

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение
высшего образования

«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ НАУК

Кафедра фундаментальной математики и механики

РЕКОМЕНДОВАНО К ЗАЩИТЕ В ГЭК

Заведующий кафедрой

 к.ф.-м.н.
Девятков А.П.
27.06 2022 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
магистерская диссертация
НЕЙРОННАЯ СЕТЬ ДЛЯ ПРОГНОЗА ЗАБОЙНОГО ДАВЛЕНИЯ

01.04.01 Математика

Магистерская программа «Вычислительная механика»

Выполнили работу
Студенты 2 курса
очной формы обучения

подпись



Мизернова
Екатерина
Сергеевна
Наумова
Кристина
Эдуардовна

подпись



Научный руководитель
д.ф.-м.н.,
профессор

подпись



Родионов
Сергей
Павлович

Рецензент
к.ф.-м.н.,
Главный инженер проекта,
ООО «ПИТЦ «Геофизика»

подпись



Марков
Павел
Владимирович

Тюмень, 2022

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	4
СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ	6
СПИСОК ТЕРМИНОВ	8
ГЛАВА 1. НАУЧНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА	12
1.1. ОПРЕДЕЛЕНИЕ НЕЙРОННЫХ СЕТЕЙ И ИХ ВИДЫ.....	12
1.1.1 Определение нейронных сетей	13
1.1.2 Виды нейронных сетей	14
1.1.3 Виды функций активаций.....	16
1.1.4 Методы обучения нейронной сети.....	20
1.1.5 Метод градиентного спуска	21
1.1.6 Метод имитации отжига.....	24
ГЛАВА 2. ОПИСАНИЕ НЕЙРОННОЙ СЕТИ ДЛЯ ПРОГНОЗА ЗАБОЙНОГО ДАВЛЕНИЯ	27
2.1. ПОДГОТОВКА ДАННЫХ ДЛЯ НЕЙРОННОЙ СЕТИ.....	27
2.2. СОЗДАНИЕ И ОБУЧЕНИЕ НЕЙРОННОЙ СЕТИ ПО ЗАБОЙНОМУ ДАВЛЕНИЮ	28
2.3. КРИТЕРИИ АДАПТАЦИИ ПО ЗАБОЙНОМУ ДАВЛЕНИЮ	31
ГЛАВА 3. ТЕСТИРОВАНИЕ НЕЙРОННОЙ СЕТИ НА ИДЕАЛЬНОМ ПРИМЕРЕ ДАННЫХ.....	32
3.1. ОБЩИЕ СВЕДЕНИЯ ОБ ОБЪЕКТЕ ТЕСТИРОВАНИЯ	32
3.2. ПОДГОТОВКА ИСХОДНЫХ ДАННЫХ	33
3.3. АДАПТАЦИЯ ПРОКСИ-МОДЕЛИ ПО ЗАБОЙНОМУ ДАВЛЕНИЮ..	36

ГЛАВА 4. ТЕСТИРОВАНИЕ АЛГОРИТМА НА РЕАЛЬНОМ ПРИМЕРЕ ДАННЫХ.....	44
4.1. ОБЩИЕ СВЕДЕНИЯ ОБ ОБЪЕКТЕ ТЕСТИРОВАНИЯ	44
4.2. ПОДГОТОВКА ИСХОДНЫХ ДАННЫХ	45
4.3. АДАПТАЦИЯ ПРОКСИ-МОДЕЛИ ПО ЗАБОЙНОМУ ДАВЛЕНИЮ..	47
ЗАКЛЮЧЕНИЕ	50
БИБЛИОГРАФИЧЕСКИЙ СПИСОК	51
СПИСОК ИЛЛЮСТРАТИВНОГО МАТЕРИАЛА	53
ПРИЛОЖЕНИЯ.....	55
Приложение.....	55
Акт о внедрении	55

ВВЕДЕНИЕ

Для управления заводнением и оптимизации режимов работы эксплуатации добывающих и нагнетательных скважин на месторождениях применяют нейронные сети (НС). Чаще всего используется прокси-модель, работа, которой основывается на математической модели взаимовлияния добывающих и нагнетательных скважин (элемент заводнения). Принцип функционирования НС заключается в определении взаимосвязи между параметрами системы (дебитом жидкости, обводненности, приемистости и забойным давлением) и подборе настроечных весов для воспроизведения фактической динамики работы скважин. Благодаря этому, появляется возможность принимать оперативные решения без построения крупномасштабной гидродинамической модели.

Для измерения параметров системы непосредственно на насосе применяются различные датчики, закрепляемые в составе глубинно-насосного оборудования. Такие датчики называются термоманометрическими системами (ТМС). Датчики ТМС предназначены для измерения и передачи на контроллер станции управления, текущих параметров работы установки электроцентробежного насоса (УЭЦН), в добывающих и нагнетательных скважинах, а именно:

- температуры окружающей (пластовой) жидкости;
- давление пластовой жидкости в районе подвески УЭЦН;
- забойное давление.

Одним из важнейших показателей, определяющим эффективность эксплуатации системы «скважина – пласт» и подлежащим непрерывному мониторингу, является забойное давление. Под забойным давлением понимается давление на забое скважины, которое замеряется во время установившейся работы скважины. Ему соответствует динамический уровень, то есть уровень жидкости в скважине.

Уровень жидкости в добывающей скважине можно контролировать с помощью системы поддержания пластового давления (ППД), то есть при закачивании воды в пласт через нагнетательные скважины.

