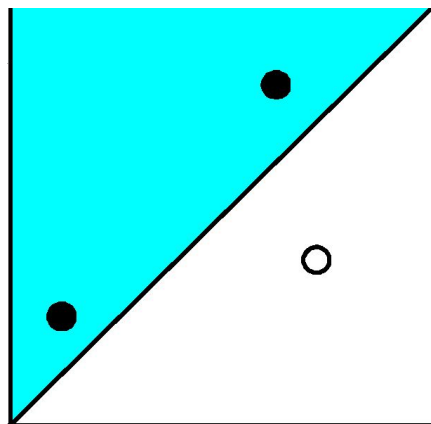
- Скалярный выход нейрона можно использовать в качестве **индикатор принадлежности входного вектора к одному из заданных классов** (т.н. *дискриминантной функции*).

Поскольку дискриминантная функция зависит лишь от линейной комбинации входов, нейрон является **линейным дискриминатором**.

# Нейрон как линейный дискриминатор



Линейно-разделимые и линейно-неразделимые классы



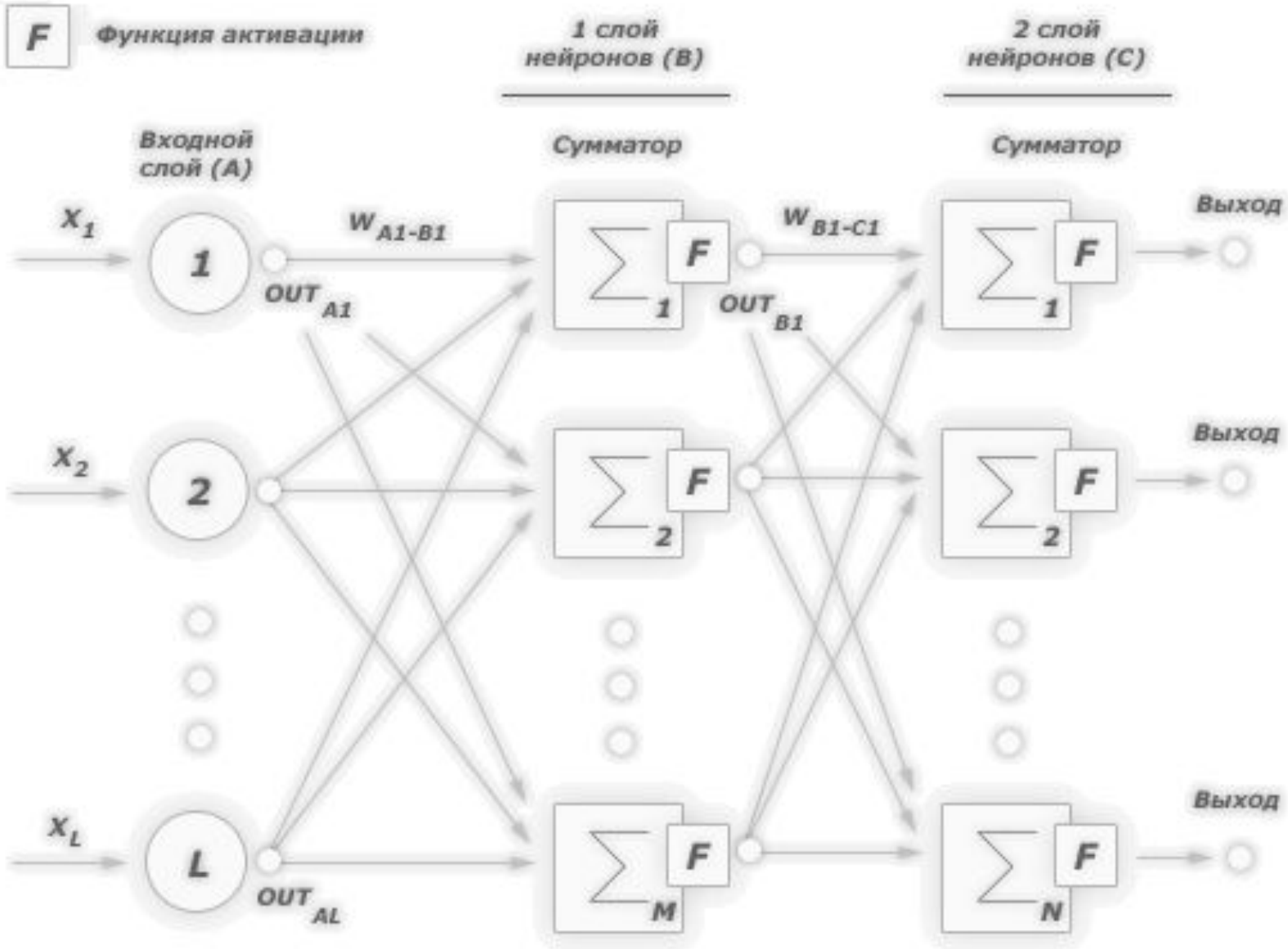
- Определен следующий важный результат в этой области: *одного скрытого слоя нейронов с сигмоидальной функцией активации достаточно для аппроксимации любой границы между классами ( или любой функции) со сколь угодно высокой точностью.*

