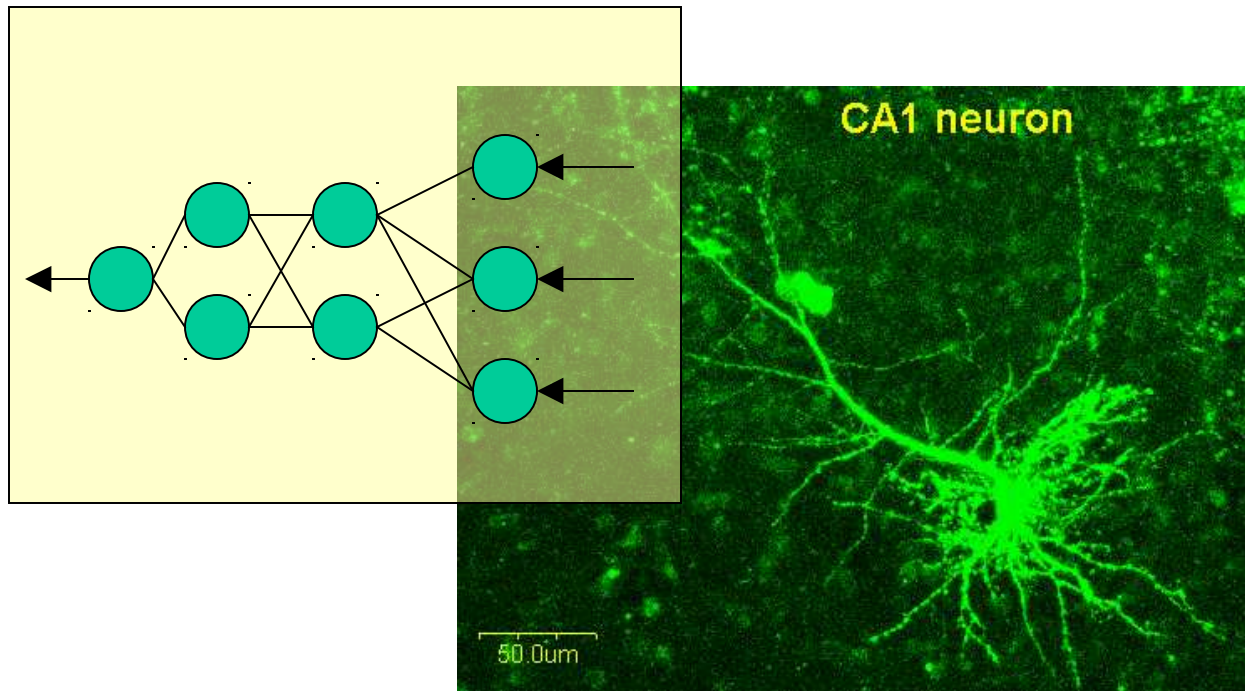


Введение в искусственные нейронные сети

Использовалась презентация
Angshuman Saha



скопировано: [www.physiol.ucl.ac.uk/fedwards/ ca1%20neuron.jpg](http://www.physiol.ucl.ac.uk/fedwards/ca1%20neuron.jpg)

Определение

Нейронные сети ANN

Широкий класс моделей, изначально создавались для моделирования работы мозга

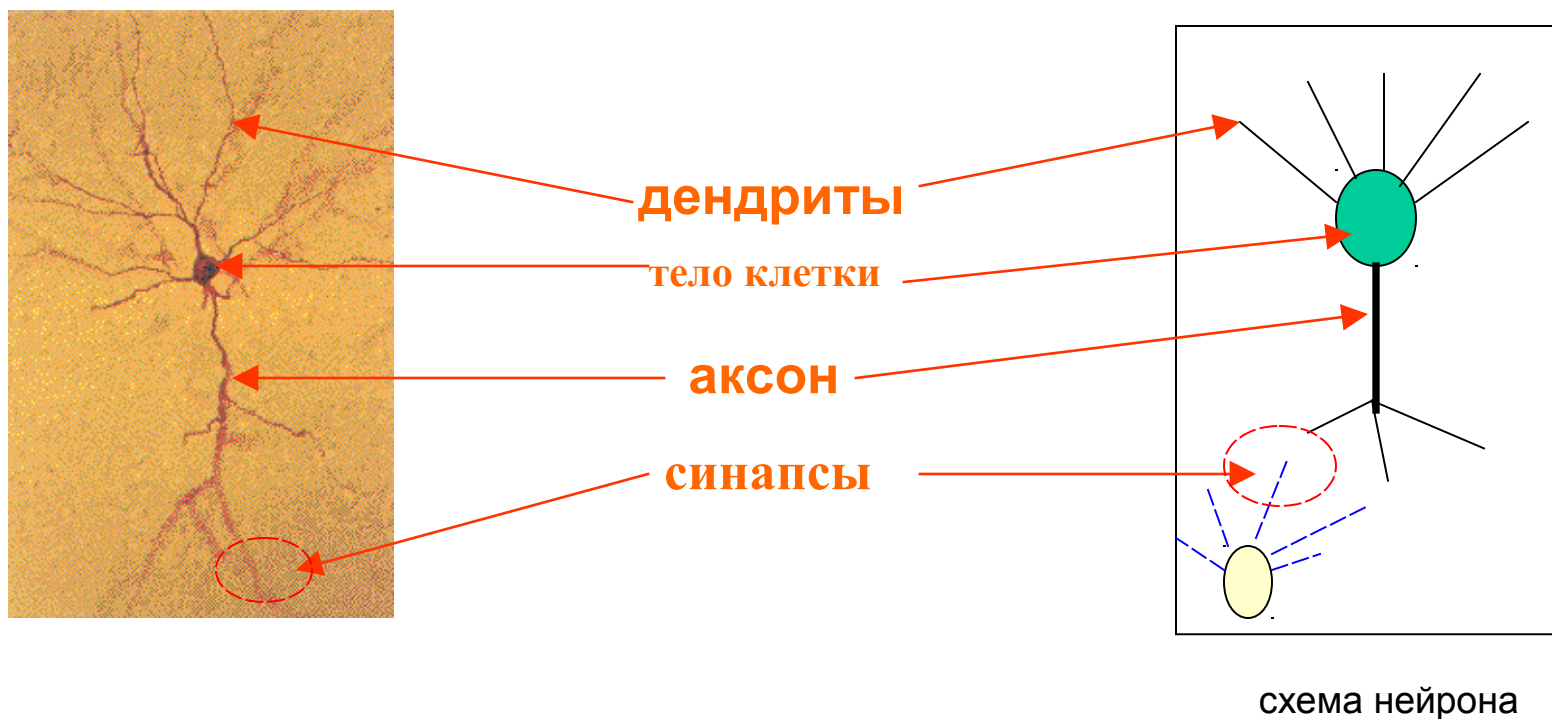
Существует много видов нейронных сетей.
Они различаются в зависимости от

