

**Institute of Electronic Engineering and Nanotechnologies
«D.GHIȚU», Republic of Moldova**

Single-layer perceptron

Maria Lupu, j.s.r.

19 April 2021

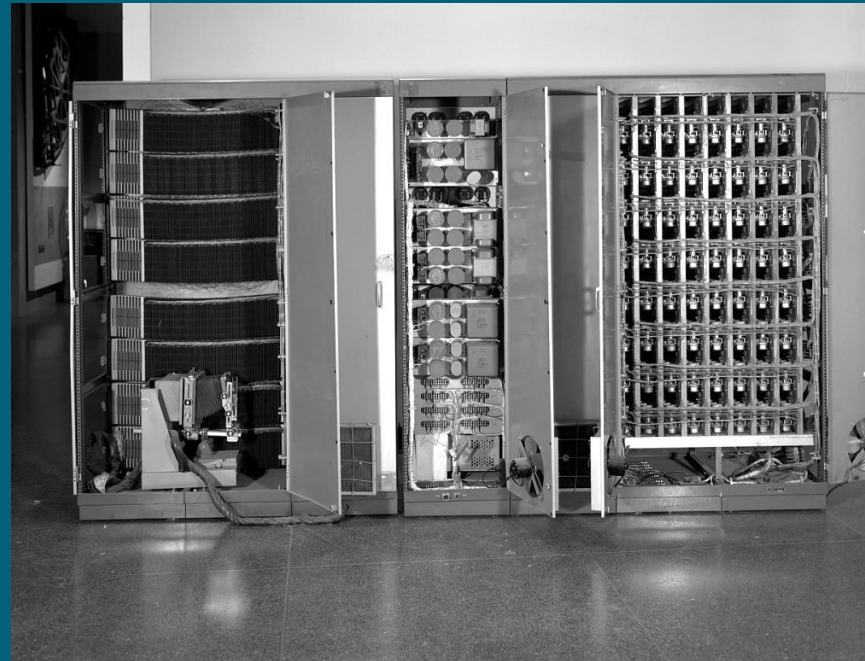
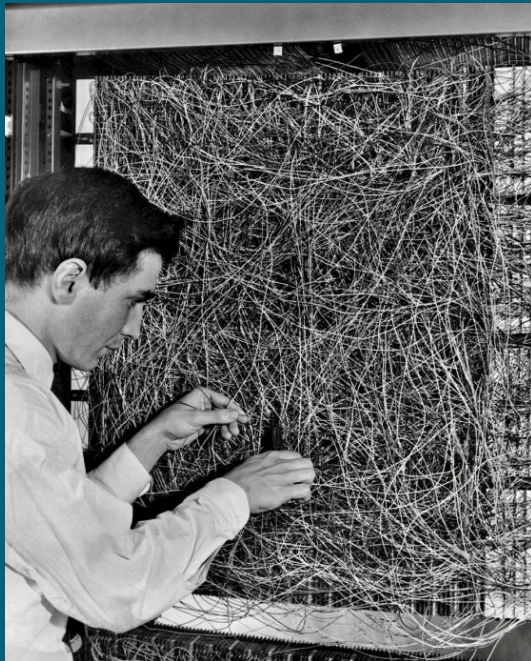
Персептрон или перцептрон (англ. perceptron от лат. perceptio — восприятие) — математическая или компьютерная модель восприятия информации мозгом (кибернетическая модель мозга), предложенная **Фрэнком Розенблаттом** в 1957 году и впервые реализованная в виде электронной машины «**Марк-1**» в 1960 году.