В условиях сокращения объема закачиваемой воды стоит проблема снижения забойного давления добывающей скважины ниже давления насыщения, приводящая к уменьшению уровня жидкости или выделению газа в скважине, что в свою очередь, несет риски отказа насосного оборудования [3-4].

Нейронная сеть по забойному давлению позволяет предусматривать ограничения работы насосного оборудования в добывающей скважине и предотвратить риски возникновения аварий.

Таким образом, целью настоящей работы является создание нейронной сети для контроля и прогноза забойного давления.

Для достижения поставленной цели были определены следующие задачи:

1. Обзор литературы соответствующей выбранной теме;
2. Анализ существующих нейронных сетей и методов адаптации;
3. Создание и обучение нейронной сети для прогноза забойного давления (подбор функции активации и настроечных параметров);
4. Тестирование нейронной сети на гидродинамической модели (идеальном примере данных);
5. Тестирование нейронной сети на реальном примере данных.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

$p_{j,t}^n$ – забойное давление добывающей скважины j на дату обучения t , [д.ед.]

$P_j^n(q_{i,t}^{inj})$ – функциональная зависимость забойного давления добывающей скважины j от приемистости нагнетательной скважины i на дату обучения t , [д.ед.]

$q_{i,t}^{inj}$ – приемистость нагнетательной скважины i на дату обучения t , [м³/сут]

$q_{i,t-1}^{inj}$ – приемистость нагнетательной скважины i на дату обучения $t - 1$, [м³/сут]

$c_{i,j}^{dec}$, $d_{i,j}^{dec}$, $c_{i,j}^{gr}$, $d_{i,j}^{gr}$ – настроечные параметры модели, которые подбираются в процессе адаптации, [ед.]

c_{dec}^{min} , c_{dec}^{max} – минимальное и максимальное значения $c_{i,j}^{dec}$, [ед.]

d_{dec}^{min} , d_{dec}^{max} – минимальное и максимальное значения $d_{i,j}^{dec}$, [ед.]

c_{gr}^{min} , c_{gr}^{max} – минимальное и максимальное значения $c_{i,j}^{gr}$, [ед.]

d_{gr}^{min} , d_{gr}^{max} – минимальное и максимальное значения $d_{i,j}^{gr}$, [ед.]

g_*^{max} – верхнее ограничение настроечных параметров, [ед.]

g_*^{min} – нижнее ограничение настроечных параметров, [ед.]

$date^{train_begin}$ – дата начала обучающей выборки, [ДД.ММ.ГГГГ]

$date^{train_end}$ – дата окончания обучающей выборки, [ДД.ММ.ГГГГ]

N – количество элементов обучающей выборки, [ед.]

$i = \overline{1 \dots M^{react}_j}$, M^{react}_j – количество нагнетательных скважин, действующих на одну добывающую скважину j , [ед.]

$Err^{общая}$ – ошибка обучения нейронной сети на всей выборке, [%]

$Err^{предпрогноз}$ – ошибка обучения нейронной сети на предпрогнозной выборке, [%]

$P_{j,t-1}$ – фактическое значение забойного давления добывающий скважины на шаге $t-1$, [атм]

$q^{inj}_{i,t}$ – значение приемистости влияющей нагнетательной скважины на шаге t , [м³/сут]

X_k – значение вектора на итерации k

n – количество компонент вектора X_k

$step_i$ – потенциальный множитель, по i -ой компоненте градиента

x_i – значение i -ой компоненты текущего решения

$grad(X)_i$ – градиент i -ой компоненты текущего решения

e – константа уменьшения множителя

α_i – независимые случайные величины, распределенные равномерно на $[0,1]$

Y – последовательность псевдослучайных значений

a, c, m – целочисленные константы которые определяют ГПСЧ

ГПСЧ – генератор псевдослучайных чисел

НС – нейронная сеть

ГНО – глубинно-насосное оборудования

ТМС – термоманометрическая система

ЭЦН – электроприводный центробежный насос

ППД – поддержания пластового давления

ПЭД – погружной электродвигатель

СПИСОК ТЕРМИНОВ

Месторождение: природное скопление минерального вещества (полезного ископаемого) на поверхности или в недрах Земли. Месторождения могут заключать газовые, жидкие, твёрдые полезные ископаемые.

Пласт: масса горной породы, ограниченная с двух сторон от других слоёв.

Скважина: горная выработка круглого сечения с диаметром во много раз меньше длины, образуемая в массиве горных пород путем бурения и крепления без доступа в нее человека, с заранее заданным положением в пространстве. Начало скважины называется устьем, дно — забоем.

Добывающая скважина: предназначена для добычи нефти и газа.

Нагнетательная скважина: предназначена для нагнетания воды (газа) в продуктивные пласты при осуществлении методов поддержания давления.

Динамика: исходные параметры скважин.

Временной ряд: последовательность зарегистрированного сигнала.

«Шахматки»: промысловый журнал для регистрации всех событий и инструментальных измерений на скважинах.

Промысловые показатели скважины: приемистость, забойное давление, динамический уровень.

Дебит жидкости: объём жидкости, стабильно поступающий в ед. времени.

Обводненность: отношение добываемой воды к общему количеству добываемой жидкости за период (год, месяц).

Приемистость: характеристика нагнетательной скважины, показывающая возможность закачки рабочего агента в пласт.

Динамический уровень: установившийся постоянный уровень воды в скважине при работе скважины.

Забойное давление: давление на забое работающей нефтяной, газовой или водной скважины.

