



# Искусственные нейронные сети

# 1. Формальный нейрон

Большинство моделей основывается на схеме формального нейрона У.С. Мак-Каллока и У.Питтса (1943 год), согласно которой нейрон представляет собой пороговый элемент (Рис.1). На входах нейрона имеются возбуждающие и тормозящие синапсы, в нейроне определяется взвешенная сумма (с учетом весов синапсов) входных сигналов, при превышении этой суммой порога нейрона вырабатывается выходной сигнал.

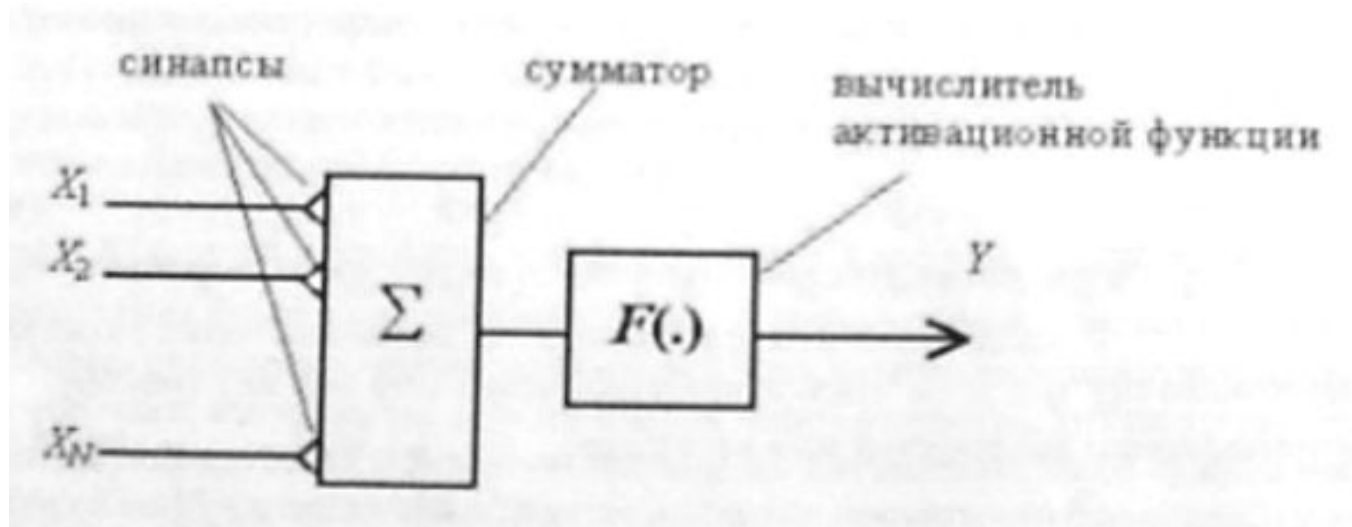


Рис. 1. Схема формального нейрона.

$X_i$ - входные сигналы,  $Y$  - выходной сигнал нейрона.

Работа формального нейрона (Рис.1) может быть описана уравнениями:

$$Y_j = F(net_j - K_j), \quad (1)$$

$$net_j = \sum_i w_{ji} X_i, \quad (2)$$

где  $j$  – номер нейрона в сети,  $X_i$  – входные сигналы,  $Y_j$  – выходной сигнал нейрона,  $w_{ji}$  – веса синапсов,  $net_j$  – суммарное входное воздействие на нейрон,  $K_j$  – порог нейрона,  $F(.)$  – активационная функция.