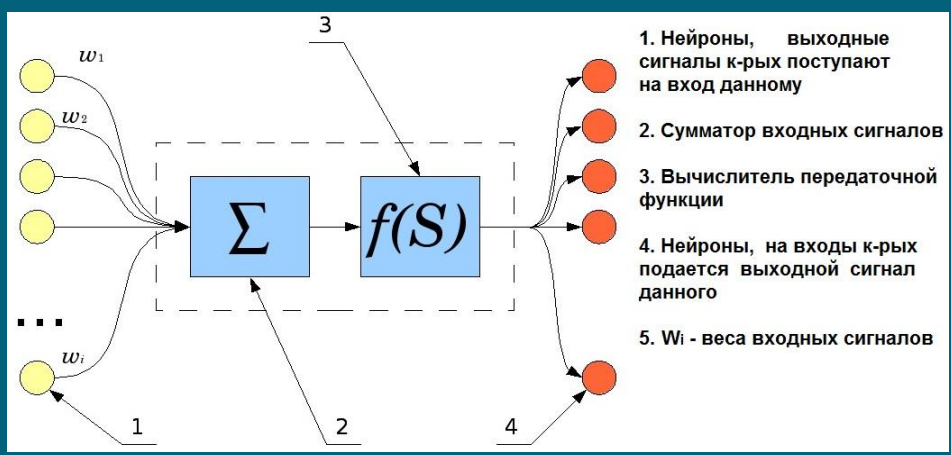
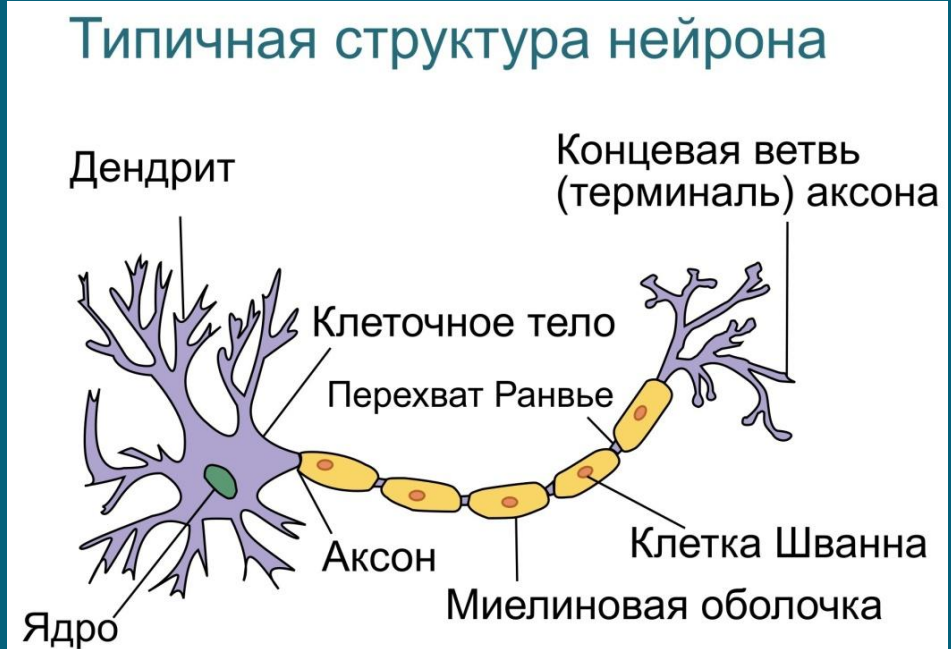
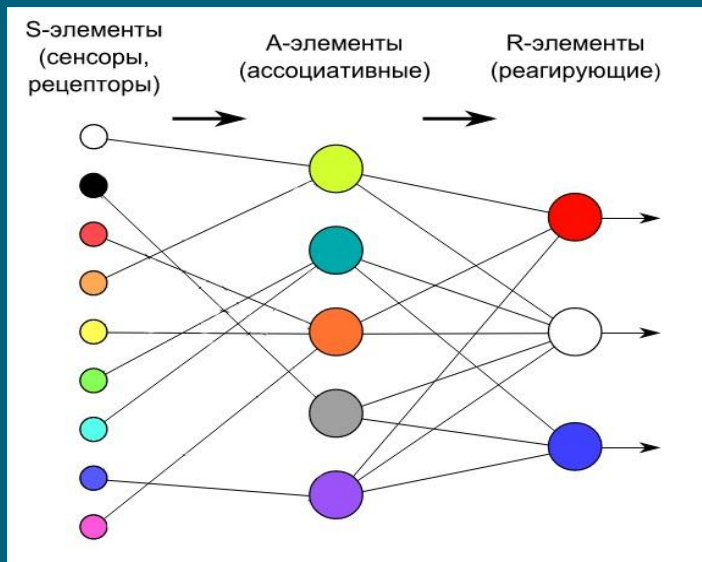
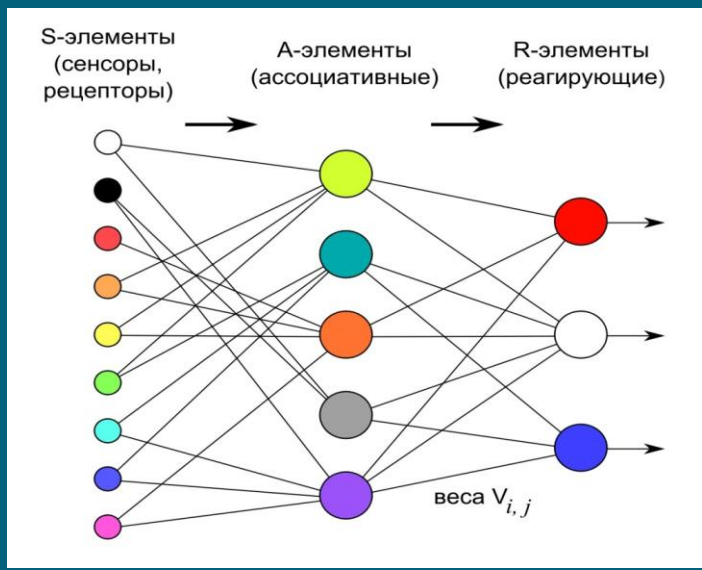
Закончив Принстонский университет, Ф. Розенblatt устроился в Корнеллскую лабораторию авиации. В 1959-м он вступил в должность директора программы по исследованию когнитивных систем в одном из кампусов Корнелла, параллельно работая лектором на факультете психологии.

