

# Введение в нейронные сети

Павел Панфилов  
nscalp@mail.ru

*В прессе про нейросети говорят, как правило, общими фразами и часто в восторженных тонах. Специалисты, занимающиеся этой проблемой, трезво оценивают возможность использования нейросетей, понимая их истинные достоинства и недостатки. Потенциал у нейронных технологий огромен, но их эффективное использование требует определенного уровня знаний и понимания*

*принципов их действия. Работающие трейдеры, как правило, не имеют специальной подготовки по нейросетевым технологиям, а академические статьи не объясняют, как ими пользоваться. В этой статье делается попытка определить круг базовых понятий, на основе которых новички приобретут некоторые навыки работы в этой интересной области.*

## История возникновения нейронных сетей

Человека всегда интересовало устройство мозга, принципы его работы и организации. Первая математическая модель нейрона (базового элемента мозга) была создана в 1943 году, когда американский ученый Уоррен Маккаллох (McCulloch W.) и его ученик У. Питтс (Pitts W.) сформулировали основные положения теории деятельности головного мозга. Ими было сделано следующее:

- разработана модель нейрона как простейшего процессорного элемента, выполнявшего вычисление переходной функции от скалярного произведения вектора входных сигналов и вектора весовых коэффициентов;
- предложена конструкция сети таких элементов для выполнения логических и арифметических операций;
- сделано основополагающее предположение о том, что такая сеть способна обучаться, распознавать образы, обобщать полученную информацию.

За прошедшие десятилетия теория нейронных сетей ушла достаточно дале-

ко вперед, однако многие утверждения У. Маккаллоха остаются актуальными и по сей день. В частности, при большом разнообразии моделей нейронов принцип их действия, заложенный У. Маккаллохом и У. Питтсом, остается неизменным. Недостатком модели У. Маккаллоха является сама модель нейрона — «пороговый» вид активационной функции. Пороговый вид функции не дает нейронной сети достаточную гибкость при обучении и настройке на поставленную задачу.

Дальнейшее развитие теория нейронных сетей получила в работах американского нейрофизиолога Френсиса Розенблата (Rosenblatt F.). В 1958 году он предложил свою модель нейронной сети. Ф. Розенблат ввел в модель У. Маккаллоха и У. Питтса способность связей к модификации, что сделало ее обучаемой. Эта модель была названа персептроном. Первоначально персептрон представлял собой однослойную структуру с жесткой пороговой функцией процессорного элемента и бинарными или многозначными входами, в дальнейшем эта модель была значительно усовершенствована.

Персептрон, воспринятый первоначально безоговорочно и с большим энтузиазмом, вскоре подвергся интенсивным нападкам со стороны крупных научных авторитетов.

В 1982 году американский биофизик Дж. Хопфилд (Hopfield J.) предложил оригинальную модель нейронной сети, в дальнейшем названную его именем. Сеть Дж. Хопфилда имеет обратные связи между слоями, что увеличивает ее обобщающие свойства, такие сети нашли широкое применение при распознавании образов. В последующие несколько лет было найдено множество эффективных алгоритмов и архитектур нейросетей, таких как сеть обратного распространения, двунаправленная ассоциативная память, самоорганизующиеся карты и др.

Построение любой нейронной сети предполагает большой объем вычислений (обучение сети обычно является итерационным процессом). Поэтому только с ростом вычислительной мощности компьютеров появилась возможность практического применения нейросетей, что дало мощный толчок к широ-

кому распространению программ, использующих принципы нейросетевой обработки данных.

### Искусственный нейрон

Искусственные нейросети чрезвычайно разнообразны по своей архитектуре, и в настоящее время создано более 20 их парадигм<sup>1</sup>. Все они имеют простой общий базовый элемент — *искусственный нейрон* (рис. 1), который имитирует свойства своего биологического родителя.

На вход искусственного нейрона (см. рис. 1) подается набор входных сигналов  $x_1, x_2, \dots, x_n$  (входной вектор  $X$ ), представляющий собой выходные сигналы других нейроноподобных элементов или вход-

водится необходимое нелинейное преобразование:

$$Y = F(S).$$

Эта функция называется *активационной*, примеры таких функций приведены на рис. 2. Наиболее часто на практике применяется сигмоидальная (логистическая) функция вида:

$$F(S) = \frac{1}{1 + e^{-\beta S}}.$$

Эта функция позволяет работать как со слабыми уровнями возбуждения (значения около 0), так и с большими уровнями возбуждения нейрона. При силь-

ти нейросети существенно выше возможностей отдельных нейронов.