Активационная функция характеризует реакцию нейрона на входное воздействие  $net_j$ , она может быть пороговой:

$$F(a) = \begin{cases} 0, & \text{при } a < 0, \\ 1, & \text{при } a > 0, \end{cases} \quad (3)$$

или некоторой непрерывной, например, линейной:  $F(a) = ka$  (3a)

или логической:  $F(a) = 1/[1 + \exp(-a)]$  (3b)

**В зависимости от реализуемого алгоритма на допустимые значения входов и выходов нейрона накладываются определенные ограничения: значения  $X_i$  и  $Y_j$  могут бинарными (т.е. равными 0 или 1), бинарными биполярными (+1 или -1), принадлежащими интервалу (0,1), неотрицательными или действительными. Аналогичные ограничения накладываются на веса синапсов нейронов  $w_{ji}$ .**

## 2. Перцептон

Перцептон состоит из элементов 3-х типов:  $S$  - элементов,  $A$  - элементов и  $R$  - элемента (Рис.2).  $S$  - элементы это - слой рецепторов. Эти рецепторы соединены с,  $A$  - элементами с помощью тормозных или возбуждающих связей. Каждый рецептор может находиться в одном из двух состояний - покоя или возбуждения.  $A$  - элементы представляют собой сумматоры с порогом (т.е. формальные нейроны). Это означаем, что  $A$  - элемент возбуждается, если алгебраическая сумма возбуждений, приходящих к нему от рецепторов, превышает определенную величину - его порог. При этом сигнал от рецептора, приходящий по возбуждающей связи, считается положительным, а приходящий по тормозной связи отрицательным. Сигналы от возбуждившихся  $A$  - элементов передаются в сумматор  $R$ , причем сигнал от  $i$ -го ассоциативного элемента передастся с коэффициентом  $k_i$ .

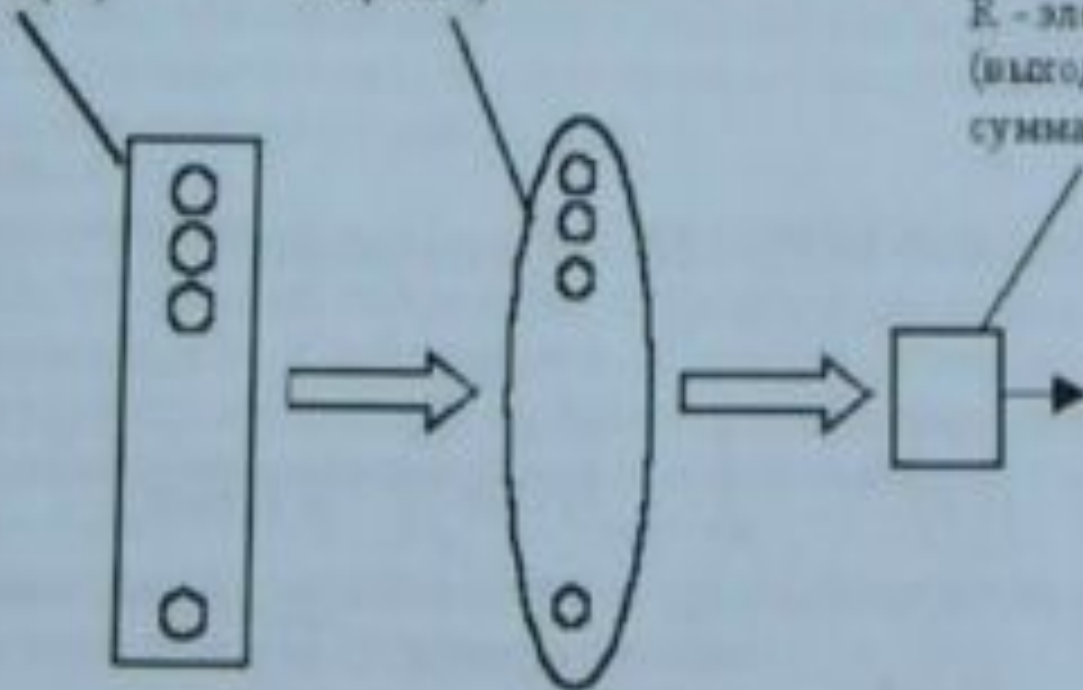
$S$  - элементы  
(рецепторы)

$A$  - элементы (слой  
ассоциативных  
нейронов)

$E$  - элемент  
(выходной  
сумматор)

Выход  
перцептрона

Рис. 2. Схема перцептрона.