В 1958—1960 годах в Корнеллском университете создал вычислительную систему «Марк-1». Это был первый нейрокомпьютер, способный обучаться в простейших задачах, он был построен на перцептроне, нейронной сети, которую Розенblatt разработал тремя годами раньше.

Перцептрон Розенblattа был первоначально программно смоделирован на компьютере **IBM 704**. Изучая нейронные сети типа перцептрона, Розенblatt надеялся «понять фундаментальные законы организации, общие для всех систем обработки информации, включая как машины, так и человеческий разум».



В основе работы перцептрона лежит математическая модель восприятия информации мозгом. Разные исследователи по-разному его определяют. В самом общем своем виде (как его описывал Розенблатт) он представляет систему из элементов трех разных типов: **сенсоров**, **ассоциативных элементов** и **реагирующих элементов**.



Первыми в работу включаются S-элементы. Они могут находиться либо в состоянии покоя (сигнал равен 0), либо в состоянии возбуждения (сигнал равен 1).

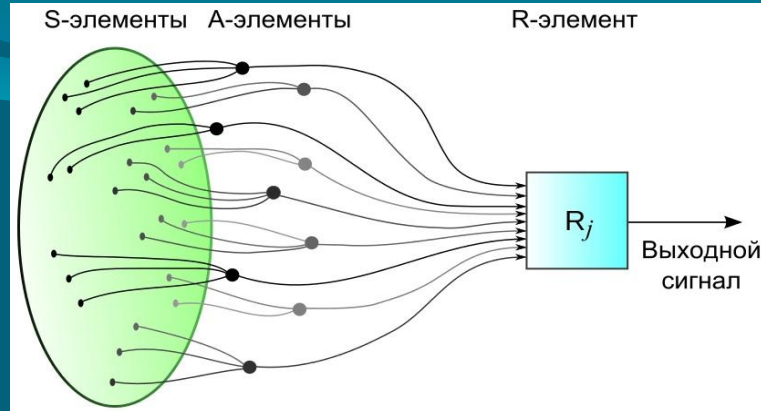
Затем сигналы от сенсорных элементов, прошедших по S-A связям попадают в A-элементы, которые еще называют ассоциативными элементами.

Ключевая особенность однослойного персептрона состоит в том, что каждый S-элемент однозначно соответствует одному A-элементу, все S-A связи имеют вес, равный 1. Если сигналы, поступившие на A-элемент, в совокупности превышают некоторый его порог $\theta = 1$, то этот A-элемент возбуждается и выдает сигнал, равный 1. В противном случае (сигнал от S-элементов не превысил порога A-элемента), генерируется нулевой сигнал.

A-элементы назвали ассоциативными, потому что A-элементы являются агрегаторами сигналов от сенсорных элементов. Например, у нас есть группа сенсоров, каждый из которых распознает кусок буквы «М» на исследуемой картинке. Однако только их совокупность (то есть когда несколько сенсоров выдали сигнал, равный 1) может возбудить A-элемент целиком. На другие буквы A-элемент не реагирует, только на букву «М». То есть он ассоциируется с буквой «М».

Можно привести и другие примеры. Наши глаза состоят из невероятного количества (около 140 млн.) S-элементов (сенсоров), улавливающих падающий свет. И есть какой-то A-элемент, который распознает конкретную часть лица. И вот мы видим на улице человека. Некоторые A-элементы, которые распознали конкретные части лица, возбуждаются.