Пластовое давление: давление, под которым находятся жидкость (нефть, вода) и газ, насыщающие поровое пространство и (или) трещины коллекторов нефтяных и газовых месторождений.

Давление насыщения: давление, при котором газ начинает выделяться из нефти.

Пластовая жидкость: смесь флюидов, находящихся в пласте.

Система поддержания пластового давления: комплекс технологического оборудования необходимый для подготовки, транспортировки, закачки рабочего агента в пласт нефтяного месторождения с целью поддержания пластового давления.

Вязкость воды: свойство жидкости, обуславливающее при движении возникновение силы трения.

Вязкость нефти: свойство, определяющее ее подвижность в пластовых условиях для добычи или при транспортировке по магистральным нефтепроводам.

Сжимаемость породы: способность горных пород изменять свой объём под воздействием всестороннего давления.

Сжимаемость воды: способность воды изменять свой объём под воздействием всестороннего давления.

Сжимаемость нефти: характеризует относительное приращение объема нефти при изменении давления на единицу.

Нейронная сеть: математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма.

Синапс: место контакта между двумя нейронами или между нейроном и получающей сигнал эффективной клеткой.

Прокси-модель: математический аппарат, который применяется для изучения и управления сложными процессами на основе технологий машинного обучения - искусственная нейронная сеть (ИНС) и алгоритмов решения оптимизационных задач.

Элемент заводнения: совокупность пар скважин, состоящая из реагирующей добывающей скважины и окружающих ее влияющих нагнетательных скважин.

Пара скважин: совокупность реагирующей добывающей и влияющей на нее нагнетательной скважины.

Влияющая скважина: нагнетательная скважина, входящая в элемент заводнения добывающей скважины, которая оказывает влияние на режим работы последней.

Реагирующая скважина: добывающая скважина, на которой регистрируется изменение показателей эксплуатации, например, увеличение обводненности или дебита жидкости, вследствие изменения режима работы нагнетательной скважины.

Коэффициент взаимовлияния: характеризующий отклик в динамике показателей эксплуатации добывающей скважины на существенные изменения режима нагнетательной скважины.

Обучающая выборка: набор исходных данных, по которым производится настройка (адаптация) модели. В нашем случае – это совокупность приемистостей нагнетательных скважин в элементе заводнения, соотнесенная по дате с дебитом жидкости и долей нефти реагирующей добывающей скважины.

Предпрогнозная выборка: значения обучающей выборки, предшествующие прогнозу.

Адаптация прокси-модели: процесс автоматической настройки рассчитанного в прокси-модели забойного давления на фактические показатели.

Фильтрация: этап предварительной обработки данных, цель которого очищение исходной динамики от шумов.

Нормировка: корректировка ряда значений в соответствии с некоторыми функциями преобразования, с целью сделать их более удобными для сравнения.

Весовой коэффициент: синоптическая связь между входными и выходными данными нейронной сети.

Функция активации: функция, которая преобразует взвешенную сумму входных сигналов в число, которое определяет выходной сигнал нейрона.

Функциональная зависимость: результат адаптация прокси модели.

Глобальный минимум: экстремум функции в точке, значение которой минимально на всей области определения.

Настроечный период: период обучения прокси-модели.

Временной лаг: показатель, отражающий отставание или опережение во времени одного явления по сравнению с другими.

ГЛАВА 1. НАУЧНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА

1.1. ОПРЕДЕЛЕНИЕ НЕЙРОННЫХ СЕТЕЙ И ИХ ВИДЫ

Нейронные сети часто воспринимаются как некая инновационная технология, часть мира будущего. Однако прорыв произошел относительно недавно.

Началом работы ученых в этом направлении можно считать статью Уорена Маккалока и Уолтера Питтса, вышедшую в 1943 году. Они разработали компьютерную модель нейронной сети, опираясь на математические алгоритмы и теорию деятельности головного мозга.

Следующим шагом стала книга Дональда Хебба «Организация поведения» 1949 года, в которой нейропсихолог из Канады описал процесс самообучения искусственной нейронной сети.

Спустя 8 лет Фрэнком Розенблаттом была представлена математическая модель обработки информации человеческим мозгом, получившая название персептрон. Еще через три года, в 1960, этот американский ученый продемонстрировал электронное устройство, которое имитировало работу мозга.

В 1969 году идея создания нейронных сетей была жестко раскритикована Минским и Пейпертом. Они указали на серьезные проблемы, препятствующие эффективному использованию искусственных нейронных сетей. Однако это направление по-прежнему представлялось весьма перспективным многим исследователям, продолжавшим попытки сформулировать принцип работы самообучаемой системы. Метод обратного распространения ошибки, предложенный в 1974 году Полом Вербосом, стал еще одним шагом на пути создания нейронной сети. Через год Фукусима разработал первую многослойную нейронную сеть – когнитрон. В 1982 году Хопфилд добился двусторонней передачи информации между нейронами,

что еще больше подогрело интерес ученых к поискам новых решений в этой отрасли.

С начала 1990-х были сделаны важнейшие шаги в деле создания нейронных сетей, а в 2007 году Джеффри Хинтон ввел понятие глубокого обучения, что позволило приступить к использованию возможностей нейросетей в утилитарных целях [1].

1.1.1 Определение нейронных сетей

Искусственная нейронная сеть – это целый комплекс параллельных вычислений, состоящий из множества простых взаимодействующих процессоров. Процессоры – это только простые элементы нейронной сети, каждый из которых лишь получает определенный сигнал от другого процессора и посылает обработанный сигнал другим процессорам. Вместе такие элементарные процессоры могут выполнять весьма сложные задачи [2].