Точность аппроксимации возрастает с числом нейронов скрытого слоя: при  $N$  нейронах ошибка оценивается как  $f(1/N)$ .

- Комарцова Л.Г., Максимов А.В. Нейрокомпьютеры: Учеб. пособие для вузов.- М.: Изд-во МГТУ им. Н.Э. Баумана, 2002 (Сер. Информатика в техническом университете).



**F** Функция активации



Пример двухслойного персептрона

# МНОГОСЛОЙНЫЙ ПЕРСЕПТРОН

При построении сети на основе нескольких слоев персептронов учитывается следующее:

- Количество входных и выходных элементов в многослойном персептоне определяется условиями задачи.
- Вопрос о том, сколько использовать промежуточных слоев и элементов в них, пока строго математически не определен.
- В качестве начального приближения можно взять один промежуточный слой, а число элементов в нем положить равным полусумме числа входных и выходных элементов.

# МНОГОСЛОЙНЫЕ ПЕРСЕПТРОНЫ

Многослойные персептроны успешно применяются для решения разнообразных сложных задач и имеют три следующих отличительных свойства:

- **Свойство 1. Каждый нейрон сети имеет нелинейную функцию активации**
- **Свойство 2. Несколько скрытых слоев**
- **Свойство 3. Высокая связность**

## МЕТОД ОБРАТНОГО РАСПРОСТРАНЕНИЯ ОШИБКИ

- Исторически наибольшую трудность на пути к эффективному правилу обучения многослойных персептронов вызывала процедура эффективного расчета градиента функции ошибки.

Ошибка сети определяется по ее выходам, поэтому нужна была процедура передачи ошибки с выходного слоя к предшествующим слоям сети, в направлении *обратном обработке входной информации*.

Поэтому такой метод, когда он был найден, получил название **метода обратного распространения ошибки** (back propagation error)(ОРО).

# САМООБУЧАЮЩИЕСЯ НЕЙРОСЕТИ

Самообучающиеся нейросети - это класс сетей, в котором обучение происходит без учителя, т.е. реализуется процесс самообучения: сеть самостоятельно формирует свои выходы, адаптируясь к поступающим на ее входы сигналам на основе минимизации некоторого целевого функционала.

В отсутствие внешней цели, "учителем" сети могут служить лишь сами данные, т.е. имеющаяся в них информация, закономерности, отличающие входные данные от случайного шума.

# САМООБУЧАЮЩИЕСЯ НЕЙРОСЕТИ

Длина описания данных пропорциональна **разрядности данных** (числу бит), определяющей возможное разнообразие принимаемых ими значений, и **размерности данных**, т.е. числу компонент входных векторов.