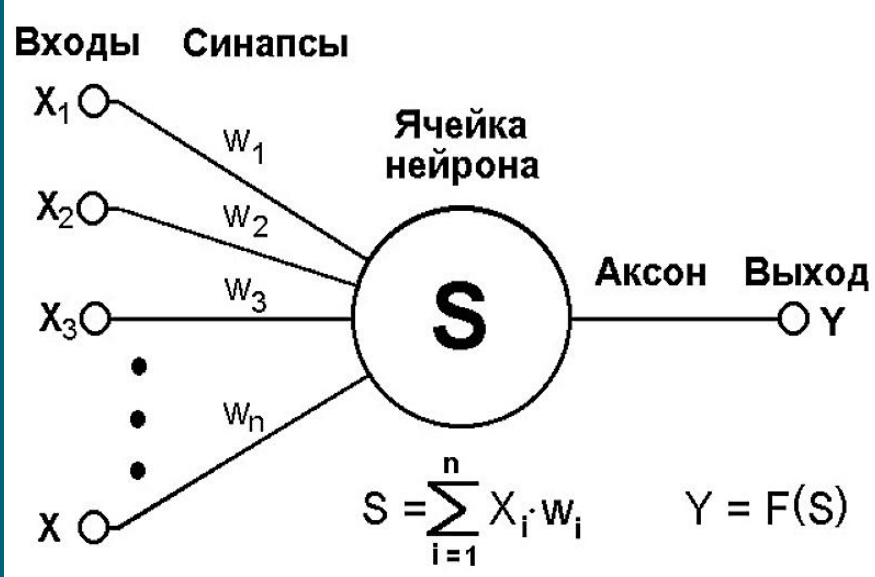
Далее сигналы, которые произвели возбужденные А-элементы, направляются к сумматору (R-элемент), действие которого уже известно. Однако, чтобы добраться до R-элемента, они проходят по A-R связям, у которых тоже есть веса. Однако здесь они уже могут принимать любые значения (в отличие от S-A связей).



R-элемент складывает друг с другом взвешенные сигналы от А-элементов и, если превышен определенный порог, генерирует выходной сигнал, равный 1. Это означает, что в общем потоке информации от глаз мы распознали лицо человека.

Если порог не превышен, то выход персептрона равен -1. То есть мы не выделили лицо из общего потока информации.

Так как R-элемент определяет выход персептрона в целом, его назвали реагирующим.

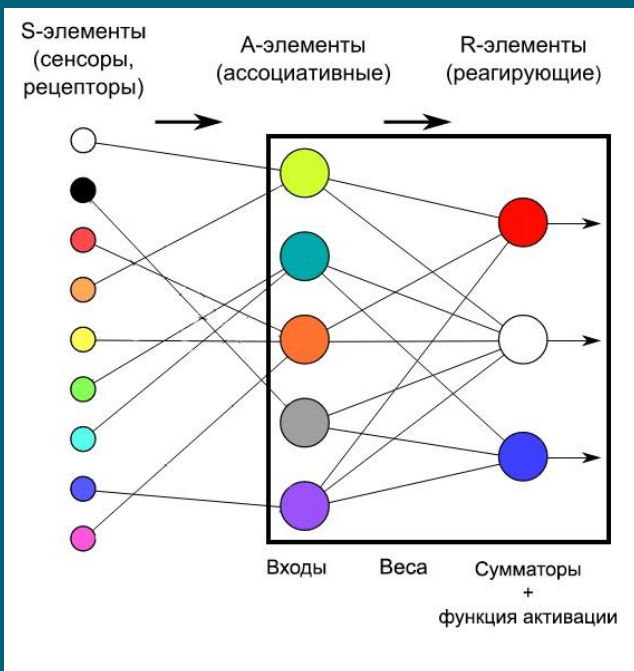


Текущее состояние нейрона определяется, как взвешенная сумма его входов.

Значение на входе синапса умножается на вес данного синапса, потом все эти значения суммируются и получаем текущее состояние нейрона.

По физическому смыслу вес синаптической связи – это электрическая проводимость данного синапса.

Таким образом становится ясно, что часть однослойного перцептрона (выделена черным прямоугольником на картинке) можно представить в виде искусственного нейрона. S-элементов в искусственном нейроне нет, в однослойном перцептроне S-элементы и A-элементы могут принимать только фиксированные значения 0 и 1, тогда как в искусственном нейроне таких ограничений нет.



Нейронные сети не программируются в привычном смысле этого слова, они обучаются. Возможность обучения — одно из главных преимуществ нейронных сетей перед традиционными алгоритмами.

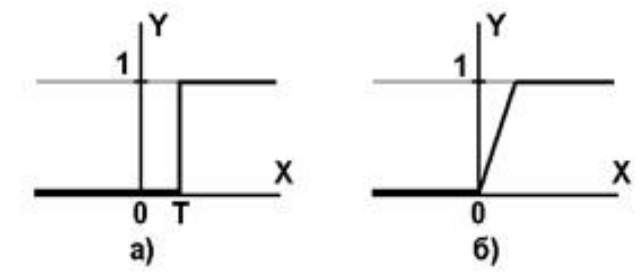
Технически обучение заключается в нахождении **коэффициентов связей между нейронами**. В процессе обучения нейронная сеть способна выявлять сложные зависимости между входными данными и выходными, а также выполнить обобщение. Это значит, что в случае успешного обучения сеть сможет получить верный результат на основании данных, которые отсутствовали в обучающей выборке, а также неполных и/или «зашумленных», частично искажённых данных.