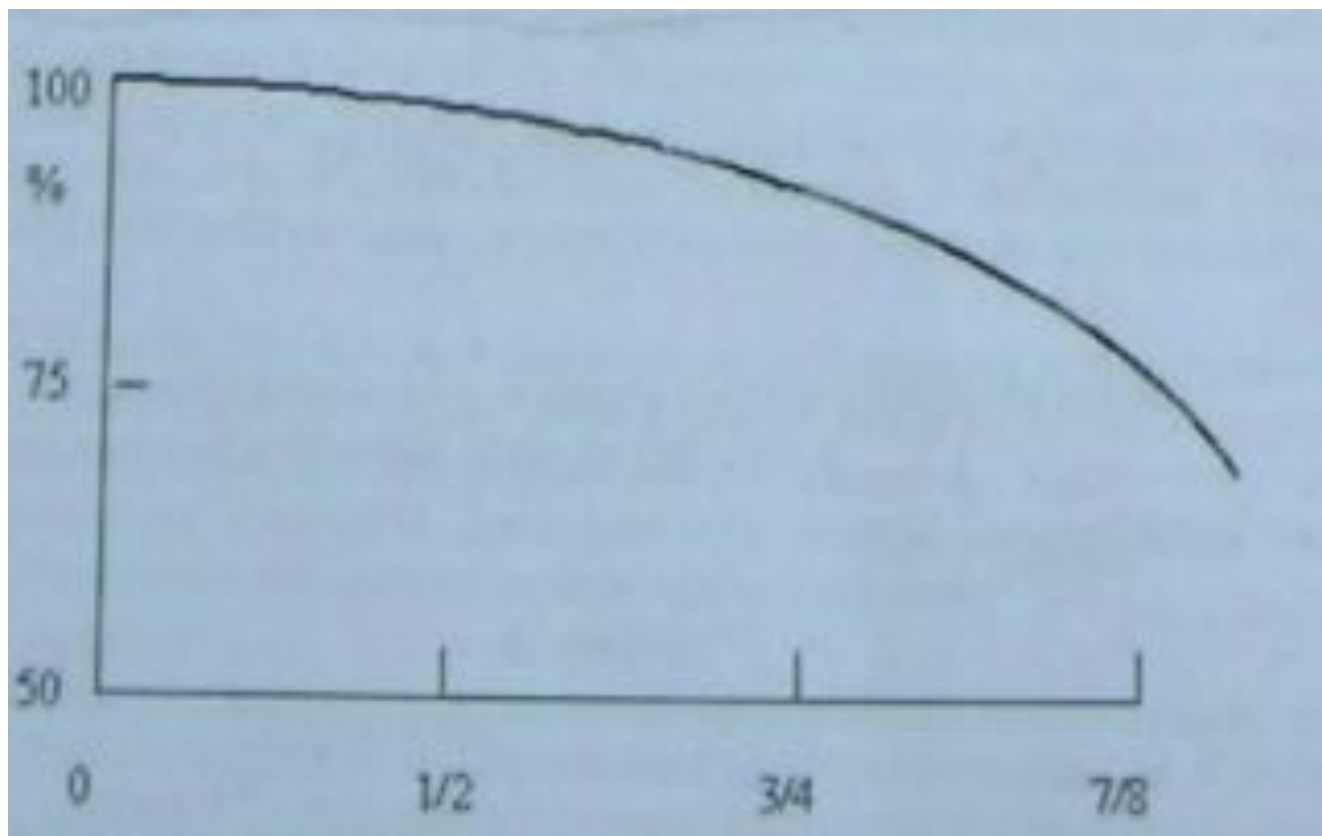
Система связей между рецепторами  $S$  и  $A$  - элементами, так же как и пороги  $A$  - элементов выбираются некоторым случайным, но фиксированным образом, а обучение состоит лишь в изменении коэффициентов  $k_i$ . Считаем, что мы хотим научить перцептрон разделять два класса объектов, и потребуем, чтобы при предъявлении объектов первого класса выход перцептрона был положителен, а при предъявлении объектов второго класса отрицательным. Начальные коэффициенты  $k_i$  полагаем равными нулю. Далее предъявляем обучающую выборку: объекты (например, круги либо квадраты) с указанием класса, к которым они принадлежат. Показываем перцептрону объект первого класса. При этом некоторые  $A$  - элементы возбуждятся. Коэффициенты  $k_i$ , соответствующие этим возбужденным элементам, увеличиваем на 1. Затем предъявляем объект второго класса и коэффициенты  $k_i$  тех  $A$  - элементов, которые возбуждятся при этом показе, уменьшаем на 1. Этот процесс продолжим для всей обучающей выборки. В результате обучения сформируются значения весов связей  $k_i$ .