- ❑ Решаемой задачи
регрессия/прогнозирование,
классификация, кластеризация
- ❑ архитектуры модели
- ❑ алгоритма обучения сети

Далее обсуждаются

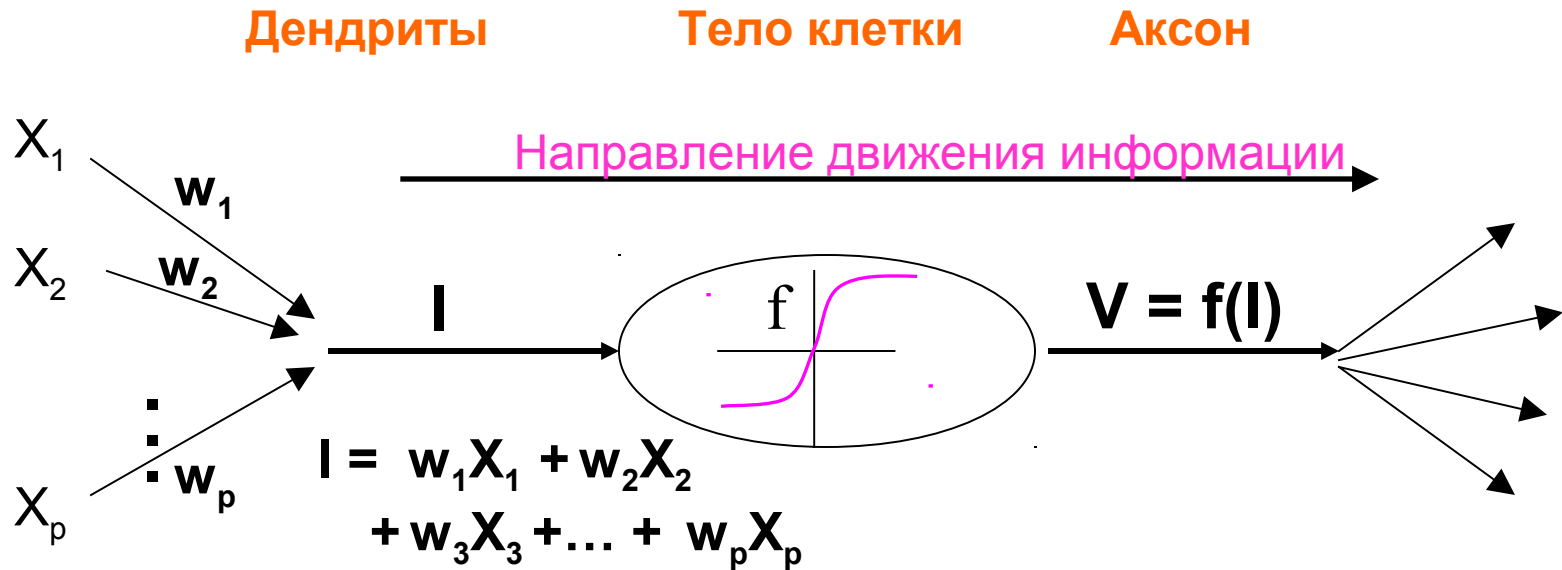
нейронные сети прямого распространения

Немного биологии . . . нейроны



- **Дендриты** – получают информацию
- **Тело клетки** – обрабатывает информацию
- **Аксон** – передает обработанную информацию другим нейронам
- **Синапсы** – соединяют аксон и дендриты других нейронов

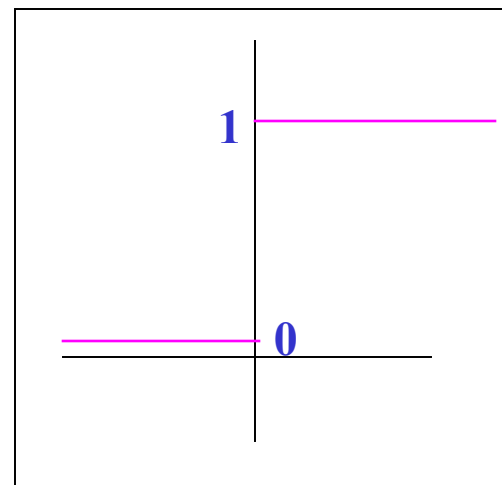
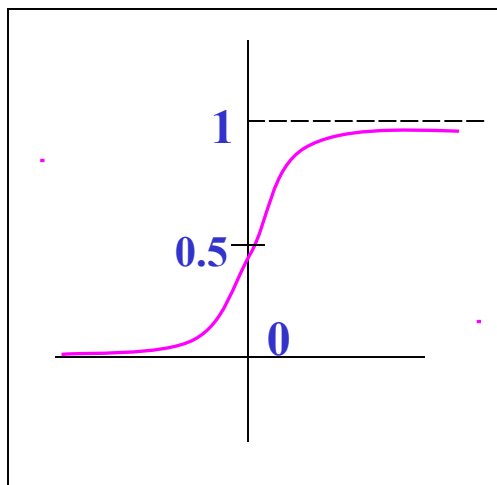
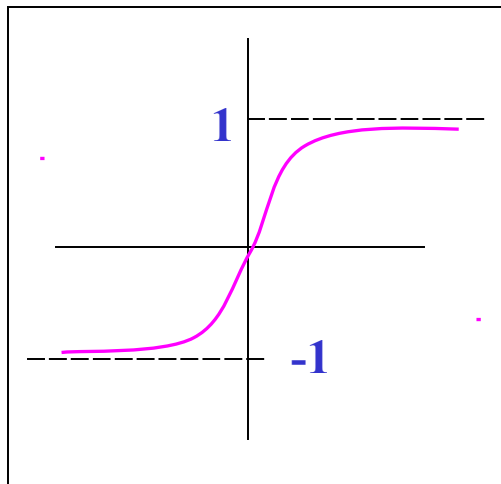
Искусственный нейрон