Соответственно, можно различить два типа кодирования, использующих противоположные способы сжатия информации:

# САМООБУЧАЮЩИЕСЯ НЕЙРОСЕТИ

- понижение размерности данных с минимальной потерей информации (анализ главных компонент данных, выделение наборов независимых признаков);
- уменьшение разнообразия данных за счет выделения конечного набора классов и отнесения данных к одному из них (кластеризация данных, квантование непрерывной входной информации).

## ПРАВИЛО ОЙА МИНИМИЗАЦИИ ОШИБКИ ОБУЧЕНИЯ СЕТИ

- Поскольку учитель в таких сетях отсутствует, то ошибка работы сети может достигать больших значений, поэтому применяют ограничение к правилу Хебба, состоящее в максимизации чувствительности выхода нейрона при ограниченной амплитуде весов: веса обученного нейрона расположены на гиперсфере:  $|W| = 1$ .



# СОРЕВНОВАТЕЛЬНОЕ ОБУЧЕНИЕ

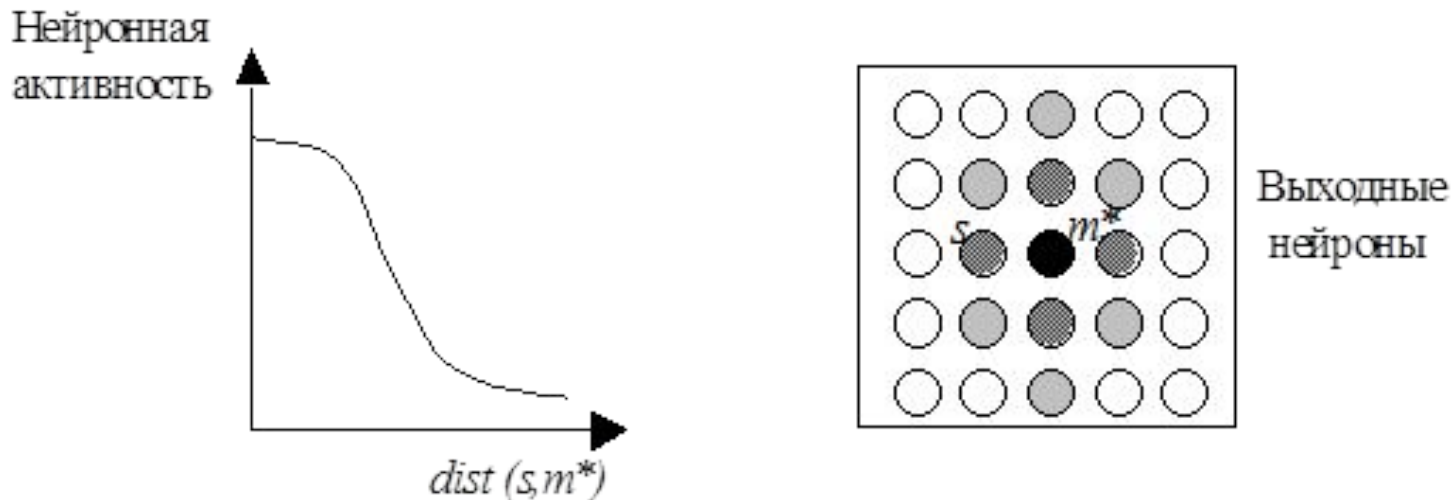
- Соревновательное обучение состоит в том, чтобы каждый нейрон был обучен усиливать свой выход и подавлять активность остальных.
- При логистической функции активации, препятствующей бесконечному росту, победителем в этой борьбе выйдет нейрон с максимальным первоначальным значением выхода. Его значение возрастет до единицы, а активность остальных нейронов затухнет до нуля.
- Такие соревновательные слои нейронов также можно использовать для сжатия информации, при этом сжатие будет основано на принципе уменьшения разнообразия входных данных при той же их размерности.