После обучения перцептрон готов работать в режиме распознавания. В этом режиме перцептрону предъявляются "не знакомые" перцептрону объекты, и перцептрон должен установить, к какому классу они принадлежат. Работа перцептрона состоит в следующем: при предъявлении объекта возбужденные  $A$  - элементы передают сигнал  $R$  - элементу, равный сумме соответствующих коэффициентов  $ki$ . Если эта сумма положительна, то принимается решение, что данный объект принадлежит к первому классу, а если она отрицательна - то второму.

Исследования перцептронов показали, что перцептроны способны обучаться, хотя способности их обучения довольно ограничены. Справедлива теорема о сходимости перцептрона, согласно которой независимо от начальных значений коэффициентов и порядка показа образцов при обучении перцептрон за конечное число шагов научится различать два класса объектов, если только существуют такие значения. Подчеркнем, что теорема ничего не говорит о том, какие классы могут быть разделены.

Исследования также показали, что слабые стороны перцептрона (в частности большое время обучения) в значительной степени связаны со случайностью связей между его элементами. Однако эта конструктивная особенность обеспечивает перцептрону и положительное качество — надежность: выход из строя заметного числа элементов перцептрона слабо сказывается на качестве его работы (Рис.3).





**Рис.3. Повеление перцептрона при выходе из строя ассоциативных элементов. По оси ординат - процент правильных ответов, по оси абсцисс - доля выключенных ассоциативных элементов. Схематично.**

### 3. Ассоциативная память

Различают автоассоциативную и гетероассоциативную память. В обоих случаях имеется режим записи и режим воспроизведения.

В случае *автоассоциативной памяти* при записи происходит запоминание набора эталонных образов, задаваемых векторами  $X_k$ ,  $k=1,2,\dots,n$ ; а при воспроизведении по искаженному образу  $X_j + \delta X$  ( $X_j$  - один из эталонов,  $\delta X$  - вектор, характеризующий искажение) восстанавливается эталон  $X_j$ .

В случае *гетероассоциативной памяти* нейронная сеть при записи запоминает отображение  $X_k \rightarrow Y_k$  между векторами  $X_k$  и  $Y_k$ ,  $k=1,2,\dots,n$ ; при воспроизведении предъявляется один из эталонных векторов  $X_j$  (возможно несколько искаженный), и с помощью нейронной сети восстанавливается парный ему вектор  $Y_j$ .