- $X_1 X_2 \dots X_p$ от других нейронов или от рецепторов
- Входы усиливаются или ослабляются умножением на веса w
- I = линейная комбинация всех входов
- Активационная функция f преобразует сумму в выходной сигнал
- Выходной сигнал поступает к другим нейронам или является выходом нейронной сети

Активационная функция

активационная функция



Гиперболический тангенс **Логистическая**

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$


$$f(x) = e^x / (1 + e^x)$$

Пороговая

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

Параллельные вычисления-1

Нейрон мозга соединен с $\geq 10,000$
соседним клеткам



В первые два года после
рождения каждую секунду
формируется около 1 миллиона
синапсов.

Параллельные вычисления-2

за секунду нейрон мозга
выполняет 100 операций

меньше операций, чем при
последовательном сложении 100
чисел.

Параллельные вычисления-3

Водитель реагирует на событие быстрее, чем за секунду. Программа из 100 тактов обеспечивает все реакции водителя.

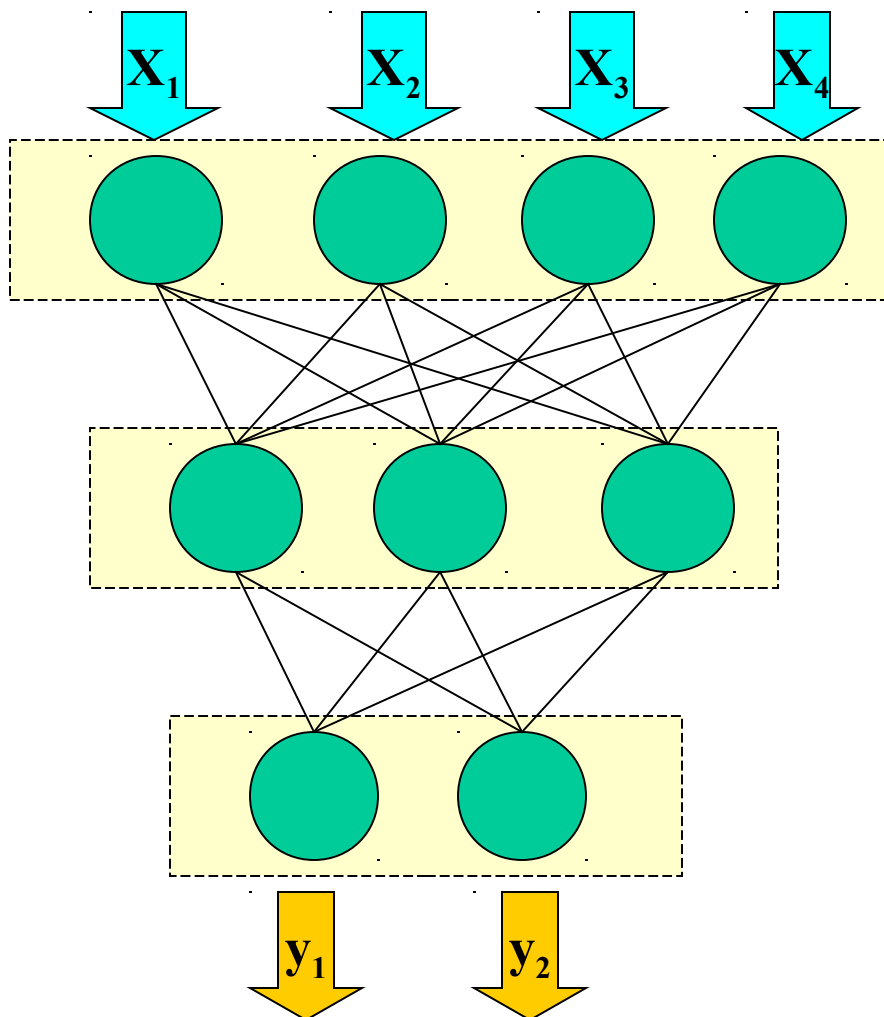
Гипотеза:, низкая скорость компенсируется количеством нейронов, количеством связей между ними, параллельными вычислениями.

в коре головного мозга порядка 10 биллионов нейронов и 60 триллионов синапсов. (Данные 1990 года).