# АЛГОРИМ «ПОБЕДИТЕЛЬ ПОЛУЧАЕТ ВСЕ» (WTA)

Разновидность соревновательного алгоритма:

- Веса латеральных связей – т.е. связей в одном (скрытом) нейронном слое – нормируются таким образом, что суммарный сигнал является усиливающим только для нейрона с максимальной исходной активностью. Остальные нейроны испытывают торможение.
- По выполнении некоторого числа итераций  $t$  для всех нейронов, кроме одного, значение корректирующей функции становится отрицательным, что обращает их активность в нуль. Единственный, оставшийся активным, нейрон является победителем. *Он и указывает на тот класс, к которому принадлежит введенный образ.*

# Распределение нейронной активности "соседей" нейрона – победителя:



$dist(s, m^*)$  - расстояние между выигравшим нейроном  $m^*$  и его "соседом"  $s$

Нейросетевая парадигма Липпмана-Хемминга, реализующая механизм WTA для решения задачи классификации данных, является моделью с прямой структурой памяти.

**Информация**, содержащаяся в исходных образах никак **не обобщается**, а непосредственно запоминается в синаптических связях.

**Память** здесь **не является распределенной**, т. е. при выходе из строя одного нейрона полностью теряется информация обо всем соответствующем ему образе памяти.

# КАРТА САМООРГАНИЗАЦИИ КОХОНЕНА (САМООРГАНИЗУЮЩАЯСЯ СЕТЬ КОХОНЕНА)

В противоположность хемминговой сети модель Кохонена (Т. Kohonen, 1982) выполняет обобщение предъявляемой информации.

В результате работы НС Кохонена получается образ, представляющий собой карту распределения векторов из обучающей выборки.

Таким образом, в сети Кохонена выполняется решение задачи нахождения кластеров в пространстве входных образов.

# САМООРГАНИЗУЮЩАЯСЯ СЕТЬ КОХОНЕНА

Данная сеть обучается без учителя на основе самоорганизации.

По мере обучения вектора весов нейронов стремятся к центрам кластеров - групп векторов обучающей выборки.

На этапе решения информационных задач сеть относит новый предъявленный образ к одному из сформированных кластеров, указывая тем самым категорию, к которой он принадлежит. При этом результат представляется в виде карты нейронной активности.

## САМООРГАНИЗУЮЩАЯСЯ СЕТЬ КОХОНЕНА

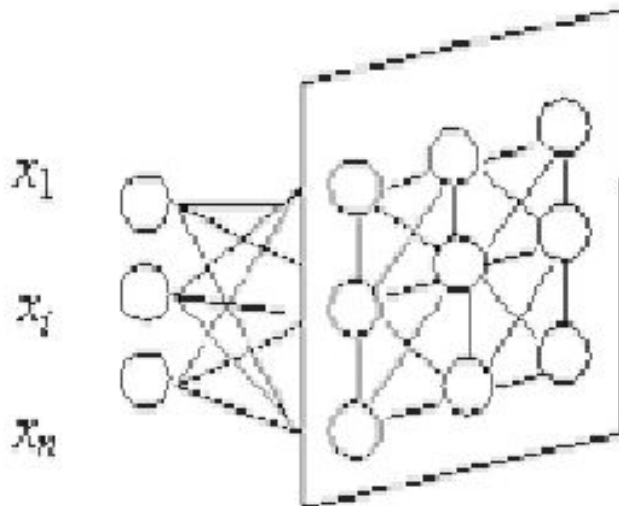
Сеть Кохонена содержит один скрытый слой нейронов. Число входов каждого нейрона равно размерности входного образа. Количество же нейронов в скрытом слое определяется той степенью подробности, с которой требуется выполнить кластеризацию набора исходных образов.

При достаточном количестве нейронов и удачных параметрах обучения НС Кохонена может не только выделить основные группы образов, но и установить "тонкую структуру" полученных кластеров.

При этом близким входным образам будут соответствовать близкие карты нейронной активности.

Нейронная сеть Кохонена (Kohonen clustering network – KСN) характеризуется следующими свойствами:

- сеть состоит из одного скрытого слоя нейронов (всего слоев 2);
- число входов каждого нейрона равно размерности входного образа ;
- количество нейронов в слое непосредственно определяет, сколько различных кластеров  $k$  сеть может распознать.