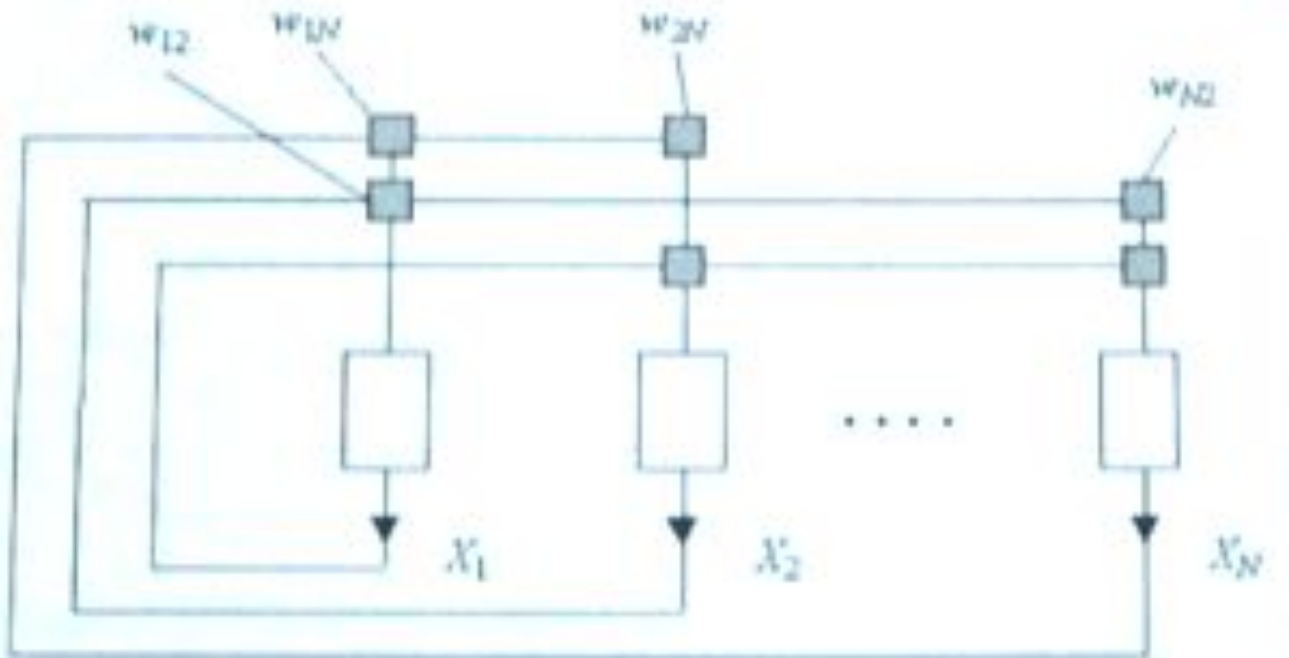
## 4. Сеть Хопфилда

Допустим, что имеется некоторая физическая система, имеющая множество устойчивых особых точек:  $X_1, X_2, \dots$ . Эти точки могут рассматриваться как память устройства, реализуемого физической системой. Начиная с некоторой точки  $X_j + \delta X$  ( $X_j$  - один из эталонов,  $\delta X$  - вектор, характеризующий искажение), при малом  $\delta X$  мы придем к  $X_j$ . Это вариант автоассоциативной памяти, в которой запомненный образ восстанавливается по искаженному образу.

Удобно построить физическую систему, имеющую локальные минимумы энергии, соответствующие устойчивым особым точкам. Для того чтобы систему можно было бы рассматривать как память, нужно иметь большое число локальных минимумов. Пример физической системы с большим числом локальных минимумов энергии — спиновые стекла, т.е. система спинов, матрица обменных взаимодействий между которыми стохастична. В [7] Хопфилд строит модель нейроподобной ассоциативной памяти, допускающую проведение явных аналогий со спиновыми стеклами. Опишем в общих чертах эту модель

# 4.1. Схема нейросети

Схема нейросети по Хопфилду показана на Рис.4. Сеть состоит из  $N$  нейронов, все нейроны связаны со всеми, каждая связь характеризуется своим весом  $w_{ji}$ , матрица весов предполагается симметричной:  $w_{ji} = w_{ij}$ . Считаем, что  $N \gg$