Сети прямого распространения

Нейроны организованы послойно

Направление потока информации



Входной слой

- каждый нейрон имеет ровно один вход от внешней среды

Скрытый слой

- входы — от каждого (?) нейрона предыдущего слоя

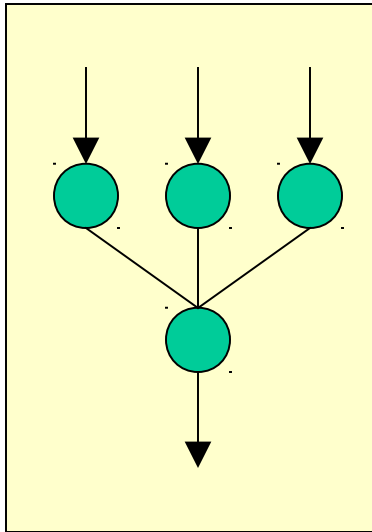
Выходной слой

- Выход каждого нейрона является реакцией сети

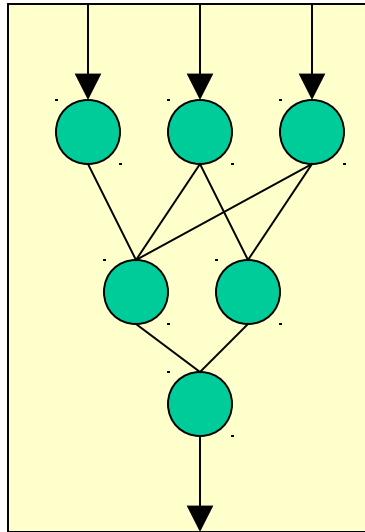
Сети прямого распространения

Число скрытых слоев может быть

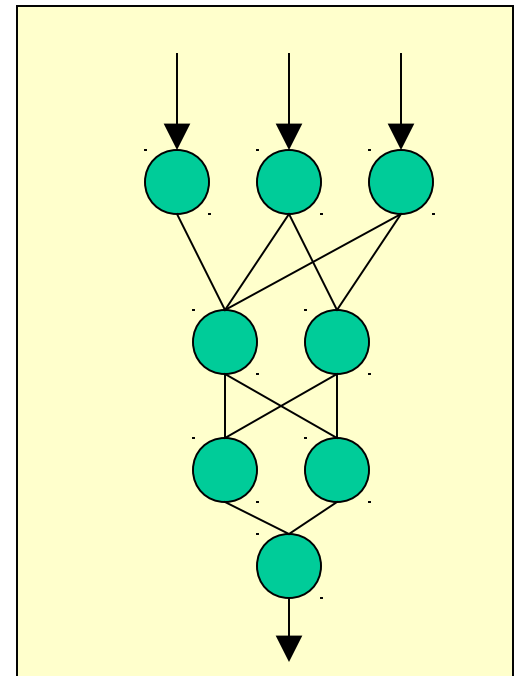
0



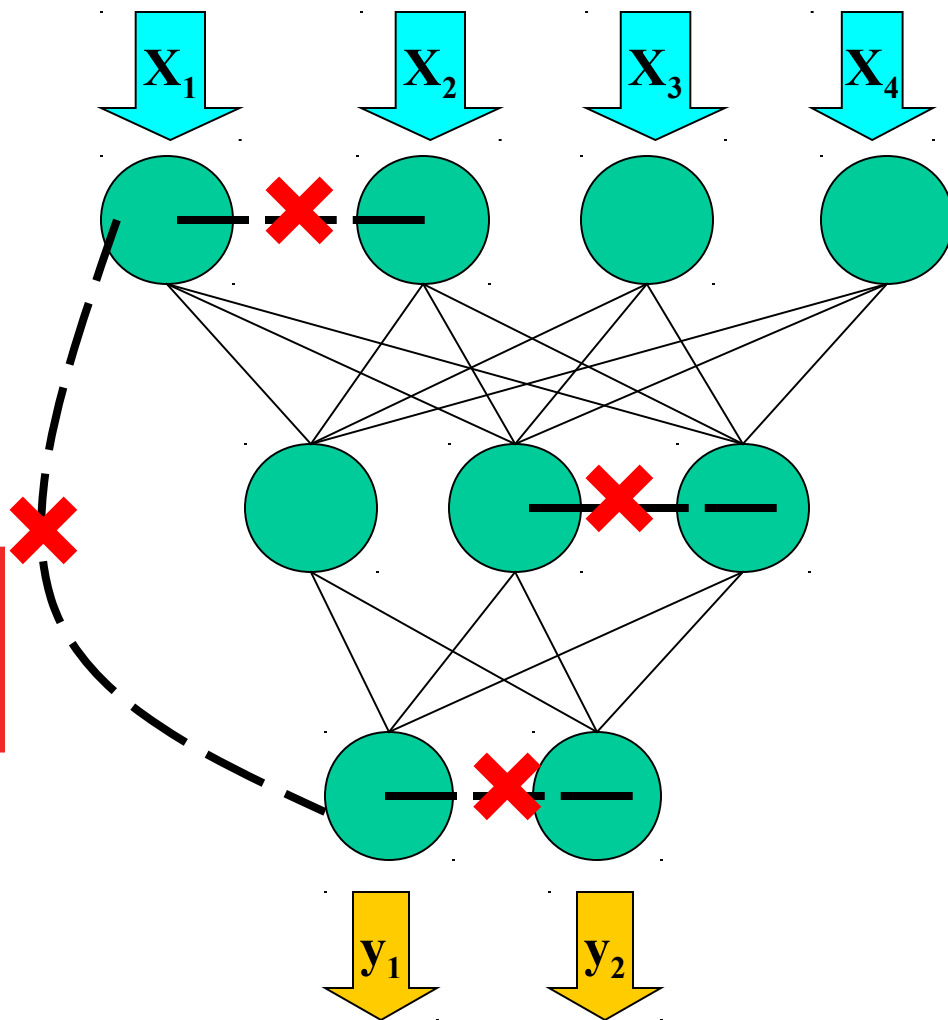
1



больше



Сети прямого распространения



- В пределах слоя нейроны не связаны.
- Нейроны передают информацию ТОЛЬКО нейронам следующего слоя. (Прямое распространение)
- Перепрыгивание через слои запрещено.