## МОДЕЛЬ ХОПФИЛДА (J.J.HOPFIELD, 1982)

Модель Хопфилда занимает особое место в ряду нейросетевых моделей:

- В ней впервые удалось установить связь между нелинейными динамическими системами и нейронными сетями.
- Образы памяти сети соответствуют устойчивым предельным точкам (аттракторам) динамической системы.
- Особенно важной оказалась возможность переноса математического аппарата теории нелинейных динамических систем (и статистической физики вообще) на нейронные сети.

При этом появилась возможность теоретически оценить объем памяти сети Хопфилда, определить область параметров сети, в которой достигается наилучшее функционирование.

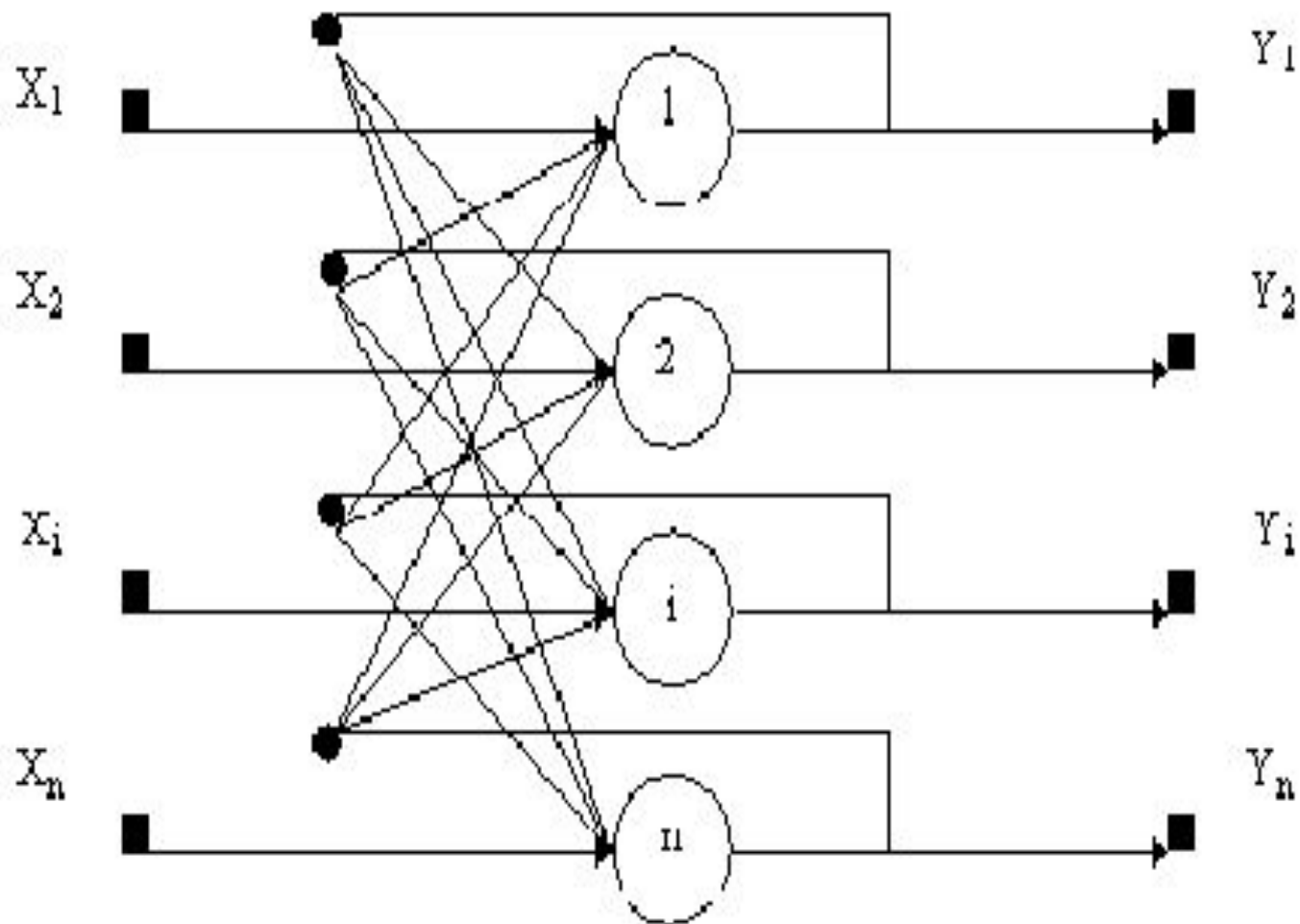
# РЕКУРЕНТНЫЕ СЕТИ ХОПФИЛДА

Сеть Хопфилда получается, если наложить на веса связей в выражении  $s_j = \sum w_{ij} x_i + w_{0j}$  следующие условия:

- 1) все элементы связаны со всеми;