**Рис.4. Схема нейронной сети по Хофилду** □ - нейроны, ■ - синапсы

**Состояние нейронной сети характеризуется вектором  $X = X_1, X_2, \dots, X_N,$**

**Каждый нейрон может находиться в двух состояниях  $X_i=0$  - покой,  $X_i=1$  - возбужденное состояние,  $i$  - номер нейрона.**

## 4.2. Режим записи и режим воспроизведения

В режиме записи формируется матрица связей между нейронами  $w_{ij}$ , равная

$$w_{ij} = \sum_k (2X_{ki} - 1)(2X_{kj} - 1), \quad k=1, 2, \dots, \quad (4)$$

при  $i \neq j$ ,  $w_{ij} = 0$  - запоминаемые бинарные векторы,  $n$  - число запоминаемых паттернов. Эталоны предполагаются случайными, компоненты векторов  $X_{ki}$  компоненты выбираются равными 0 либо 1 с равной вероятностью:  $p\{X_{ki}=0\} = p\{X_{ki}=1\} = 0,5$ . Отметим, что  $(2X_{ki} - 1) = +1$  либо  $-1$ .

В процессе воспроизведения нейронная сеть функционирует в дискретном времени. Работа нейронной сети происходит в так называемом ассинхронном режиме, т. е. в один такт времени только один нейрон может изменить свое состояние. Нейрон, которому разрешено менять состояние, выбирается стохастически. При этом нейроны работают как нейроны Мак-Каллока и Питтса с пороговой активационной функцией. А именно, полагаем:

$$X_j(t+1)=1 \text{ при } net_j > 0 \text{ и } X_j(t+1)=0 \text{ при } net_j < 0 \quad (5)$$

$$net_j = \sum_i w_{ji} X_i, \quad (6)$$

где  $t$  – момент "пересмотра" состояния  $j$ -го нейрона

Работу нейронной сети в режиме воспроизведения можно рассматривать как динамическую систему. Покажем, что если число эталонов не слишком высоко, то эталоны соответствуют устойчивым состояниям (особым точкам) рассматриваемой динамической системы.

Оценим величину

$$nets_j = \sum_i w_{ji} X_{si}(t), \quad (7)$$

где  $X_{si}(t)$ -компоненты какого-либо эталонного вектора. С учетом (4) имеем:

$$nets_j = \sum_k (2X_{ki}-1) [\sum_i (2X_{kj}-1) X_{si}(t)]. \quad (8)$$

В силу случайности эталонов среднее значение величины в квадратных скобках равно 0 при  $s \neq k$  и равно  $N/2$  при  $s=k$ . Следовательно, в силу такой псевдоортогональности имеем:

$$nets_j \approx (2X_{sj}-1)N/2 \quad (9)$$

Итак, имеем:  $nets_j > 0$  при  $X_{sj}=1$  и  $nets_j < 0$  при  $X_{sj}=0$ .

Следовательно, в пренебрежении шумом, возникающим от случайных флуктуации, запомненные эталоны устойчивы.