После обучения перцептрон готов работать в режиме **распознавания** или **обобщения**. В этом режиме перцептрону предъявляются ранее неизвестные ему объекты, и перцептрон должен установить, к какому классу они принадлежат.

В искусственных нейронных сетях **функция активации нейрона** определяет выходной сигнал, который определяется входным сигналом или набором входных сигналов.

Стандартная компьютерная микросхема может рассматриваться как цифровая сеть функций активации, которые могут принимать значения «ON» (1) или «OFF» (0) в зависимости от входа. Это похоже на поведение линейного персептрона в нейронных сетях. Однако только нелинейные функции активации позволяют таким сетям решать нетривиальные задачи с использованием малого числа узлов. В искусственных нейронных сетях эта функция **также называется передаточной функцией**.

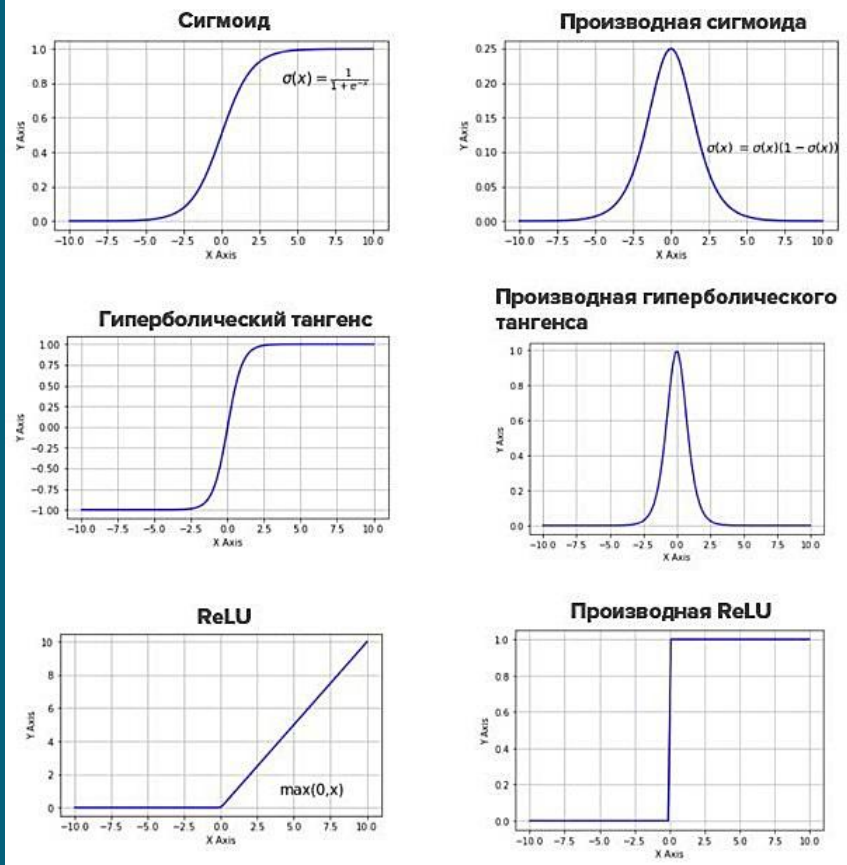
Активационная функция может иметь различный вид:



а) функция единичного скачка
б) линейный порог(гистерезис)

Каждая функция имеет свои отличительные свойства, преимущества и недостатки. Ни одна из функций не является универсальной, однозначно сказать, в каком случае следует использовать линейный выпрямитель, сигмойду или гиперболический тангенс, невозможно.

Зная некоторые характеристики функции, которую требуется аппроксимировать, следует выбирать такую активационную функцию, которая аппроксимирует искомую функцию максимально точно и приведёт к быстрому обучению.



Например, линейная функция, позволяет получать спектр значений, а не только бинарный ответ. Если все они по своей природе линейные, то финальная функция активации в последнем слое будет просто линейной функцией от входов на первом слое, т.о. теряется возможность делать наборы из слоев.

Сигмоида — нелинейна по своей природе, а комбинация таких функций производит тоже нелинейную функцию. Её хорошо использовать в задачах классификации. Например, аппроксимацию классифицирующей функции комбинацией сигмоид можно провести легче, чем применяя ReLU.

Гиперболический тангенс имеет те же характеристики, что и у сигмоиды. Эта функция так же нелинейна, она хорошо подходит для комбинации слоёв. Поэтому можно не беспокоиться, что активационная функция перегрузится от больших значений. Однако стоит отметить, что градиент тангенциальной функции больше, чем у сигмоиды (производная круче). Решение о том, выбрать сигмоиду или тангенс, зависит от требований к амплитуде градиента.