Выбор модели ANN

Входы: $X_1 X_2 X_3$ Выход: Y Модель: $Y = f(X_1 X_2 X_3)$

Нейронная сеть :

- # входных нейронов
- # скрытых слоев
- # нейронов в каждом скрытом слое
- # выходных нейронов
- активационная функция каждого нейрона
- ВЕСА каждого соединения

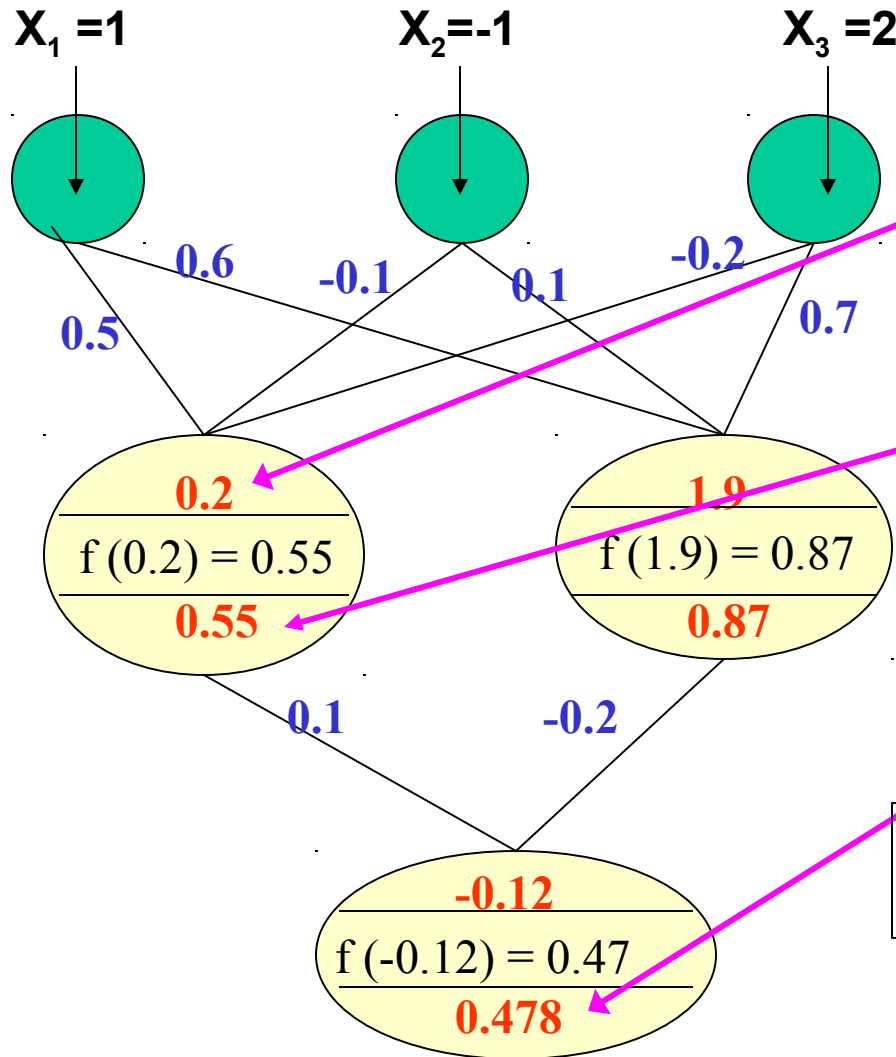
Обучение нейронной сети = определение значений весов.
Остальные параметры задаются аналитиком заранее.

Вычисления при прямом распространении

ВХОДЫ: $X_1 X_2 X_3$

ВЫХОД: Y

МОДЕЛЬ: $Y = f(X_1 X_2 X_3)$



$$0.2 = 0.5 * 1 - 0.1 * (-1) - 0.2 * 2$$

$$f(x) = e^x / (1 + e^x)$$
$$f(0.2) = e^{0.2} / (1 + e^{0.2}) = 0.55$$

ВЫХОД $Y = 0.47$

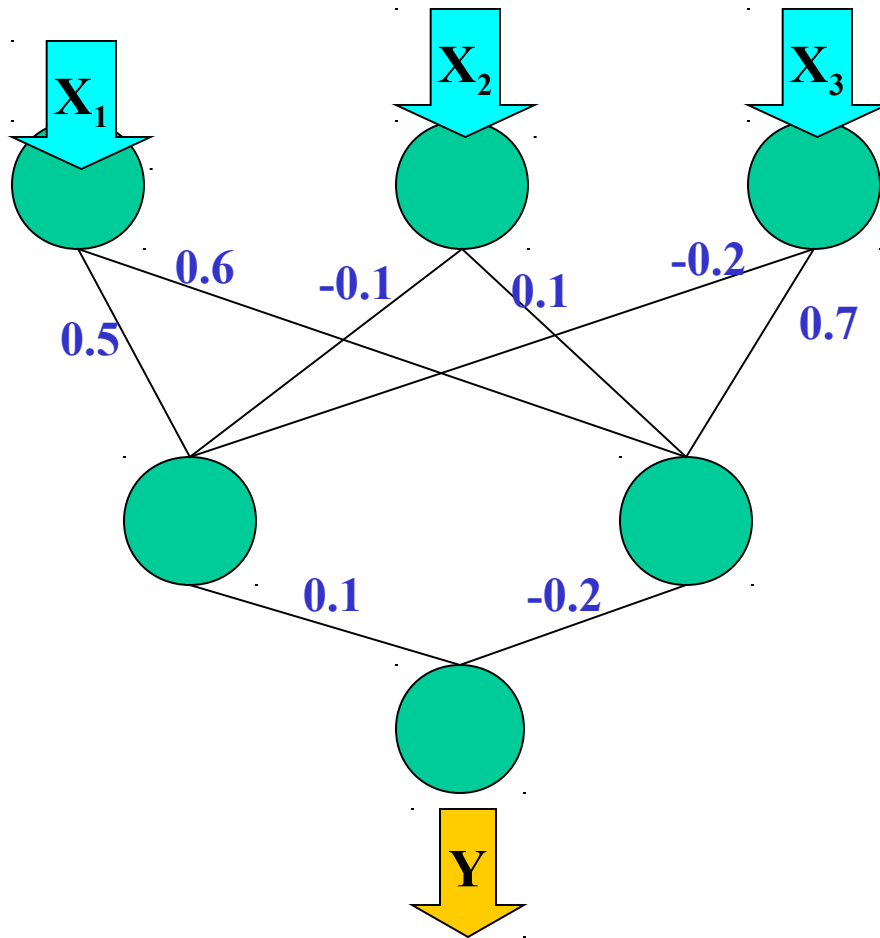
Если на выходе $Y = 2$, то
ошибка сети = $(2 - 0.47) = 1.53$