## 4.3 Минимизация энергии нейронной сети

Определим "энергию" нейронной сети как

$$E = -0,5 \sum_{ij} w_{ji} X_i X_j, \quad i \neq j \quad (10)$$

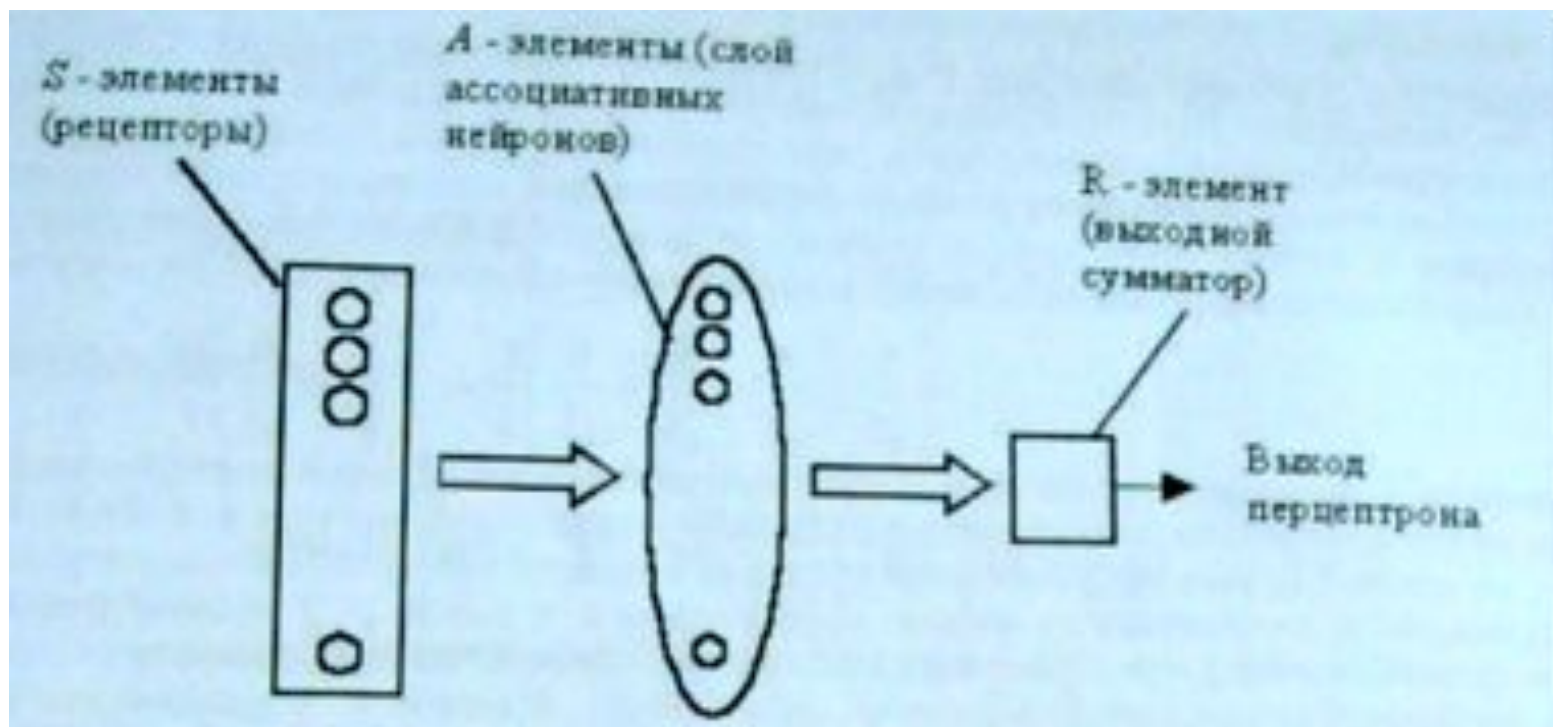
Так как  $w_{ji} = w_{ij}$ , то изменение энергии, обусловленное текущим изменением состояния нейрона, равно

$$\Delta E = -\Delta X_j \sum_{ij} w_{ji} X_i = -\Delta X_j \text{net}_j, \quad i \neq j \quad (11)$$

В силу (5)  $\Delta E \leq 0$ . Таким образом, алгоритм изменения состояний нейронов приводит к монотонному уменьшению энергии. Состояния изменяются до тех пор, пока не будет достигнут локальный минимум энергии. Энергия играет роль функции Ляпунова для рассматриваемой динамической системы. Из (9), (11) следует, что эталоны соответствуют минимумам энергии  $E$ .