Может показаться, что функция ReLU имеет те же проблемы, что и линейная функция, так как линейна в первом квадранте. Но на самом деле ReLU нелинейна по своей природе, и комбинация ReLU также нелинейна. Такая функция является хорошим аппроксиматором. ReLU менее требовательна к вычислительным ресурсам, чем гиперболический тангенс или сигмоида, так как производит более простые математические операции. Поэтому имеет смысл использовать ReLU при создании глубоких нейронных сетей.

Помимо формул, представленных выше, существуют множество других функций активации, исследователи в области искусственного интеллекта постоянно ищут новые способы улучшить нейронные сети, в том числе модифицируя уже существующие способы и приемы обработки входных сигналов.

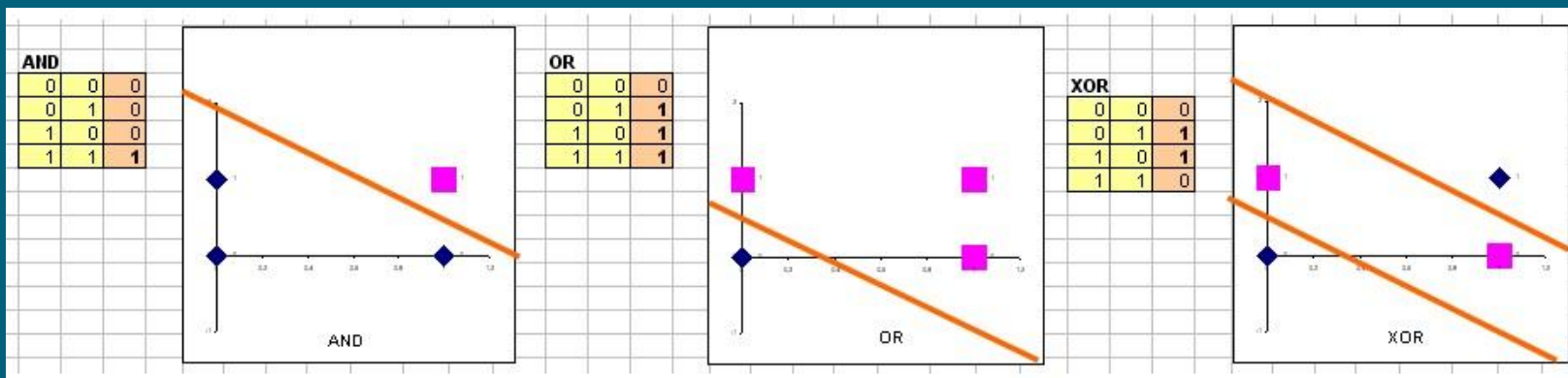
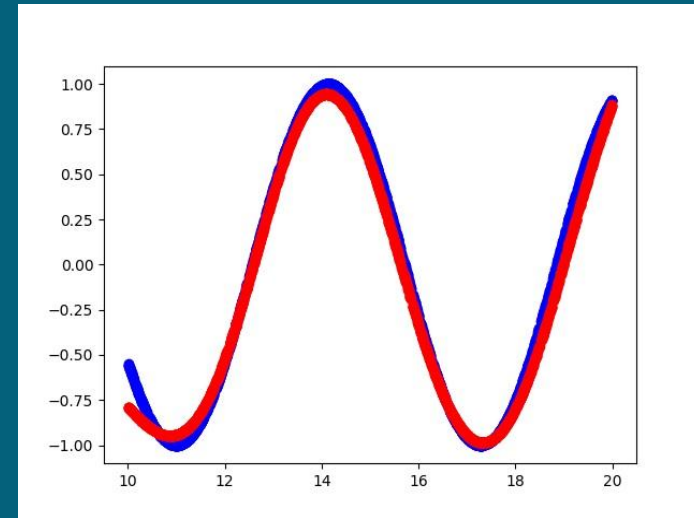
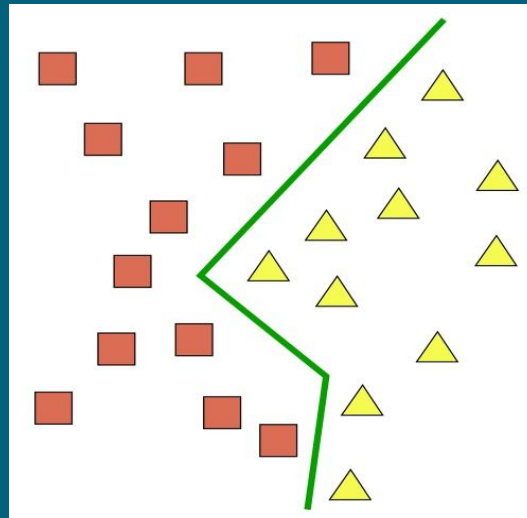
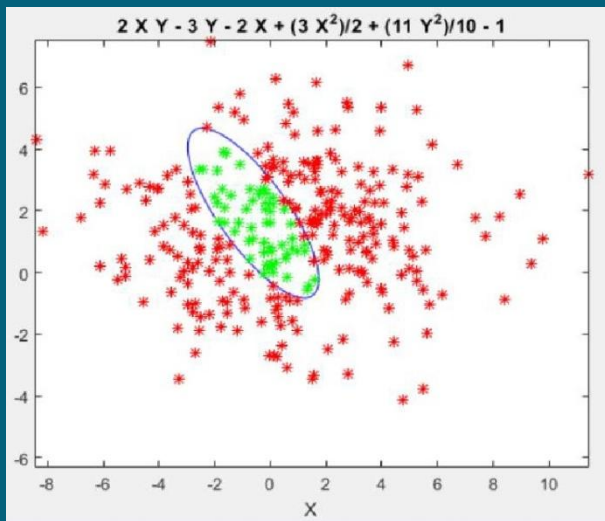
При решении сложных задач с использованием многослойных сетей возможно комбинировать функции на разных слоях, чтобы добиться наиболее высокой точности и скорости обучения.