Выше описанная модель нейрона пренебрегает многими известными характеристиками биологического прототипа, которые некоторые исследователи считают критическими. Например, в ней не учитывают нелинейность пространственно-временной суммации, которая особенно существенна для сигналов, приходящих по возбуждающим и тормозящим синапсам; различного рода временные задержки; эффекты синхронизации и частотной модуляции; рефлекторность и т. п. Невзирая на эти отклонения, сети, построенные на основе таких нейронов, демонстрируют ассоциативные свойства, напоминающие свойства биологических систем, и успешно применяются на практике.

### Многослойная нейронная сеть

Как и человеческий мозг, нейросеть состоит из множества простейших элементов — нейронов, взаимодействующих друг с другом. Именно их совокупность обеспечивает уникальность свойств нейронных сетей, таких как способность к обобщению, работа с зашумленными и неполными данными и др.

Среди всего разнообразия архитектур нейронных сетей наибольшее распространение получили *многослойные*. Фактически эти сети стали «классическими», и когда в литературе говорится о нейросетях, обычно подразумевают многослойные полно связанные сети. Поэтому рассмотрим их подробнее и приведем примеры использования.

В такой архитектуре нейроны объединены в так называемые *слои*. Под слоем понимается совокупность нейронов, входы которых соединены с выходами нейронов предыдущего слоя, а в свою очередь, выходы нейронов этого слоя являются входами для следующего слоя. Обычно слоистые сети являются полно связанными, то есть входы каждого нейрона слоя связаны со всеми выходами нейронов предыдущего слоя. Пример двухслойной полносвязанной нейросети приведен на рис. 3. В нейропрограммах пользователь может легко задать требуемую структуру нейронной сети (рис. 4), определив для этого лишь число слоев и

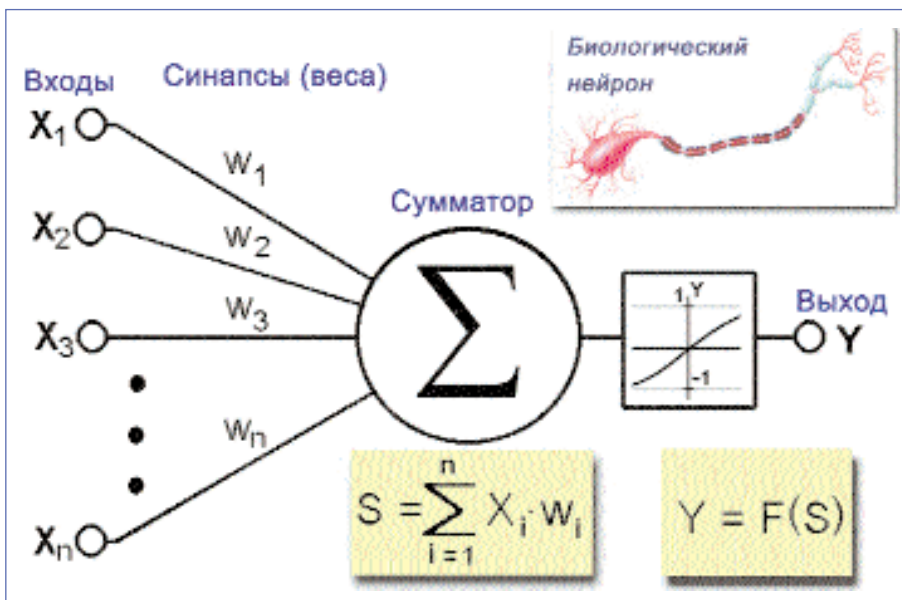


Рис. 1. Модель нейрона

ные сигналы нейросети. Каждый входной сигнал умножается на соответствующий *вес связи*  $w_1, w_2, \dots, w_n$  — аналог эффективности синапса. Вес этой связи является положительным для возбуждающей и отрицательным для тормозящей связи нейрона. Взвешенные весами связей, входные сигналы поступают на блок суммации (по своей сути являющийся аналогом тела клетки в биологическом нейроне), где вычисляется их алгебраическая сумма и определяется уровень возбуждения нейрона  $S$ :

$$S = \sum_{i=1}^n x_i w_i + \theta,$$

где:  $\theta$  — порог возбуждения нейрона. Затем над уровнем возбуждения  $S$  произ-

ных уровнях сигнала происходит насыщение активационной функции<sup>2</sup>, и таким образом, нейрон функционирует в широком диапазоне входных сигналов. Коэффициент  $\beta$  определяет крутизну сигмоида.