Любая нейронная сеть состоит из элементов. Эти элементы соединены между собой определенным образом для обеспечения взаимодействия. Элементы нейронной сети – это процессоры, которые называются также нейронами (узлами). Вычислительные способности таких узлов (нейронов) невелики и ограничены лишь вычислениями входных сигналов (по некоторому правилу их комбинирования) и выходным сигналом (по правилу активации – правилу обработки входных сигналов). Элемент нейронной сети представлен на рисунке 1.

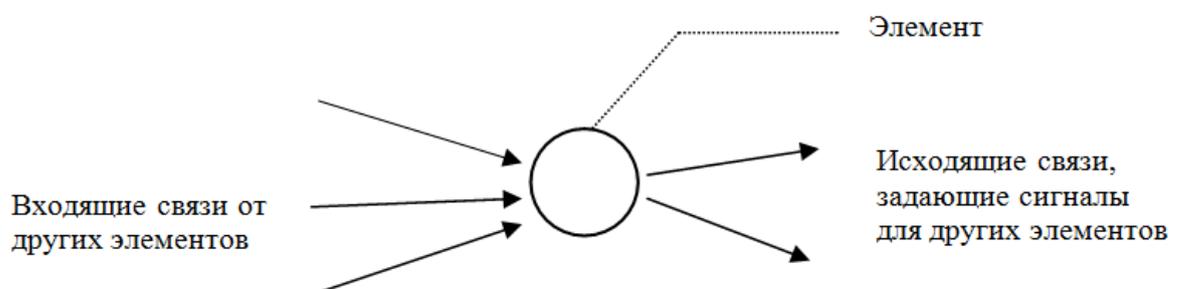


Рисунок 1– Элемент нейронной сети

Выходной сигнал посылается другим элементам по соответствующим связям, каждая из которых имеет определенный весовой коэффициент (вес), т.е. выходной сигнал умножается на соответствующее значение веса связи и поступает на вход к другому элементу. Согласно значению весового коэффициента, выходной сигнал либо усиливается, либо подавляется.

В зависимости от того, каким образом соединены между собой элементы нейронной сети, выделяют различные структуры связей. Структура связи отражает конструкцию сети, какие элементы соединены, в каком направлении работают связи, вес каждого соединения (см. рисунок 2). При формировании структуры связей сначала определяют какие элементы и в каком направлении должны быть связаны, и только потом связям задают соответствующие весовые коэффициенты.

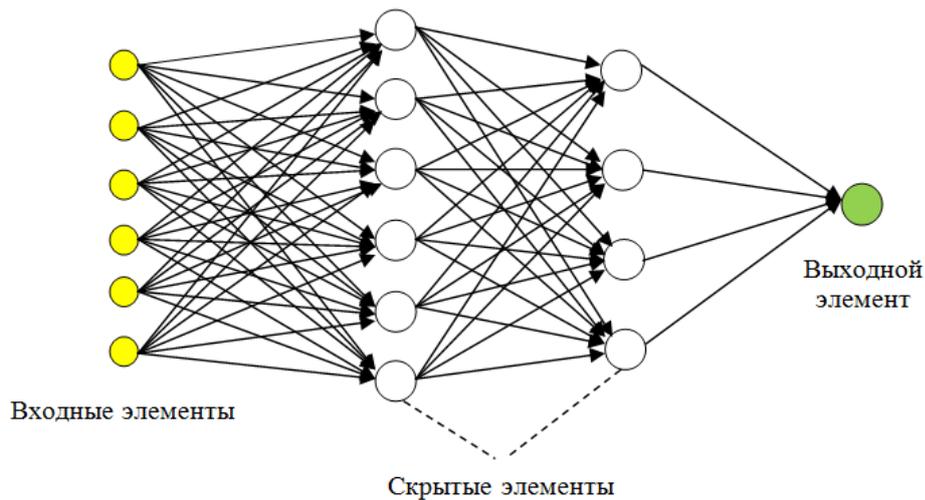


Рисунок 2 – Структура связей нейронной сети

Структура связей нейронной сети, представленной на рисунке 2, состоит из четырех слоев: слой входных элементов (шесть элементов), слой скрытых элементов (два слоя по шесть и четыре элемента соответственно) и слой выходных элементов (1 элемент).

1.1.2 Виды нейронных сетей

Любая искусственная нейронная сеть содержит первый, или входной, слой нейронов, который принимает сигналы и распределяет их по другим нейронам. Этот признак объединяет все сети, а разные типы выделяются в

зависимости от того, какая её дальнейшая структура и по какому принципу различные слои нейронной сети взаимодействуют между собой.

Однослойная нейронная сеть. Представляет собой структуру взаимодействия нейронов, в которой сигналы с входного слоя сразу направляются на выходной слой, который, в конечном счёте, не только преобразует сигнал, но и сразу же выдаёт ответ. Первый входной слой только принимает и распределяет сигналы, а необходимые вычисления происходят уже во втором (выходном) слое. Входные нейроны являются объединёнными с основным слоем при помощи синапсов с разными весами, обеспечивающими качество связей.