- 2)  $w_{ij} = w_{ji}$  - прямые и обратные связи симметричны;
- 3)  $w_{ii} = 0$  - диагональные элементы матрицы связей равны нулю.

Последнее условие обычно добавляется, чтобы исключить непосредственную обратную связь с выхода нейрона на собственный вход.

# КОНФИГУРАЦИЯ СЕТИ ХОПФИЛДА

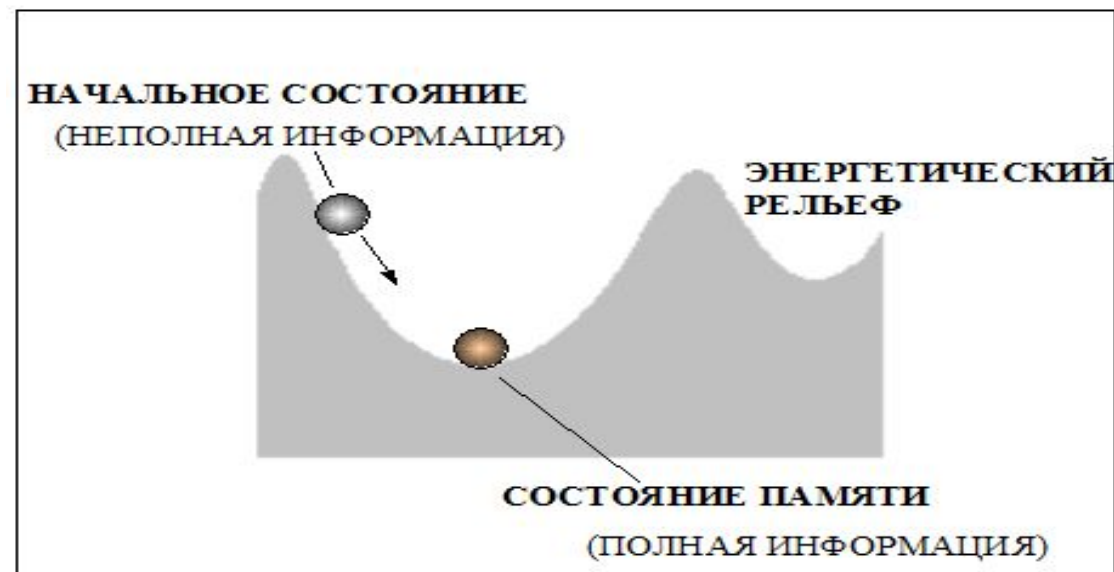


Одно из достоинств симметричной квадратной матрицы связей, характерной для сети Хопфилда, состоит в том, что поведение сети можно описать через стремление к минимуму простой целевой функции

$$\square E = - \sum w_{ij} x_i x_j = \min, \quad i \neq j$$

Обычно  $E$  интерпретируется как некоторая "обобщенная энергия". Такая интерпретация берет начало от *физической модели Изинга*, в которой совокупность взаимодействующих магнитных диполей (спинов) стремится принять такую конфигурацию, в которой суммарная энергия будет минимальна.