Нелинейность функции активации  $F(S)$  очень важна и принципиальна; если бы нейроны были линейными элементами (использовали линейную функцию активации), то любая последовательность нейронов также производила бы линейное преобразование и вся нейросеть была бы эквивалентна одному нейрону (или одному слою нейронов — в случае нескольких выходов). Нелинейность разрушает линейную суперпозицию и приводит к тому, что возможнос-

<sup>1</sup> Парадигма (здесь) — исходная концептуальная схема нейронной сети (базовый вариант).

<sup>2</sup> Насыщение активационной функции — состояние нейрона, при котором значительное изменение параметров входа приводит к очень малым изменениям параметров выхода.

нейронов в них. Практическое распространение таких сетей объясняется хорошо отработанными алгоритмами обучения и широким классом решаемых ими задач.

классификации, при которой выделяемые классы можно разделить гиперплоскостями (для нейросети с двумя входами — прямыми линиями).

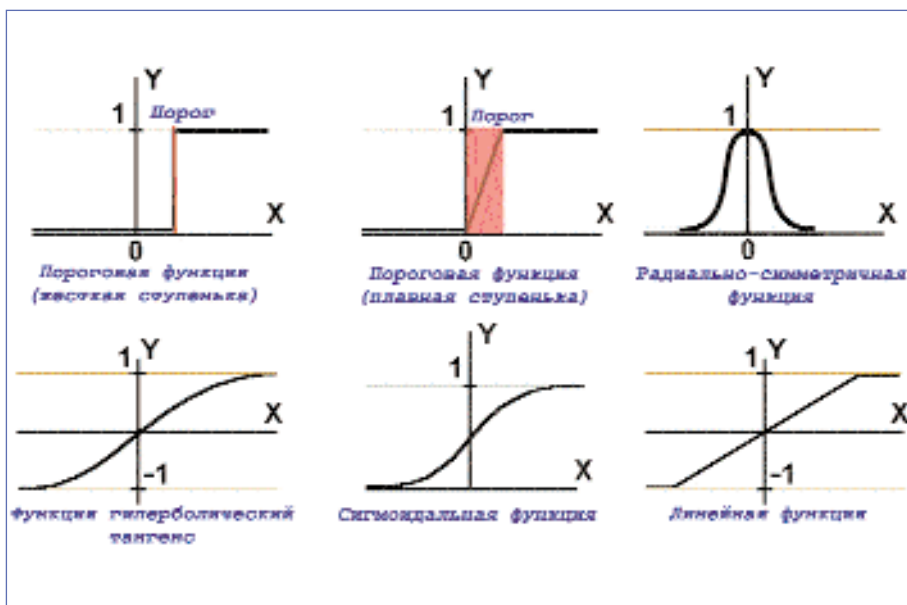


Рис. 2. Функции активации

Не полносвязанные нейросети являются частным случаем. Отсутствие связи эквивалентно нулевому коэффициенту веса связи.

Чаще применяются нейросети имеющие два и более слоев. Однослойные нейронные сети имеют один существенный недостаток, они не решают задачу

### Обучение нейросети

Одно из важнейших свойств нейронной сети — ее способность к самоорганизации, самоадаптации с целью улучшения качества функционирования. Эта цель достигается путем обучения сети, алгоритм задается набором пра-

(веса синапса) между одновременно активными нейронами. Часто используемые связи усиливаются, и этим объясняется феномен обучаемости нейросетей путем повторения и привыкания. В настоящее время выработано достаточное количество разнообразных обучающих правил (алгоритмов обучения).

Все это многообразие можно разделить на две условные группы — «обучение без учителя» и «обучение с учителем». При «обучении без учителя» нет эталонных выходных значений, и нейросеть сама обучается на входных примерах, выявляя в них закономерности и проводя самоорганизацию. Наиболее известной архитектурой такой нейросети являются самоорганизующиеся карты Кохонена (Self-Organizing Maps), о которых будет рассказано в дальнейшем.