Пример модели

ВХОДЫ: X_1 X_2 X_3

ВЫХОД: Y


модель: $Y = f(X_1 X_2 X_3)$



Коннективизм

Гипотеза бабушки
(grandmother-cell hypothesis)

Память обеспечивается синапсами.



Мозг устроен как электропроводка, у которой все время меняется толщина проводов.

Выбор модели нейронной сети

Как выбрать модель ?

входы: $X_1 X_2 X_3$ **выход:** Y **модель:** $Y = f(X_1 X_2 X_3)$

входных нейронов = # входов = **3** # выходных нейронов = # выходов = **1**

скрытых слоев = ???

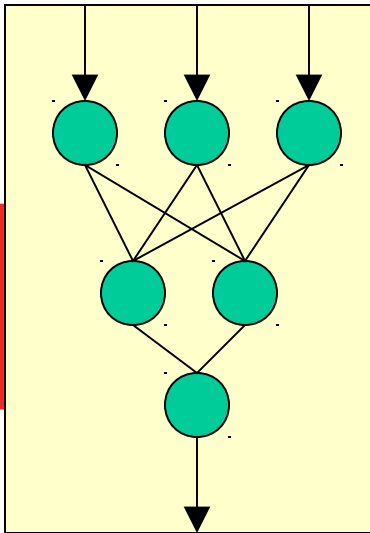
Попробуем **1**

Нет рецепта.

нейронов в скрытом слое = ???

Попробуем **2**

Пробы и ошибки.



Архитектура определена ... Как найти веса ???

В данной модели надо найти 8+3 значений весов.

$$\underline{W} = (W_1, W_2, \dots, W_{11})$$

Данные: $(Y_i, X_{1i}, X_{2i}, \dots, X_{pi}) \quad i = 1, 2, \dots, n$

Зададимся какими-то весами \underline{W} , получим для Y 'ов

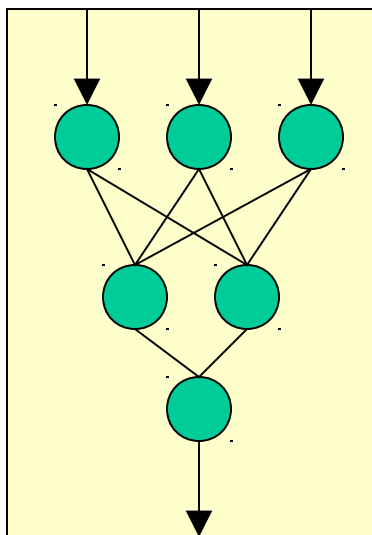
$$(V_1, V_2, \dots, V_n)$$

Эти значения - *функции* от \underline{W} .

Выберем \underline{W} так, чтобы минимизировать общую ошибку E

$$E = \sum (Y_i - V_i)^2$$

Метод обратного распространения

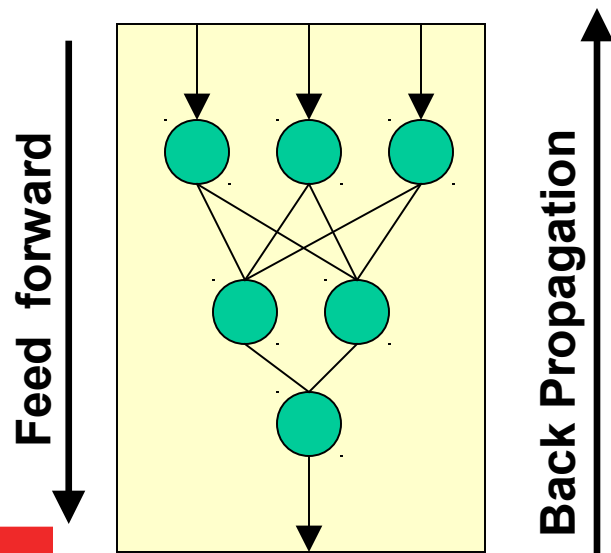


$$E = \sum (Y_i - V_i)^2$$

Метод обратного распространения (Back Propagation) задает поправки весов.

Обучение нейронной сети

Как обучить нейронную сеть ?



$$E = \sum (Y_i - V_i)^2$$

- Начнем со случайно выбранных весов.
- Подадим на вход сети 1 наблюдение X_1
 $X_1 \rightarrow V_1 \rightarrow (Y_1 - V_1)$
На выходе сети V_1 , Ошибка = $(Y_1 - V_1)$
- Подправим веса так, чтобы ошибка уменьшилась
- Подадим на вход сети 1 наблюдение.
Подправим веса так, чтобы ошибка 2 наблюдения уменьшилась
- Продолжаем так до последнего наблюдения
- Первый цикл обучения закончен.
- Выполняем эти циклы, пока общая ошибка E не станет маленькой.