Многослойная нейронная сеть. В отличие от однослойной, помимо выходного и входного слоёв, имеются ещё несколько скрытых промежуточных. Таким образом, структура взаимодействия нейронов следующая: сигналы с входного слоя поступают на промежуточные, где выполняются необходимые расчёты, а затем выходной слой представляет полученный результат. Число скрытых слоёв зависит от степени сложности нейронной сети. Соответствующие решения обладают большими возможностями, если сравнивать с однослойными, ведь в процессе обработки данных каждый промежуточный слой — это промежуточный этап, на котором осуществляется обработка и распределение информации.

Кроме количества слоёв, нейронные сети можно классифицировать по направлению распределения информации по синапсам между нейронами.

Нейросети прямого распространения (однонаправленные). В такой структуре сигнал перемещается строго по направлению от входного слоя к выходному. Движение сигнала в обратном направлении не осуществляется и в принципе невозможно.

Рекуррентные нейронные сети (с обратными связями). Здесь сигнал двигается как в прямом, так и в обратном направлении. По итогу, результат выхода способен возвращаться на вход. Выход нейрона определяется весовыми характеристиками и входными сигналами, плюс дополняется

предыдущими выходами, снова вернувшись на вход. Этим нейронным сетям свойственна функция кратковременной памяти, на основании чего сигналы восстанавливаются и дополняются во время их обработки.

1.1.3 Виды функций активаций

Функцией активации называется функция, которая преобразует взвешенную сумму входных сигналов и порогового значения в число, которое определяет выходной сигнал нейрона.

Основными и самыми распространенными являются следующие функции активации:

Пороговая функция

Пороговая функция активации меняет свое значение скачкообразно, когда аргумент функции достигает значения порога нейрона. Используют два варианта пороговой функции — пороговую биполярную и пороговую бинарную функцию активации.

На рисунке 3 приведён пример пороговой биполярной функции, изменяющей свое значение от -1 до 1 .

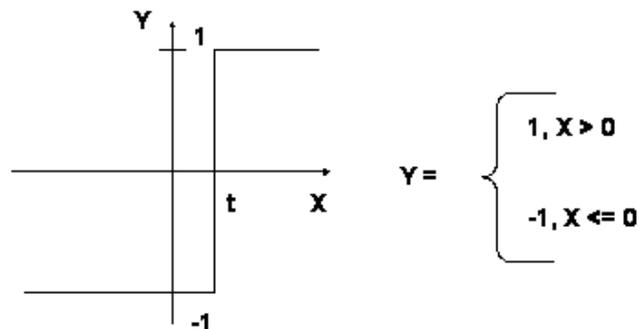


Рисунок 3 – Пороговая биполярная функция активации

Помимо биполярной пороговой функции активации применяется так называемая бинарная пороговая функция активации. Она может принимать значения 0 или 1, как это показано на рисунке 4.

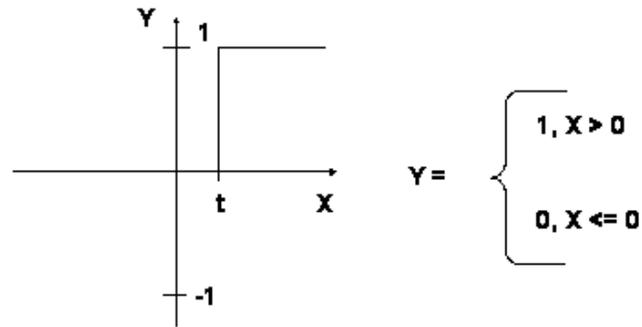


Рисунок 4 – Пороговая бинарная функция активации

Линейная функция активации

Линейная функция представляет собой прямую линию, то есть $A = cx$, а это значит, что результат этой функции активации пропорционален переданному аргументу. В отличие от предыдущей функции, она позволяет получить диапазон значений на выходе, а не только бинарные 0 и 1, что решает проблему классификации с большим количеством классов. Но у линейной функции есть две основных проблемы:

1. Использование метода обратного распространения ошибки невозможно. Так как в основе этого метода обучения лежит градиентный спуск, а для того чтобы его найти, нужно взять производную, которая для данной функции активации — константа и не зависит от входных значений. То есть при обновлении весов нельзя сказать улучшается ли эмпирический риск на текущем шаге или нет.
2. Рассмотрим нейронную сеть с несколькими слоями и линейной функцией активации. Так как для каждого слоя выходное значение линейно, то они образуют линейную комбинацию, результатом которой является линейная функция. То есть финальная функция активации на последнем слое зависит только от входных значений на первом слое. А это значит, что любое количество слоев может быть заменено всего одним слоем, и, следовательно, нет смысла создавать многослойную сеть.

Главное отличие линейной функции от остальных в том, что ее область значений не ограничена: $(-\infty; +\infty)$. Следовательно, ее можно использовать, когда выходное значение нейрона, принадлежит действительным числам, а не ограниченному интервалу (Рисунок 5).