В процессе «обучения с учителем» на входы нейросети подаются входные значения и затем производится сравнение полученных выходных значений с требуемыми. Фактически требуемые (эталонные) значения выходов играют роль «учителя», каждый раз давая урок нейросети. Рассматриваемые в этой статье многослойные сети обучаются по этому принципу.

Когда в сети есть только один слой, то алгоритм ее обучения с учителем довольно очевиден, так как правильные выходные состояния нейронов единственного слоя заведомо известны и подстройка весов (синаптических связей) идет в направлении минимизации ошибки на выходе. В многослойных же сетях оптимальные выходные значения нейронов всех слоев, кроме последнего, как правило, не известны, и их уже невозможно обучить, руководствуясь только величинами ошибок на выходах НС.

Тренировка многослойных нейронных сетей проводится с помощью алгоритма обратного распространения ошибки (*back propagation*). Этот алгоритм является обобщением одной из процедур обучения простого персептрона, известной как дельта-правило, и требует наличия обучающей выборки. Выборка состоит из набора пары векторов (входной и выходной), между которыми надо установить соответствие.

Алгоритм обратного распространения ошибки является одной из разновидностей градиентного спуска, минимизирующий суммарную квадратичную ошибку:

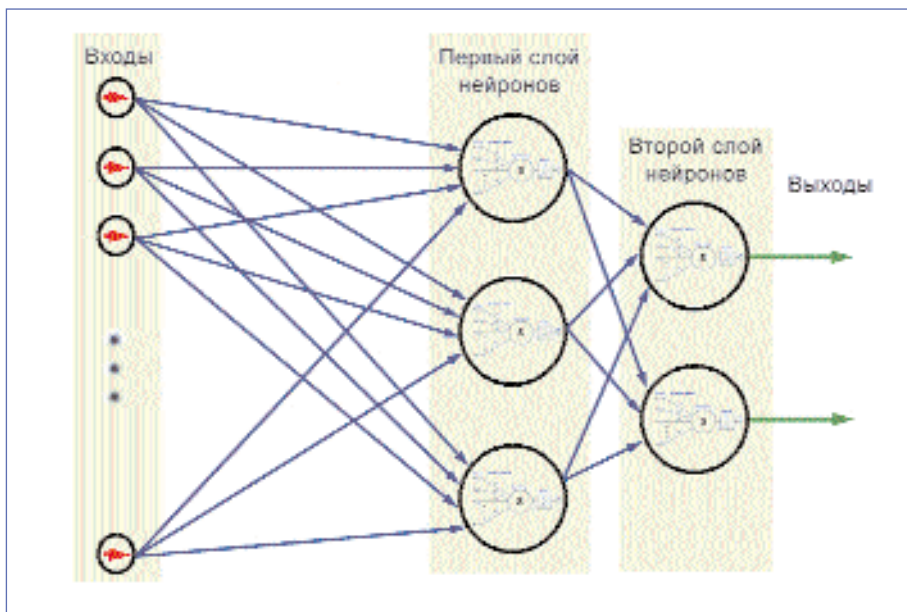


Рис. 3. Двуслойная нейронная сеть

классификации линейно неразделимых классов, которая в литературе носит название задачи «исключающего ИЛИ». Линейно разделимая задача — задача

вил. Обучающие правила определяют, каким образом изменяются связи в ответ на входное воздействие. Обучение основано на увеличении силы связи



$$E = \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^I (y_{ik}^{[N]} - Y_{ik})^2$$

где:  $y_{ik}^{[N]}$  — реальное выходное состояние нейрона  $i$ -го выходного слоя  $N$  нейронной сети при подаче на ее входы  $k$ -го образа;  $Y_{ik}$  — идеальное (желаемое) выходное состояние этого нейрона. Идея алгоритма состоит в том, чтобы вычислить чувствительность выхода нейрона к изменению весов в случае возникнове-

Перед началом обучения связям присваиваются небольшие случайные значения. Каждая итерация процедуры состоит из двух фаз. Во время первой фазы (прямой проход) на сеть подается входной вектор путем установки в нужное состояние входных элементов. Затем входные сигналы распространяются по сети, порождая выходной вектор.