Название	Формула	Область значений
Пороговая	$\Psi = F(\psi) = \begin{cases} 0, \psi < \theta, \\ 1, \psi \geq \theta. \end{cases}$	0, 1
Знаковая	$\Psi = F(\psi) = \begin{cases} 1, \psi > 0, \\ -1, \psi \leq 0. \end{cases}$	-1, 1
Сигмовидная	$\Psi = F(\psi) = \frac{1}{1 + e^{-\psi}}$	(0, 1)
Полулинейная	$\Psi = F(\psi) = \begin{cases} \psi, \psi > 0, \\ 0, \psi \leq 0. \end{cases}$	(0, ∞)
Линейная	$F(\psi) = \psi$	(-∞, ∞)
Радиальная базисная	$\Psi = F(\psi) = e^{-\psi^2}$	(0, 1)
Полулинейная с насыщением	$\Psi = F(\psi) = \begin{cases} 0, \psi \leq 0, \\ \psi, 0 < \psi < 1, \\ 1, \psi \geq 1. \end{cases}$	(0, 1)
Линейная с насыщением	$\Psi = F(\psi) = \begin{cases} -1, \psi \leq -1, \\ \psi, -1 < \psi < 1, \\ 1, \psi \geq 1. \end{cases}$	(-1, 1)
Гиперболический тангенс	$\Psi = F(\psi) = \frac{e^{\psi} - e^{-\psi}}{e^{\psi} + e^{-\psi}}$	(-1, 1)
Треугольная	$\Psi = F(\psi) = \begin{cases} 1 - \psi , \psi \leq 1, \\ 0, \psi > 0. \end{cases}$	(0, 1)

Применение однослойного персептрона в практических задачах:

- **Возможность классификации.** Сеть с одним скрытым слоем, содержащим N нейронов со ступенчатой функцией активации, способна осуществить произвольную классификацию Nd точек d -мерного пространства (то есть классифицировать Nd примеров). Одного скрытого слоя нейронов с сигмоидной функцией активации достаточно для аппроксимации любой границы между классами со сколь угодно высокой точностью.

- **Возможность аппроксимации.** Одного скрытого слоя нейронов с сигмоидной функцией активации достаточно для аппроксимации любой функции и ее производных со сколь угодно высокой точностью.



Задачи, решаемые с помощью ИНС

Классификация образов. Задача состоит в указании принадлежности входного образа (например, речевого сигнала или рукописного символа), представленного вектором признаков, одному или нескольким предварительно определенным классам. К известным приложениям относятся распознавание букв, распознавание речи, классификация сигнала электрокардиограммы, классификация клеток крови.

Кластеризация/категоризация. При решении задачи кластеризации, отсутствует обучающая выборка с метками классов. Алгоритм кластеризации основан на подобии образов и размещает близкие образы в один кластер. Известны случаи применения кластеризации для извлечения знаний, сжатия данных и исследования свойств данных.

Аппроксимация функций. Предположим, что имеется обучающая выборка $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$ (пары данных вход-выход), которая генерируется неизвестной функцией (x) , искаженной шумом. Задача аппроксимации состоит в нахождении оценки неизвестной функции (x) . Аппроксимация функций необходима при решении многочисленных инженерных и научных задач моделирования.

Предсказание/прогноз. Пусть заданы n дискретных отсчетов $\{y(t_1), y(t_2), \dots, y(t_n)\}$ в последовательные моменты времени t_1, t_2, \dots, t_n . Задача состоит в предсказании значения $y(t_{n+1})$ в некоторый будущий момент времени t_{n+1} . Предсказание/прогноз имеют значительное влияние на принятие решений в бизнесе, науке и технике. Предсказание цен на фондовой бирже и прогноз погоды являются типичными приложениями техники предсказания/прогноза.