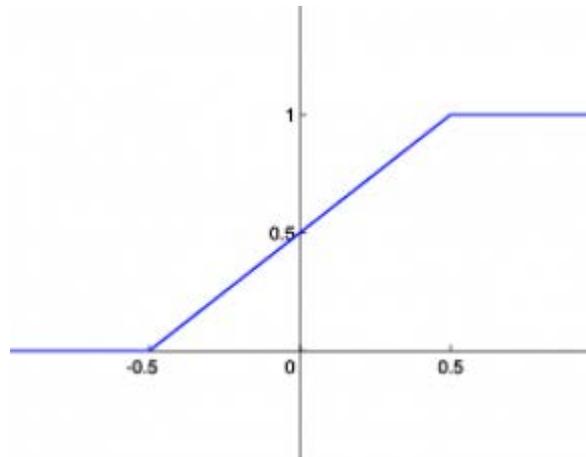


Рисунок 5 - Линейная функция активации

Сигмоидная (логистическая) функция

Сигмоидная функция, которую также называют логистической, является гладкой монотонно возрастающей нелинейной функцией:

$$A = \frac{1}{1+e^{-x}} \quad (1)$$

Так как эта функция не линейна, то ее можно использовать в нейронных сетях с множеством слоев, а также обучать эти сети методом обратного распространения ошибки. Сигмоида ограничена двумя горизонтальными асимптотами $y=1$ и $y=0$, что дает нормализацию выходного значения каждого нейрона (Рисунок 6). Кроме того, для сигмоидной функции характерен гладкий градиент, который предотвращает «прыжки» при подсчете выходного значения. Функции имеет еще одно преимущество, для значений $x > 2$ и $x < -2$, y «прижимается» к одной из асимптот, что позволяет делать четкие предсказания классов.

Несмотря на множество сильных сторон сигмоидной функции, у нее есть значительный недостаток. Производная такой функции крайне мала во всех точках, кроме сравнительно небольшого промежутка. Это сильно усложняет процесс улучшения весов с помощью градиентного спуска. Более

того, эта проблема усугубляется в случае, если модель содержит много слоев. Данная проблема называется проблемой исчезающего градиента.

Что касается использования сигмоидной функции, то ее преимущество над другими — в нормализации выходного значения. Иногда, это бывает крайне необходимо. К примеру, когда итоговое значение слоя должно представлять вероятность случайной величины.

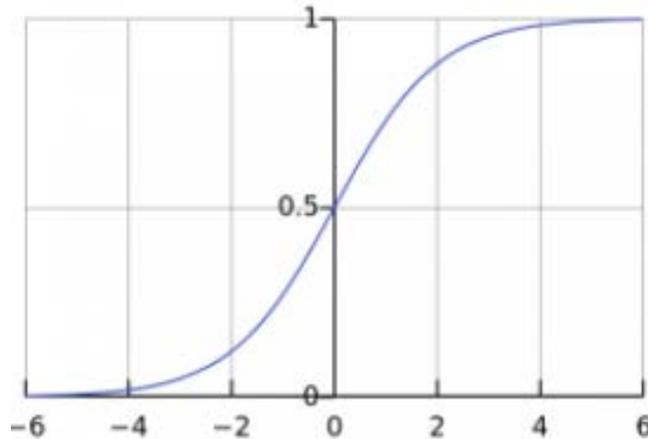


Рисунок 6 - Сигмоидная функция активации

Гиперболический тангенс

Функция гиперболического тангенса имеет вид:

$$\tanh(x) = \frac{2}{1+e^{-2x}} - 1 \quad (2)$$

Она является скорректированной сигмоидной функцией:

$$\tanh(x) = 2 \cdot \text{sigmoid}(2x) - 1 \quad (3)$$

То есть сохраняет те же преимущества и недостатки, но уже для диапазона значений $(-1;1)$.

Обычно, \tanh является предпочтительнее сигмоиды в случаях, когда нет необходимости в нормализации (Рисунок 7). Это происходит из-за того, что область определения данной функции активации центрирована относительно нуля, что снимает ограничение при подсчете градиента для перемещения в определенном направлении. Кроме того, производная гиперболического тангенса значительно выше вблизи нуля, давая большую амплитуду градиентному спуску, а, следовательно, и более быструю сходимость.

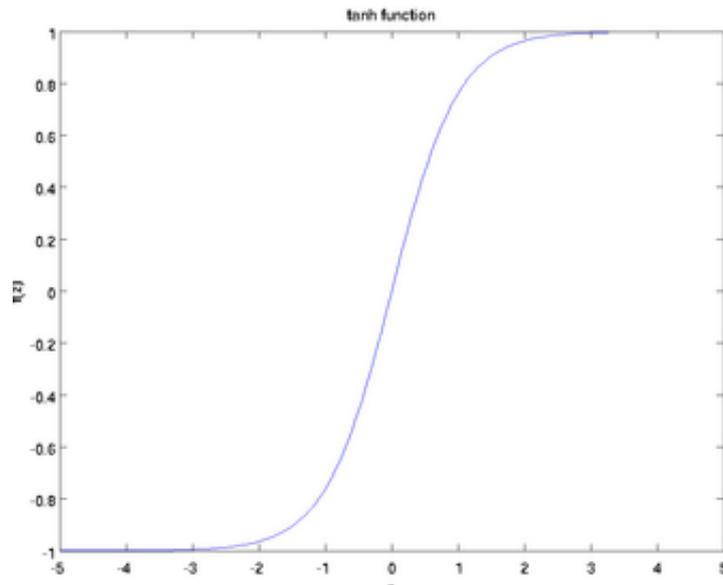


Рисунок 7 - Функция активации гиперболического тангенса

1.1.4 Методы обучения нейронной сети

Обучение нейронной сети – это процесс, в котором параметры нейронной сети настраиваются посредством моделирования среды, в которую эта сеть встроена. Тип обучения определяется способом подстройки параметров.

Для того, чтобы нейронная сеть была способна выполнить поставленную задачу, ее необходимо обучить. Различают алгоритмы обучения с учителем и без учителя.