Полученный выходной вектор сравнивается с требуемым (обратный проход). Если они совпадают, обучения не происходит. В противном случае вычисляется

- в процессе обучения может возникнуть ситуация, когда большие положительные или отрицательные значения весов переведут нейрон в область насыщения активационной функции;
- применение метода градиентного спуска не гарантирует, что будет найден глобальный, а не локальный минимум целевой функции;
- нет никаких предварительных оценок требуемого времени для обучения нейросети.

Наличие этих и другие, проблем подвигло разработчиков, в свою очередь, к созданию большого числа улучшенных алгоритмов обратного распространения ошибки, таких как использование производных второго порядка, пакетной обработки и т. д.

Отдельной проблемой при обучении с учителем является переобучение нейросети. Под переобучением понимается такое обучение, при котором теряются обобщающие свойства и нейросеть просто «запоминает» выходные данные. Другими словами на обучающей выборке сеть показывает великолепные результаты, а на новых данных или на подтверждающей выборке, результаты гораздо хуже. На рис. 5 показан момент наступления переобучения нейросети. При дальнейшем уменьшении ошибки на обучающей выборке на подтверждающей выборке ошибка увеличивается.

Основным методом «борьбы» с переобучением является введение механизма калибровки в алгоритм обучения. Калибровка подразумевает выделение из обучающего множества — специального калибровочного (подтверждающего). Процесс обучения проводится на обучающем множестве, а мониторинг и принятие решения об остановке обучения принимается по результатам прогона через нейросеть калибровочного множества примеров.

В большинстве случаев ошибка прогноза нейросети складывается из двух основных составляющих. Первая из них определяется недостаточным качеством аппроксимации и наличием «шума» во входных данных, вторая — недостатком информации, поданной на вход нейросети при обучении. Недостаток информации обычно вызван либо отсутствием во входах важных для прогноза переменных или/и малым размером обучающей выборки, не отражающей всей полноты картины.

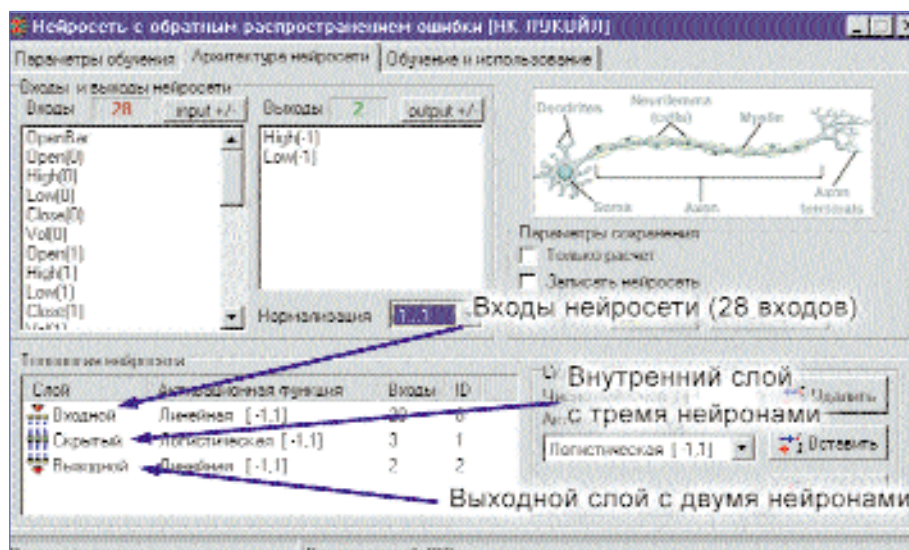


Рис. 4. Создание нейросети в программе NeuroScalp

ния ошибки. Для этого вычисляются частные производные от ошибки по весам. Для работы алгоритма требуется, чтобы активационные функции нейронов были неубывающими и имели ограниченную производную. При проведении данной операции обычно используют сигмоидальную активационную функцию (рис. 2.).

разница между фактическими и требуемыми выходными значениями, которая передается последовательно от выходного слоя к входному. На основании информации об ошибке производится модификация весов связей.