- Поведение системы в пространстве состояний напоминает движение шарика, который стремится скатиться в точку минимума некоторого потенциального рельефа со множеством локальных минимумов. Эти минимумы будут устойчивыми состояниями памяти, а окружающие точки на склонах - переходными состояниями. Начальное состояние шарика соответствует вектору, содержащему неполную информацию об образе памяти, которому отвечает дно лунки.



- Характер рельефа определяется видом целевой функции  $E$  и формируется в процессе обучения сети.
- Обучение производится путем демонстрации эталонных образов, которые сеть должна запомнить, хранить и потом воспроизводить (узнавать).
- Алгоритм обучения (формирование весовых коэффициентов  $w_{ij}$ ) основывается на правиле Хебба.

- Главное свойство сети Хопфилда - способность восстанавливать возмущенное состояние равновесия - "вспоминать" искаженные или потерянные биты информации.
- Восстановление полной информации по какой-либо ее части - вспоминание по ассоциации - наделяет модель Хопфилда свойством ассоциативной памяти.

- Одна и та же сеть с одними и теми же весами связей может хранить и воспроизводить несколько различных эталонов.
- Каждый эталон является аттрактором (стационарным оптимальным состоянием), вокруг которого существует область притяжения (от англ. attract – притягивать)
- Любая система с несколькими аттракторами, к которым она тяготеет, может рассматриваться как содержательно-адресуемая память, т.е. память, из которой информация об эталоне извлекается путем задания нескольких признаков эталона.



Ассоциативный характер памяти сети Хопфилда качественно отличает ее от обычной, адресной, компьютерной памяти.

При использовании ассоциативной памяти доступ к информации производится непосредственно по ее *содержанию*, т.е. по частично известным искаженным фрагментам.

Важным свойством памяти, представленной набором аттракторов сети, является ее *распределенность*. Это означает, что все нейроны сети участвуют в кодировании всех состояний памяти. Поэтому небольшие искажения значений отдельных весов не сказываются на содержании памяти, что повышает устойчивость памяти к помехам.

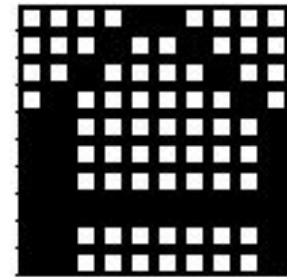
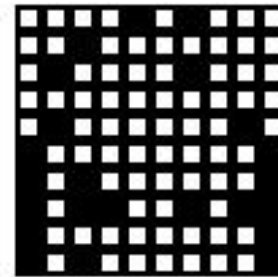
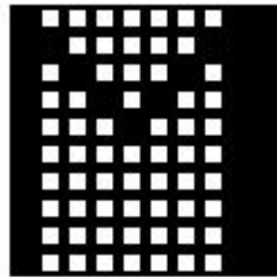
- Преобразование информации рекуррентными нейронными сетями типа сети Хопфилда, минимизирующими энергию, может приводить к появлению в их пространстве состояний аттракторов, далеких по форме от образов внешнего для сети окружения.
- Таким образом, появляется возможность использовать рекуррентные сети для активной кластеризации, при которой сеть "творчески" относится к входным векторам, осуществляя нетривиальные обобщения поступающих на ее вход сигналов
- Итак, минимизирующие энергию нейронные сети типа сети Хопфилда могут использоваться для предсказания существования новых классов объектов.

Хотя сети Хопфилда получили применение на практике, им свойственны определенные недостатки, ограничивающие возможности их применения:

- модель Хопфилда предполагает симметрию связей между элементами; без этого условия понятие энергии не может быть введено и эта простая физическая метафора, которой модель во многом обязана своими достижениями, перестает работать;
- условность понятия энергии, которая может исказить суть происходящих процессов.

Представляется, что место энергии как целевой функции нейросети должна занять информация

- Нейросетевая система, построенная по модели Хопфилда, способна распознавать практически полностью зашумленные образы. Практическое решение задач демонстрирует замечательное свойство сети Хопфилда ассоциативно узнавать образ по его небольшому фрагменту.



**А**

**Б**

Образы с информационным шумом удачно распознаются.