Процесс обучения с учителем представляет собой предъявление сети выборки обучающих примеров. Каждый образец подается на входы сети, затем проходит обработку внутри структуры НС, вычисляется выходной сигнал сети, который сравнивается с соответствующим значением целевого вектора, представляющего собой требуемый выход сети. Затем по определенному правилу вычисляется ошибка, и происходит изменение весовых коэффициентов связей внутри сети в зависимости от выбранного алгоритма. Векторы обучающего множества предъявляются последовательно, вычисляются ошибки и веса подстраиваются для каждого вектора до тех пор, пока ошибка по всему обучающему массиву не достигнет приемлемо низкого уровня.

При обучении без учителя обучающее множество состоит лишь из входных векторов. Обучающий алгоритм подстраивает веса сети так, чтобы получались согласованные выходные векторы, то есть чтобы предъявление достаточно близких входных векторов давало одинаковые выходы. Процесс обучения, следовательно, выделяет статистические свойства обучающего множества и группирует сходные векторы в классы. Предъявление на вход вектора из данного класса даст определенный выходной вектор, но до обучения невозможно предсказать, какой выход будет производиться данным классом входных векторов. Следовательно, выходы подобной сети должны трансформироваться в некоторую понятную форму, обусловленную процессом обучения. Это не является серьезной проблемой. Обычно не сложно идентифицировать связь между входом и выходом, установленную сетью.

Для обучения нейронных сетей было рассмотрено два метода: метод градиентного спуска и метод имитации отжига. Алгоритм имитации отжига похож на градиентный спуск, но за счёт случайности выбора промежуточной точки должен будет попадать в локальные минимумы реже, чем градиентный спуск. При использовании закона понижения температуры может быть гарантировано нахождение глобального минимума, с вероятностью, стремящейся к единице, однако на практике это требует слишком большого числа итераций, поэтому напрямую такой подход не используется.

1.1.5 Метод градиентного спуска

Градиентный спуск — метод вычисления градиента, который используется при обновлении весов персептрона, для минимизации значения функции ошибки.

При обучении нейронной сети требуется изменять весовые коэффициенты сети так, чтобы минимизировать среднюю ошибку на выходе нейронной сети при подаче на вход последовательности обучающих входных данных. Чтобы сделать всего один шаг по методу градиентного спуска

(сделать всего одно изменение параметров сети), необходимо подать на вход сети последовательно абсолютно весь набор обучающих данных, для каждого объекта обучающих данных вычислить ошибку и рассчитать необходимую коррекцию коэффициентов сети и уже после подачи всех данных рассчитать сумму в корректировке каждого коэффициента сети (сумма градиентов) и произвести коррекцию коэффициентов «на один шаг».

Функция используется, чтобы контролировать ошибку в прогнозах модели машинного обучения. Поиск минимума означает получение наименьшей возможной ошибки или повышение точности модели.

Итак, градиентный спуск нужен для минимизации функции ошибки. Суть алгоритма – процесс получения наименьшего значения ошибки.

Алгоритм является одним из классических алгоритмов оптимизации и имеет большое количество модификаций [9].

1. Метод автоматического определения множителя скорости градиентного спуска для градиентных методов

Одним из параметров метода является множитель λ , задающий скорость градиентного спуска.

$$X_{k+1} = X_k - \lambda \text{grad}(X_k) \quad (4)$$

Так как для каждого параметра процесса адаптации функции определена область допустимых значений $(0, 1)$, то необходимо рассчитывать множитель скорости таким образом, чтобы не покинуть ОДЗ.

Используемый подход описан ниже:

$$\lambda = \min_{i \in [1, n]} (\text{step}_i) - e, \quad (5)$$

$$\text{step}_i = \begin{cases} \frac{(0-x_i)}{-\text{grad}(X)_i}, & \text{если } \text{grad}(X)_i > 0 \\ +\infty, & \text{если } \text{grad}(X)_i = 0 \\ \frac{(1-x_i)}{-\text{grad}(X)_i}, & \text{если } \text{grad}(X)_i < 0 \end{cases}, \quad (6)$$

где:

X_k – значение вектора на итерации k ,

n – количество компонент вектора X_k ,

$step_i$ – потенциальный множитель, по i -ой компоненте градиента,

x_i – значение i -ой компоненты текущего решения,

$grad(X)_i$ – градиент i -ой компоненты текущего решения,

ϵ – константа уменьшения множителя.

Дополнительное вычитание константы из минимального потенциального шага необходимо для предотвращения попадания следующего решения на граничные значения ОДЗ: 0 или 1.

2. Метод оврагов

Градиентные методы медленно сходятся в тех случаях, когда поверхности уровня целевой функции сильно вытянуты. Этот факт известен в литературе как «эффект оврагов».

Метод градиентного спуска оказывается очень медленным при движении по оврагу, причём при увеличении числа переменных целевой функции такое поведение метода становится типичным. Для борьбы с этим явлением используется метод оврагов, суть которого очень проста. Сделав два шага градиентного спуска и получив три точки, третий шаг следует сделать в направлении вектора, соединяющего первую и третью точку, вдоль дна оврага.

Для учета ограничений в градиентном методе используется метод барьерных функций, таким образом, перед началом обучения составляется смешенная целевая функция.

Пример работы метода представлен на рисунке 8.