Описанный алгоритм хотя и является широко распространенным, но все же имеет некоторые недостатки:

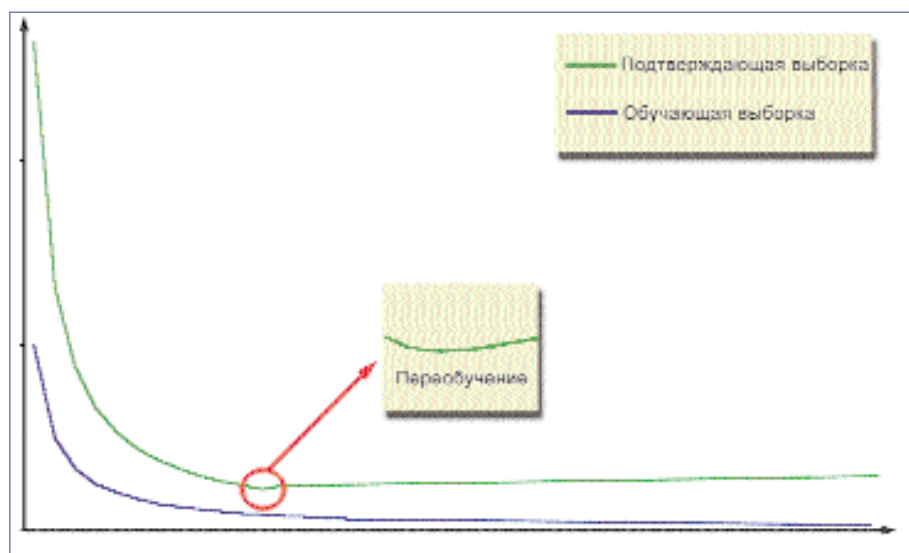


Рис. 5. Обучение нейронной сети

### **Задачи, решаемые нейронными сетями**

Нейронные системы представляют собой не что иное, чем то, что мы привыкли подразумевать под компьютерами. Нейронная сеть используется для изучения

образцов данных, выявления их взаимосвязей и закономерностей. Входными данными могут быть, например, котировки ценных бумаг на определенный день на некоторой бирже или же результаты производственного процесса при разных производственных условиях. Независимо от

типа данных или специфики наблюдаемого явления нейронная сеть является неким абстрактным методом решения, в отличие от традиционных.

Первоначально нейронные сети применялись в задачах классификации, перцептрон Ф. Розенблата решал задачу класси-

мации. В таблице приведены некоторые прикладные экономические задачи, решаемые нейросетями сегодня.

## Возможность прогнозирования временных рядов

Для трейдинга наиболее простой, из перечисленных задач и не менее актуальной является задача предсказания цен активов и их анализ. Как известно, существуют два основных подхода к анализу рынка — технический и фундаментальный. Первый из них базируется на теории Доу, который выдвинул еще в начале 20 века лозунг: «Цены учитывают все», и, соответственно, технический аналитик использует только цены актива и различные индикаторы (функции цен). Фундаментальный анализ, наоборот, ищет взаимосвязь цен актива и внешних событий и данных типа макроэкономических показателей и финансовой отчетности корпораций и т. д.

Применение нейронных сетей в качестве дополнения дает уникальную возможность объединить эти два метода. Такой анализ, в отличие от технического, не имеет никаких ограничений по характеру входной информации. Это могут быть как индикаторы данного временного ряда, так и сведения о поведении других рыночных инструментов, и внешние события. Недаром нейросети активно используют на западе именно крупные консервативные инвесторы (например, пенсионные фонды и страховые компании), работающие с большими портфелями, для которых особенно важны корреляции между различными рынками и которые тяготеют к фундаментальному анализу.

В отличие от теханализа, основанного на общих рекомендациях и опыте трейдера, нейросети способны находить оптимальные для данного актива индикаторы и строить по ним оптимальную модель прогнозирования, более того, она (модель) адаптивна, и меняется вместе с рынком, что особенно важно для современных высоко динамичных финансовых рынков. ■

фикации графических изображений букв. Это объясняется ступенчатой активационной функцией нейронов. В дальнейшем, с введением в нейрон непрерывных активационных функций (рис. 2) появилась возможность решения обширного класса задач прогнозирования и апрокси-

Решаемая задача	Тип задачи	Обучение
Прогнозирование цен активов и их индикаторов	Временные ряды	Обучение с учителем
Составление рейтингов	Классификация	Обучение без учителя
Нейронные торговые системы	Классификация	Обучение с учителем
Определение «кредитоспособности» заемщика	Классификация	Обучение с/без учителем
Макроэкономический анализ	Смешанный	Обучение с/без учителем
Оптимизация портфеля	Смешанный	Обучение с учителем
Безопасность транзакций по пластиковым карточкам	Смешанный	Обучение с/без учителем