Изменения весов в ходе обратного распространения

Формула для изменения весов по методу обратного распространения

V_i – на выходе сети для i -го наблюдения –

функция весов сети $\underline{W} = (W_1, W_2, \dots)$

Общая ошибка E тоже функция W

$$E(\underline{W}) = \sum [Y_i - V_i(\underline{W})]^2$$

Метод скорейшего спуска :

Каждый вес W_i изменяется по правилу

$$W_{\text{new}} = W_{\text{old}} + \alpha * (\partial E / \partial W) |_{W_{\text{old}}}$$

α = параметр обучения (между 0 и 1)

Возможен вариант формулы

$$W_{(t+1)} = W_{(t)} + \alpha * (\partial E / \partial W) |_{W_{(t)}} + \beta * (W_{(t)} - W_{(t-1)})$$

β = момент (между 0 и 1)

Алгоритм обучения

Выберите архитектуру нейронной сети
(# скрытых слоев, # нейронов в каждом слое)

Выберите скорость обучения и момент

Инициализируйте сеть, задав случайные веса

Do while, сработало правило остановки

For I = 1 to # наблюдений в выборке

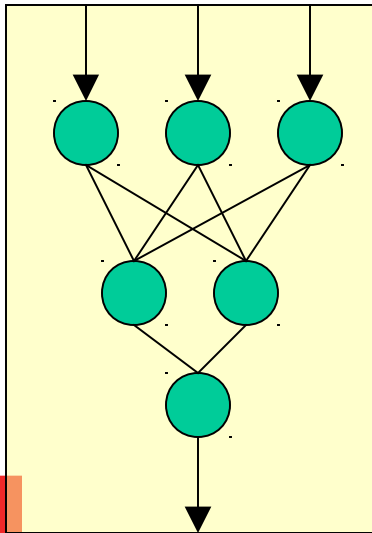
Feed forward the I-th observation thru the Net
Compute the prediction error on I-th observation

Back propagate the error and adjust weights

Next I

Проверить правило остановки

End Do



$$E = \sum (Y_i - V_i)^2$$

Критерий сходимости

Когда прекратить обучение сети ?

В идеале – когда достигнут **глобальный минимум** на поверхности ошибок

Как узнать, что достигнут ?

Неизвестно ...

В первом приближении:

1. Остановиться, когда изменение общей ошибки **мало**.
2. Остановиться, когда изменение весов **мало**.

Описание проблемы:

Ошибка убывает. Хорошая подгонка **на обучающей выборке**.

НО ... Сеть плохо **предсказывает** на **новых данных**

Название проблемы – **Перепогонка (over fitting)** на обучающей выборке

Говорят, что сеть **запоминает** обучающую выборку.

- в результате когда на входе X из обучающей выборки, сеть честно воспроизводит соответствующий Y .

-однако для X' ов которых сеть раньше не встречала, плохой прогноз.

Критерий сходимости

Улучшенное правило остановки:

Разбейте данные на **обучающую выборку** и **контрольную выборку**

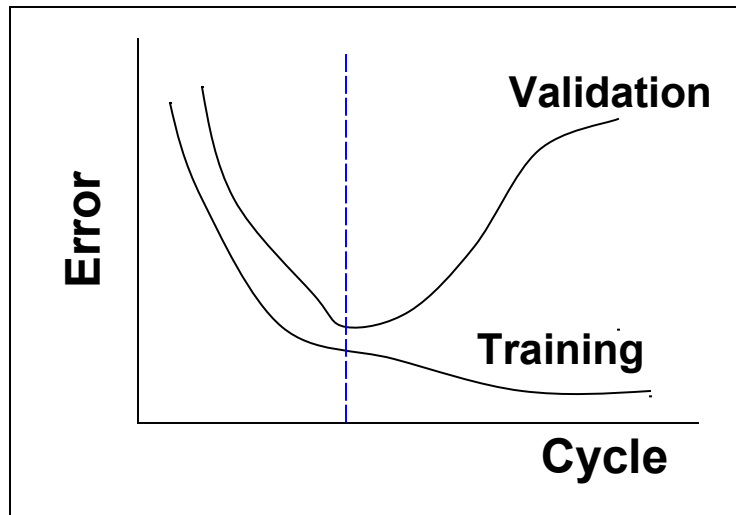
Используйте обучающую выборку для обучения сети

контрольную выборку для проверки сети на новых данных

Обычно в ходе обучения

Ошибка на обучающей выборке убывает.

Ошибка на контрольной выборке сначала убывает, потом возрастает.



Остановите обучение, когда
ошибка на тестовом множестве
начала возрастать

Выбор параметров обучения

Скорость обучения и момент

- задаются аналитиком, должны быть между 0 и 1

Как найти оптимальные значения этих параметров ?

- Ясных рекомендаций нет.
- Однако последствия неудачного выбора хорошо изучены.

Скорость обучения

Слишком большая – большие скачки в пространстве весов – можно пропустить глобальный минимум.

Слишком маленькая –

- Долго сходится
- Попад в локальный минимум трудно из него выбраться.