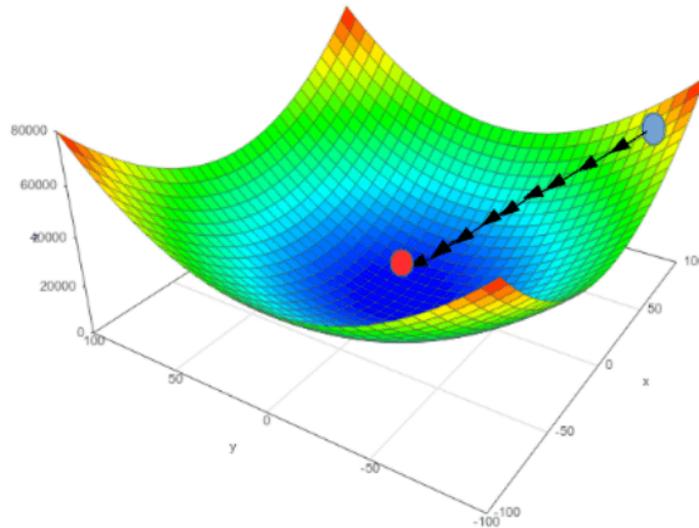


Рисунок 8 – Метод градиентного спуска

1.1.6 Метод имитации отжига

Метод имитации отжига — эвристический метод решения задачи глобальной оптимизации, основанный имитации физического процесса, который происходит при кристаллизации вещества, в том числе при отжиге металлов.

Используя понятие постепенного охлаждения, имитация отжига медленно уменьшает вероятность принятия худших решений по ходу исследования их пространства. Так как допущение худших решений формирует более обширное пространство поиска глобального минимума, имитация отжига предполагает, что возможности соответствующим образом представлены и изучены в первых итерациях. С течением времени алгоритм постепенно переходит от изучения к применению (Рисунок 9).

Ниже приведено примерное описание работы алгоритма имитации отжига:

1. Температура устанавливается в некое изначальное положительное значение и прогрессивно смещается к нулю.
2. В каждый шаг времени алгоритм произвольно выбирает решение, близкое к текущему, измеряет его качество и перемещается к нему в

зависимости от текущей температуры, определяющей вероятность принятия лучших или худших решений.

3. В идеале к моменту достижения нулевой температуры алгоритм сводится к решению глобального минимума.

Имитация отжига особенно хорошо проявляет себя в сценариях, где приблизительное решение нужно получить за короткий период времени, опережая при этом медлительный градиентный спуск. Её можно использовать в тандеме с градиентным спуском, чтобы задействовать преимущества обоих способов: скорость имитации отжига и точность градиентного спуска.

Существует большое количество модификаций алгоритма [9].

1. Метод «тушения»

Модификация метода имитации отжига, которая позволяет работать алгоритму на неоднородных данных и допускает очень быстрое моделирование нового состояния системы через рекуррентную формулу:

$$x'_i = x_i + z_i \cdot (x_i^{max} - x_i^{min}), \quad (7)$$

$$z_i = \text{sign}(\alpha_i - 0.5) \cdot T_i \cdot ((1 + 1/T_i)^{|2\alpha_i - 1|}), \quad (8)$$

где α_i – независимые случайные величины, распределенные равномерно на $[0,1]$.

Использование линейный закон падения температуры позволяет найти решение близкое к оптимальному решению за короткий промежуток времени.

2. Генератор псевдослучайных чисел

Для генерации случайных величин α используется линейный конгруэнтный метод. Генератор определяется рекуррентным соотношением:

$$Y_{n+1} = (aY_n + c) \bmod m, \quad (9)$$

где Y последовательность псевдослучайных значений;

a, c, m – целочисленные константы которые определяют генератор.

Если поделить последовательность Y на m то получится последовательность независимых случайных величин, распределённая равномерно на $[0,1]$.

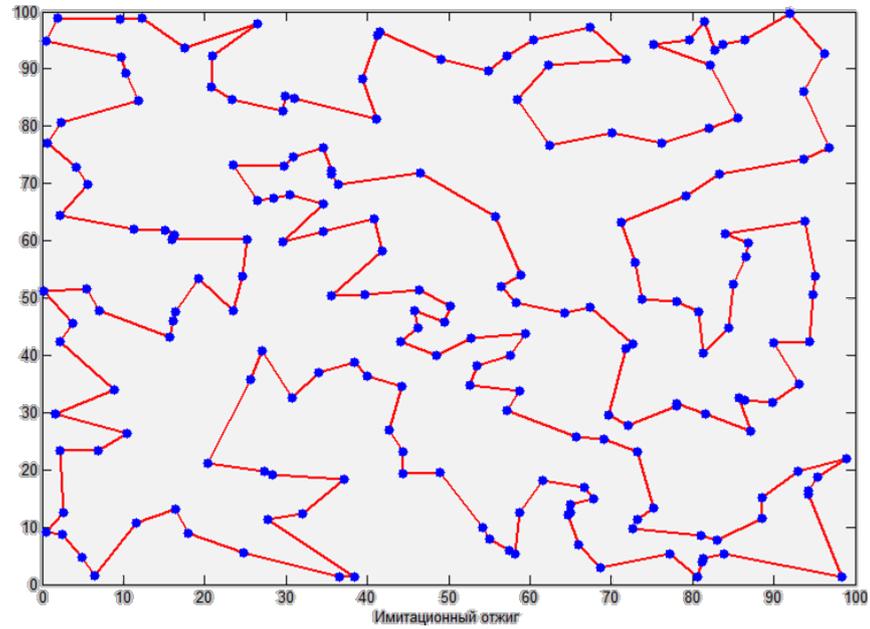


Рисунок 9 – Метод имитации отжига