Оптимизация. Многочисленные проблемы в математике, статистике, технике, науке, медицине и экономике могут рассматриваться как проблемы оптимизации. Задачей алгоритма оптимизации является нахождение такого решения, которое удовлетворяет системе ограничений и максимизирует или минимизирует целевую функцию. Задача коммивояжера является классическим примером задачи оптимизации.

Память, адресуемая по содержанию. В модели вычислений фон Неймана обращение к памяти доступно только посредством адреса, который не зависит от содержания памяти. Более того, если допущена ошибка в вычислении адреса, то может быть найдена совершенно иная информация. Ассоциативная память, или память, адресуемая по содержанию, доступна по указанию заданного содержания. Содержимое памяти может быть вызвано даже по частичному входу или искаженному содержанию. Ассоциативная память чрезвычайно желательна при создании мультимедийных информационных баз данных.

Управление. Рассмотрим динамическую систему, заданную совокупностью $\{u(t), y(t)\}$, где $u(t)$ является входным управляющим воздействием, а $y(t)$ - выходом системы в момент времени t . В системах управления с эталонной моделью целью управления является расчет такого входного воздействия $u(t)$, при котором система следует по желаемой траектории, диктуемой эталонной моделью. Примером является оптимальное управление двигателем. 11

Возможности и ограничения перцептронов

Сам Розенблатт выделил два фундаментальных ограничения для трёхслойных перцептронов (состоящих из одного S-слоя, одного А-слоя и R-слоя):

- отсутствие у них способности к обобщению своих характеристик на новые стимулы или новые ситуации;
- неспособность анализировать сложные ситуации во внешней среде путём расчленения их на более простые.

В 1969 году М. Минский и С. Паперт опубликовали книгу «Перцептроны», где математически показали, что перцептроны, подобные розенблаттовским, принципиально не в состоянии выполнять многие из тех функций, которые хотели получить от перцептронов.

Способности ИНС и, в частности, перцептрона, не столь велики, как это рекламируется. И это связано даже не с устройством, или алгоритмом, осуществляющим прогноз, а с самим происходящим явлением. Только в том случае, когда во внимание берутся существенные параметры, на основании которых будет строиться прогноз, будет иметь место некоторый успех. Как только такие параметры определены, можно начать статистическую обработку данных и построить модель явления. Но данная модель будет лишь показывать зависимость выбранных входных параметров от выходных, которые имели место в прошлом. Минский пытался показать, что перцептрон не имеет в этом отношении серьёзных преимуществ по сравнению с другими статистическими методами прогноза.

Но разницу составляет то, что классические статистические методы требуют расчёта многих сложных уравнений, а перцептрон более естественно решает требуемые уравнения, что связано с его устройством. Если на перцептрон посмотреть глазами математика, то окажется, что это по сути способ решения систем уравнений с большим числом неизвестных коэффициентов. Алгоритм поиска этих коэффициентов технически более быстрый, чем у аналогичных классических способов решения уравнений.

Построив систему уравнений, охватывающую наиболее значимые параметры, можно говорить о том, что мы нашли закон, по статусу близкий к законам физики, но только оперирующий большим числом переменных. Именно такие модели позволяют описывать системы с большим числом состояний — биологические, социальные и т. п. Именно в этом смысле можно говорить о прогнозе.

Качество сделанного прогноза или точность построенной модели зависит от количества знаний, используемых при её построении. Если мы хотим, чтобы на основании половины всех необходимых знаний модель была способна достроить (спрогнозировать) вторую половину неизвестных нам знаний, то желательно иметь информацию, равномерно распределённую по всему пространству возможных состояний. В таком случае персептрон способен спрогнозировать неизвестные, но близкие к известным, результаты с определённой вероятностью правильности.

Человек с такой задачей справляется быстро, так как находит определённую аналогию. Для персептрона (а также для ряда других ИНС) данная задача в полной мере слишком сложна. Это связано с основным ограничением ИНС — невозможность найти инвариант, вследствие этого персептрон работает только как статистическая машина, он не способен самостоятельно находить инварианты, которые были бы основой для принятия решений.

Такие задачи возникают, например, если нам требуется построить машину для чтения печатных букв или цифр так, чтобы эта машина могла распознавать их независимо от положения на странице; или если нам нужно определить, из скольких частей состоит фигура; или находятся ли две фигуры рядом или нет.

Таким образом, рассмотрев базовые принципы строения и функционирования однослойного персептрона, можно сделать выводы о возможности его применения в различных областях и усовершенствования рабочих характеристик с целью создания моделей, которые способны максимально точно выполнять поставленные задачи и максимально быстро обучаться.

Спасибо за внимание!