Советы

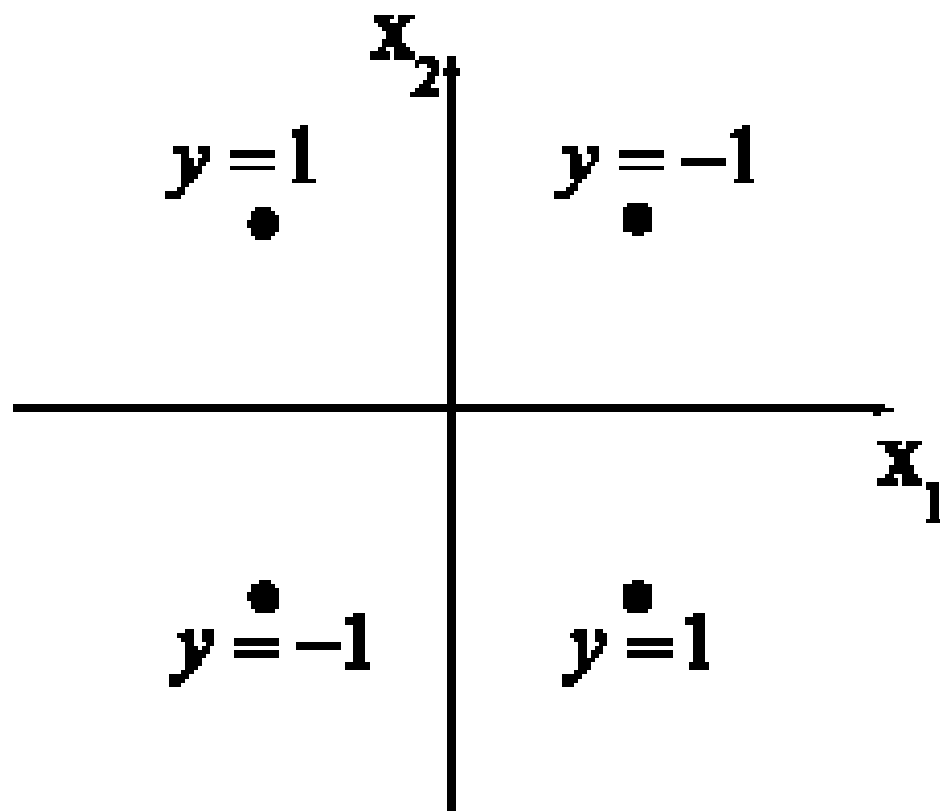
Метод проб и ошибок – пробуйте разные значения скорости обучения и момента

Выберите вариант, дающий наименьшие ошибки

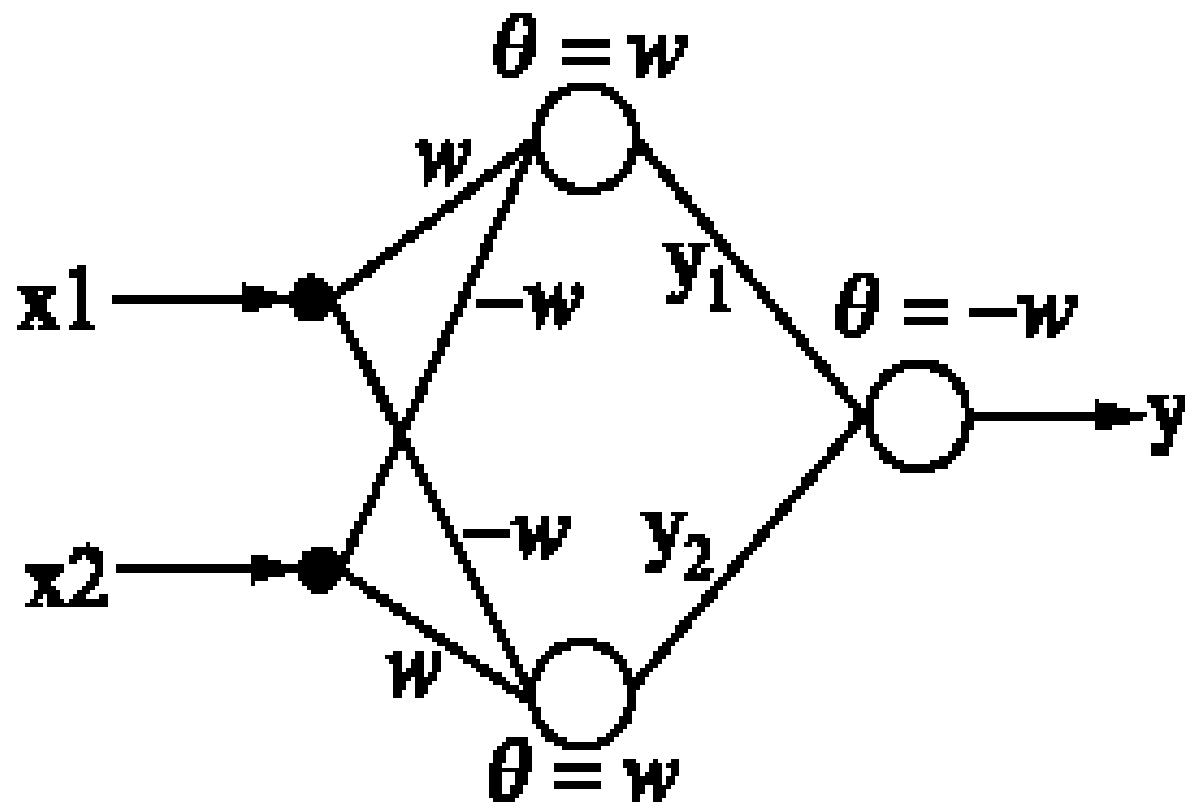
Кратко еще раз

- ❑ Широкий спектр применений – прогноз, классификация, кластеризация, ..
- ❑ Важный подкласс ANN – сети прямого распространения
 - ❖ слоистая структура – входной, скрытые, выходной слои
 - ❖ каждый слой – из нейронов
 - ❖ нейроны слоя связаны только с нейронами соседних слоев
 - ❖ Связи между нейронами имеют веса
- ❑ Обучение ANN == изменение весов.
- ❑ Нужна обучающая выборка. Веса находятся методом обратного распространения, чаще всего модификацией метода скорейшего спуска.
- ❑ Архитектуру и веса нейронной сети находятся методом проб и ошибок. Лучший вариант имеет наименьшую ошибку прогноза.

Исключающее или



Исключающее или





x_1	x_2	NET_1	NET_2	OUT_1	OUT_2	NET	
-1	-1	0	0	-1	-1	$-2w$	
-1	1	$-w$	w	-1	1	w	
1	-1	w	$-w$	1	-1	w	
1	1	0	0	-1	-1	$-2w$	