## **5. Метод обратного распространения ошибок**

**Предполагается, что нейронная сеть имеет многослойную структуру (Рис.5). Нейроны сети представляют собой формальные нейроны с логической активационной функцией (см. формулу (3b)). Сеть не имеет обратных связей: при вычислении выхода нейронной сети по ее входу сигналы нейронов передаются от слоя к слою, слева направо. Схема такой сети напоминает перцептрон (Рис.2). Иногда ее так и называют "обобщенный перцептрон".**



**Рис.5. Схема нейронной сети, используемой в методе обратного распространения ошибок. Стрелками показано направление движения сигналов в процессе расчета выхода нейронной сети. При коррекции весов синапсов сигналы перемешаются в *обратом* направлении.**

Нейронная сеть фактически реализует функцию гетероассоциативной памяти. Имеется режим *обучения* и режим воспроизведения. В режиме обучения нейронная сеть запоминает отображение  $X_k \rightarrow Y_k$  между векторами  $X_k$  и  $Y_k$ ,  $k=1,2,\dots,n$ ; при воспроизведении предъявляется один из эталонных векторов  $X_j$  (возможно несколько искаженный), и с помощью нейронной сети восстанавливается парный ему вектор  $Y_j$ .

Режим воспроизведения фактически тривиален: на вход нейронной сети поступают входные векторы  $X$  и в процессе функционирования формальных нейронов вычисляется выходной вектор  $Y$

Обучение многослойной нейронной сети осуществляется путем оптимизации весов синапсов методом градиентного спуска. Это обучение может быть представлено как последовательность следующих операций:

1. Выбрать очередную пару векторов  $X_k$  и  $Y_k$  из обучающей выборки.
2. Вычислить выход сети  $Y$ .
3. Вычислить разность между выходом сети  $Y$  и требуемым выходным вектором  $Y_k$  (целевым вектором обучающей пары). Т.е. определить ошибку нейронной сети.

4. Подкорректировать веса сети  $w_{ji}$  так, что бы минимизировать ошибку.
5. Повторять шаги с 1 по 4 для каждой пары обучающей выборки до тех пор, пока ошибка не достигнет приемлемого уровня.

Все вычисления производятся послойно: в процессе расчета выхода нейронной сети по входу расчет идет слева направо, при корректировке весов рассчитываются всех нейронов и по этим ошибкам корректируются веса нейронов, при этом ошибки как бы "распространяются" в обратном направлении - справа налево. Процедура коррекции весов описана в многочисленной литературе. Здесь мы не будем воспроизводить ее, отметим только, что фактически коррекция весов означает минимизацию ошибки нейронной сети, а поиск минимума ошибки производится методом градиентного спуска.

Таким образом, в настоящее время активно продолжаются исследования нейронных сетей. Отметим, что, несмотря на все экономические трудности, работы по нейронным сетям в нашей стране активно продолжаются. В Москве уже более 20 лет под руководством В.Л. Дунина-Барковского ежемесячно проводится семинар по нейронным сетям. На базе этого семинара сформирована Российская ассоциация нейроинформатикн, объединяющая энтузиастов-нейросетевиков. Ассоциация имеет широкие международные связи, под эгидой ассоциации проведен ряд конференций, симпозиумов, выставок, совещаний.

Необходимо подчеркнуть, что современные исследования нейронных сетей характеризуются внедрением в нейроинформатику высокоэффективных математических методов, заимствованных из статистической физики, синергетики, теории вероятностей, дифференциальной геометрии.

**Необходимо максимально полно понимать работу биологических систем обработки информации и свойств организмов, обеспечиваемых этими системами. Одним из важных направлений исследований, способствующих такому пониманию, может быть анализ того, как в процессе биологической эволюции возникали "интеллектуальные" свойства биологических организмов. По-видимому, наиболее интересные исследования, направленные на изучение интеллектуальных свойств биологических организмов, ведутся сейчас в направлениях исследований "Искусственная жизнь" и "Адаптивное поведение"**